FACULTEIT PSYCHOLOGIE EN
PEDAGOGISCHE WETENSCHAPPEN

PARENTING AND SPECIAL EDUCATION
RESEARCH UNIT

The Neural Substrates of Children's Arithmetic Fluency

Brecht Polspoel

Doctoral thesis offered to obtain the degree of
Doctor of Educational Sciences (PhD)

Supervisor: Prof. dr. Bert De Smedt
Co-supervisor: Prof. dr. Maaike Vandermosten

2019

Brecht Polspoel

The neural substrates of children's arithmetic fluency

Dissertation submitted to obtain the degree of Doctor in Educational Sciences, 2019.

Supervisor: Prof. dr. Bert De Smedt – Co-supervisor: Prof. dr. Maaike Vandermosten

---

Arithmetic is an essential skill for further mathematical and educational development, and comprises a large part of our daily lives. Even tough, over the past few years, an increase in both functional and structural developmental neuroscientific research on mathematical cognition and arithmetic has occurred, a lot is still not fully understood. This doctoral dissertation aimed to further identify which brain regions are important for typically developing children's arithmetic fluency.

Neuroimaging work in adults has shown that strategy use (i.e., fact retrieval vs. procedures) modulates the arithmetic brain network, however, this was never clearly studied in children. A first fMRI study investigated children's neural activation associated with the use of these strategies in subtraction and multiplication, and observed distinct neural networks associated with each strategy, but no differences between operations when taking strategy into account. Next, within children's multiplication fact retrieval, performance can be influenced by various effects, such as the problem size and interference effect, as evidenced by behavioral research. A second fMRI study investigated the neural basis of both effects, and, concurring with previous studies, revealed clear behavioral effects of problem size and interference, but at the neural level, only a clear effect of problem size. The interference effect was not detected; no clear neural distinctions were observed between low and high interfering items.

Other than its functionality, the structure of grey matter regions has also been associated to cognitive skills. Accordingly, a third study investigated the structural grey matter correlates of children's arithmetic fluency, by looking at both volume and cortical complexity, and observed associations with various cortical grey matter structures. Furthermore, as the grey matter regions of the arithmetic brain network are spatially distant, it is also crucial to study the structural white matter connections between these regions. In a fourth study, the white matter integrity of previously observed arithmetic-related white matter pathways was correlated to children's arithmetic fluency, implementing spherical deconvolution, a novel non-tensor method which goes beyond classic DTI to tackle its methodological constraints. Clear associations were observed between the right inferior longitudinal fasciculus and arithmetic.

Finally, a fifth study investigated the added value of the collected structural brain imaging data on top of well-known behavioral measures in predicting individual differences in the children's arithmetic fluency, and revealed that the neuro-anatomical measures provided the best prediction of performance, highlighting the value of brain imaging measures for the prediction of cognitive skills and striving towards a bridge between cognitive neuroscience and education.

In all, the combination of both functional and structural neuroimaging in these studies has led to results which further expand our understanding of the neural substrates of children's arithmetic fluency.

Brecht Polspoel

De neurale basis van de rekenvaardigheid van kinderen

Proefschrift aangeboden tot verkrijgen van de graad Doctor in de Pedagogische Wetenschappen, 2019.

Promotor: Prof. dr. Bert De Smedt – Copromotor: Prof. dr. Maaike Vandermosten

---

Rekenen is een essentiële vaardigheid voor verdere ontwikkeling en omvat een groot deel van ons dagelijkse leven. Ondanks dat er de voorbije jaren een toename in functioneel en structureel neurowetenschappelijk onderzoek rond numerieke cognitie en rekenvaardigheid bij kinderen heeft plaats gevonden, is er nog veel onduidelijk. Dit doctoraat tracht de hersengebieden die belangrijk zijn voor de rekenvaardigheid van typisch ontwikkelende kinderen verder te identificeren.

Neurowetenschappelijke studies bij volwassenen hebben aangetoond dat strategiegebruik (rekenfeiten of procedurele manipulaties) het hersennetwerk voor rekenen moduleert, maar dit werd nog nooit bestudeerd bij kinderen. Een eerste fMRI studie onderzocht de neurale activatie geassocieerd met deze strategieën tijdens aftrekken en vermenigvuldigen en observeerde diverse neurale netwerken voor beide strategieën, maar geen verschillen tussen de bewerkingen. Vervolgens heeft gedragsmatig onderzoek aangetoond dat prestatie op rekenfeiten bij vermenigvuldigingen beïnvloed wordt door verscheidene effecten, zoals het probleemgrootte- en interferentie-effect. Een tweede fMRI studie onderzocht de neurale basis van deze effecten en vond een duidelijk gedragsmatig effect van zowel probleemgrootte als interferentie, maar alleen een duidelijk neuraal effect van probleemgrootte. Het interferentie-effect werd niet gedetecteerd; er werden geen duidelijke verschillen in neurale activatie gevonden tussen laag en hoog interfererende rekenproblemen.

Buiten functie, werd de structuur van grijze stof ook eerder gekoppeld aan cognitieve vaardigheden. Zo onderzocht een derde studie hoe structurele eigenschappen van grijze stof samenhangen met rekenvaardigheid, door te kijken naar zowel volume als corticale complexiteit. Meerdere associaties werden zo gevonden met verscheidene corticale structuren. Daarenboven liggen de hersengebieden van het rekenkundige hersennetwerk niet naast elkaar, waardoor het ook cruciaal is om de structurele witte stofbanen die deze regio's met elkaar verbinden te bestuderen. In een vierde studie werd de integriteit van deze witte stofbanen bestudeerd aan de hand van spherical deconvolution, een nieuwe non-tensor methode die verder gaat dan klassieke DTI om bepaalde methodologische beperkingen tegen te gaan. Er werden duidelijke associaties gevonden tussen de rechter inferieur longitudinale fasciculus en kinderen hun rekenvaardigheid.

Tot slot onderzocht een vijfde studie de toegevoegde waarde van de geobserveerde structurele neurale correlaten bovenop gedragsmatige predictoren in het voorspellen van individuele verschillen in kinderen hun rekenvaardigheid. Deze studie toonde aan dat de neuro-anatomische maten de beste predictie van prestatie geven, wat de waarde van deze correlaten benadrukt en bijgevolg streeft naar een brug tussen cognitieve neurowetenschappen en onderwijs.

Bij elkaar genomen heeft de combinatie van zowel functionele en structurele neurale maten gezorgd voor resultaten die ons begrip van de neurale basis van de rekenvaardigheid van kinderen uitbreiden.

Acknowledgements

It is hard to describe how excited I am to have finished this doctoral project after years of hard work. It has been an incredible journey in which, throughout a series of rises and falls, I learned so much about scientific research, but also about myself. Consequently, I would like to take the time here to thank a couple of people who have helped me get where I am today.

First and foremost, I want to thank my supervisor, Bert, for believing in me and giving me the opportunity to work on this project. I cannot express how grateful I am for your countless hours of reading and rereading my work, for all the advice you have given me over the years, and for guiding me throughout this project. These past few years of working alongside you have been among the best of my life. I admire your enthusiasm for our field, as well as your work ethic, not to mention your guitar skills, but most of all, I am proud to be able to call you a friend. I also want to take the time to thank my co-supervisor, Maaike, for her time and for always being ready to help me out. Thank you for always providing me with insightful feedback which consistently widened my point of view. I could not have asked for a better team of supervisors and strongly believe that the world is better place with the two of you in it.

I would also like to thank the members of the jury, Prof. Dr. Jérôme Prado, Prof. Dr. Pol Ghesquière, and Prof. Dr. Hans Op de Beeck for their comprehensive reading of this dissertation, and for discussing it with me during my defense. Furthermore, I would like to express additional gratitude to the members of my supervisory committee, Prof. Dr. Pol Ghesquière and Prof. Dr. Hans Op de Beeck, for their feedback throughout this doctoral project.

In the past few years, I have also had the great pleasure to work alongside some wonderful people. Jolijn and Lien, it has been years already, but thank you for mentoring me during my research internship and my master's thesis. The two of you really awakened my interest for research, which is definitely a part of why I started this PhD project, and for which I am very thankful. Elien, a.k.a. Alien Ballon, thank you for being a fantastic colleague and for not going crazy sharing an office with me for two years, but most of all, thank you for being there when I really needed somebody. This Brochette will be forever grateful for that. I also wish to thank all other members of Mathlab for your feedback on my work, as well as other colleagues of Ortho, like Filip and Martijn. I may only have attended a handful of coffee breaks over the last four years, but the small conversations we have had, as well as your general collegiality were greatly appreciated. I have also had the pleasure of doing research with people from other research facilities. Roland, Stephan, Alexander, and Alice, thank you ever so much for your insightful remarks and for genuinely making my research better.

I also wish to thank my family, and most of all my mom and dad for having given me the opportunity to go to university and for standing by me throughout my education. Mom, thank you for showing me the value of hard work, and for, during my education, listening to me whine endlessly about subject matter you probably never even cared for. Dad, thank you for always telling me to do what I want to do in life. I hope both of you know how much I appreciate everything you have ever done for me. On the topic of parents, I also want to thank my surrogate parents, Christel and Michel. Thank you for everything you have done for Laura and me over the years, and for reminding me that life is too short to not laugh on the daily.

Thank you to all my friends for all the good times we've had over the past years (the first names that come to mind are Melvin, Willem, Gert, Karel, Thomas, Lionel… The list goes on). Most of all, I want to thank my *boyz* from Dear. Bert, Bart, and Yorick, thank you guys so much for making every Monday a party. All of you mean the world to me.

Finally, I want to express my sincerest gratefulness for Laura. I could have easily placed you in any of the categories above, as we are coworkers, friends, and family, but you very much deserve a category of your own. Thank you for believing in me, and for making me believe in myself. Thank you for your love and patience, and for making me laugh every single day. It is always better when we are together, and I for one cannot wait to spend the rest of my life with you.

Thank you.

Brecht Polspoel

Table of Contents

Overview of Included Figures

Overview of Included Tables

CHAPTER 1

General Introduction & Aims

Arithmetic refers to the ability to do addition, subtraction, multiplication or division, and is constantly present in our daily lives. For example, we maintain a budget when shopping and compare the prices of different products, or might double recipes based on the amount of people we are cooking for. This ubiquitous presence is especially true for children, who encounter arithmetic during various classes throughout their scholastic careers. The importance of arithmetic and the general ability to efficiently process numerical information, however, really becomes clear when looking at the impact it might have later in life. For example, a study by Gerardi, Goette, & Meier (2013) matched the numerical and cognitive abilities of subprime mortgage borrowers to their administrative mortgage records and provided empirical evidence for a negative association between efficient numerical processing and the tendency to default on one’s mortgage, which could affect a person’s life direly. Furthermore, problems with arithmetic are a key component of dyscalculia (American Psychiatric Association, 2013), in which people experience persistent deficits in acquiring basic mathematical competencies, making the study of arithmetic in general crucial for the study of this learning disorder.

Being able to fluently process numbers and do arithmetic is thus important, yet arithmetic is not as one-sided as one might think. An important aspect of arithmetic fluency lies in the notion that arithmetic problems can be solved, not only through complex procedural manipulations, but also through the fast retrieval of arithmetic facts form long-term memory (e.g., Siegler, 1996; Siegler, Adolph, & Lemaire, 1996). Although a vast body of behavioral literature exists on arithmetic fluency and these arithmetic strategies (e.g., De Smedt, 2016; Geary, Bow-Thomas, & Yao, 1992; Siegler, 1996; Vanbinst, Ghesquière, & De Smedt, 2012), relatively little is known about its neural basis, especially in children, even though knowledge on these neural correlates is highly important from a perspective of educational neuroscience. Placing themselves at the intersection of psychology, education, and cognitive neuroscience, developmental brain imaging studies are exceedingly promising to understand the biological processes that play a role for educationally relevant skills (De Smedt, 2018a; De Smedt & Grabner, 2015; Howard-Jones et al., 2016). They make it possible to investigate the brain during the learning phase of cognitive skills and to see if brain imaging measures are able to predict subsequent learning gains (Hoeft et al., 2011), or if they can predict responses to educational interventions (Supekar et al., 2013). Against this background, the current doctoral project aims to increase our knowledge on the functional and structural neural correlates of typically developing children’s arithmetic fluency.

In this introduction, different arithmetic problem-solving strategies are discussed, elucidating how these strategies are formed and how they can be implemented to fluently and efficiently solve arithmetic problems. This is followed by a summary of the brain imaging techniques used throughout this project. Next, the existing literature on the neural correlates of arithmetic in adults and typically developing children is presented, standing still at shortcomings or gaps of the existing literature. The introduction ends by a disclosure of the concrete aims of this doctoral dissertation.

1.1. Arithmetic strategy use

Throughout arithmetic development, changes occur in the strategies children use to solve arithmetic problems, which has often been studied at the behavioral level (e.g., Bailey, Littlefield, & Geary, 2012; Barrouillet, Mignon, & Thevenot, 2008; Siegler, 1996). At the very beginning of arithmetic development, children tend to count all numbers in a problem in order to get to the solution (e.g., 2 + 3 = 1, 2, 3, 4, 5). As arithmetic development progresses, these counting procedures progress as well, becoming more advanced, such as counting onwards from the largest operand in the problem (e.g., 2 + 3 = 3, 4, 5; Geary et al., 1992). Through the repeated use of such counting strategies, children then develop solid associations of the arithmetic problem and its solution, and thus develop representations of basic arithmetic facts, stored in long-term memory, making it possible to very quickly and automatically solve the problem at hand (Siegler & Shrager, 1984). These arithmetic facts are also relied upon in later arithmetic development, when more advanced procedural problem-solving strategies are needed, such as a tie strategy (e.g., 6 + 7 = 6 + 6 + 1 = 13, in which case the answer to 6 + 6 is retrieved from memory), or a decomposition of operands strategy (e.g., 6 + 7 = 6 + 4 + 3 = 13, in which case the answer to 6 + 4 is retrieved from memory; Siegler, 1996).

Recent studies, however, have observed linear increases in the reaction time of addition problems, along the magnitude of both operands, involving operands up to 4 (Barrouillet & Thevenot, 2013). This suggests that, instead of fact retrieval, children and adults may actually use fast automatized counting procedures for small addition items, as the use of a retrieval strategy would not show such a linear effect (Barrouillet & Thevenot, 2013; Thevenot, Barrouillet, Castel, & Uittenhove, 2016). In other words, the very fast responses for these small items, that are most often interpreted as reflecting the direct retrieval of the answer from long-term memory, might actually be due to compiled automated procedures that are even faster than retrieval, providing an answer while the subject remains unaware of the process, and mistaking it for fact retrieval (Uittenhove, Thevenot, & Barrouillet, 2016).

Either way, arithmetic problems can thus be solved by means of fact retrieval or fast automatized counting procedures, or through a variety of procedural manipulations, such as counting, tie, or decomposition strategies, in which, over development, a shift is observed towards the increased use of a fact retrieval strategy (Siegler, 1996), or from slow to quick counting procedures (Thevenot et al., 2016). However, the strategies used to solve arithmetic items are often dependent on aspects such as problem size (i.e., smaller single-digit items such as 3 + 4 or 3 × 4 are more likely to be solved through fact retrieval than larger multidigit items such as 13 + 8 or 13 × 8) or operation (i.e., as associations between multiplication problems and their answers are explicitly learned by rote in elementary school, multiplication problems are more likely to be retrieved than, for example, subtraction or division problems; Dehaene & Cohen, 1995; Siegler et al., 1996). Approaches for studying arithmetic strategy use based on problem size or operation, however, are rather limited and have been criticized over the past years, as items of a certain operation or size are not necessarily solved through the same strategy

(De Smedt, 2016; Siegler & Stern, 1998). This issue can be resolved though, by measuring strategy use through self-reports on a trial-by-trial basis (Siegler & Stern, 1998). Furthermore, a crucial aspect of arithmetic strategy development lies in the ability to alternate the strategies used to solve certain items. Such strategy flexibility or adaptivity is often defined as using a variety of solution strategies, without any further qualification (e.g. Heirdsfield & Cooper, 2002). However, the mere use of different solution strategies on similar arithmetic problems, without directing any attention to which strategies the children use best, or to the efficiency of the chosen strategy for the problem at hand, can hardly be characterized as adaptive, making it necessary to define such flexibility or adaptivity in terms of subject and task characteristics (Verschaffel, Torbeyns, De Smedt, Luwel, & Van Dooren, 2007). Accordingly, throughout typical arithmetic development, children learn to link each strategy to a particular type or arithmetic problem, leading towards fluent and efficient problem-solving.

Additionally, large subject-variability exists in the use of these strategies, as, with age, an increased use of fact retrieval is observed (Siegler, 1996), and as children differ in the amount they rely on such a fact retrieval strategy (Dowker, 2005; Imbo & Vandierendonck, 2007). This is especially apparent in children with dyscalculia, who rely on retrieval strategies to a much lesser extent (American Psychiatric Association, 2013). Moreover, even within a fact retrieval strategy, individual differences on performance exist, as performance can be affected by various effects, to which people may be more or less sensitive. For example, the problem size effect implies that smaller arithmetic problems are solved faster and more accurately (e.g., Berteletti, Prado, & Booth, 2014; De Brauwer, Verguts, & Fias, 2006; De Smedt, Holloway, & Ansari, 2011; Prado et al., 2013; Prado, Mutreja, & Booth, 2014). This problem size effect is most often explained by the notion that smaller problems are more likely to be retrieved in comparison to larger problems (Zbrodoff & Logan, 2005), but within fact retrieval itself, it could be explained by the frequency theory, which emphasizes that smaller problems appear more frequently and are therefore solved faster and more accurately (Ashcraft & Christy, 1995). Alternatively, this problem size effect could be explained by the distribution of associations model, which states that each problem is associated with previously computed answers, making the amount of errors increase as problem size increases (Siegler, 1988), or the network interfering theory, which suggests that magnitude representations follow a psychophysical scale that is more compressed as magnitude increases, making the representations of large answers more similar to one another than representations of small answers (Campbell, 1995). Another, more recently observed, effect on performance within fact retrieval, and more specifically within multiplication, is the interference effect, which states that when retrieving the answer to a problem, the more similar the problem is to a previously learned one, the more that previous problem will interfere and impact storing in long-term-memory, leading to poorer performance (De Visscher & Noël, 2013; De Visscher & Noël, 2014a; De Visscher & Noël, 2014b). Accordingly, an interference parameter can be calculated for each problem, representing the weight of proactive interference. For example, the first multiplication problems children learn are $2 \times 2 = 4$ and $2 \times 3 = 6$, which have no interference as they do not share two or more digits with one another. However, the next item, 2 × 4 = 8, shares the digits 2 and 4 with 2 × 2 = 4, and is thus subject to interference in memory. This, however, does not mean that problems encountered later are always more interfering, but means that the more similar a certain problem is to previously learned problems, the larger the impact of interference will be. Accordingly, the more sensitive a person is to this interference in multiplication facts storing, the more detrimental the effect can be for arithmetic development.

Individual differences in arithmetic strategy use and performance could also be explained by various other arithmetic-related cognitive skills, such as the domain-specific ability to represent numerical magnitudes (i.e., children with better access to magnitude representations from symbolic digits retrieve more facts from memory and are faster in executing fact retrieval as well as procedural strategies; e.g., Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Schneider et al., 2017; Vanbinst et al., 2012). These individual differences might also be explained by domain-general cognitive correlates of arithmetic, such as working memory, which refers to the capacity of storing information for short periods of time when engaging in cognitively demanding activities (Baddeley, 1986). Here, working memory is important as arithmetic often involves the processing and storing of information simultaneously (e.g., remembering numbers and solutions when using a decomposition strategy for multi-digit problems; Peng et al., 2017). Another relevant domain-general skill is rapid automatized naming (i.e., the fast retrieval of phonological information from long term memory; Koponen et al., 2013), which is important for arithmetic as poor (access to) phonological representations in long-term memory can interfere with the retrieval, manipulation, and retention of phonological codes, which in turn means that if phonological representations for number words or facts in long-term memory are weak, they will be more difficult to retrieve quickly and accurately (Koponen et al., 2013; Simmons & Singleton, 2008).

Finally, it is also worth mentioning that the choice of the used strategy, be it fact retrieval or procedure, or be it the type of procedural strategy used, is highly dependent on cultural aspects, such as the general educational environment in which these skills evolve, as well as the emphasis on automatization processes within the mathematics curriculum (De Smedt, 2016). Behavioral studies have clearly shown cross-cultural differences in retrieval use depending on the emphasis of the math curriculum on fact retrieval and automatization (e.g., Campbell & Xue, 2001). For example, the mandatory guidelines of the Flemish education system, in which all participants of this doctoral dissertation are receiving education, puts a high emphasis on the use of fact retrieval strategies for single digit problems (especially for multiplication), and on the use of a decomposition of operands strategy for larger problems, which is often coupled with a limited attention for, or even prohibition of counting, and differs from other cultures, such as North America. A comparison of the fact retrieval frequencies in single-digit addition and subtraction in Belgian (Torbeyns, Verschaffel, & Ghesquière, 2004) and American (Geary, Hoard, Byrd-Craven, & Desoto, 2004) third-graders even revealed a relative retrieval frequency of 88% for the Belgian children, but only a retrieval frequency of 38% for the American children, clearly emphasizing these cultural differences. It should thus be noted that all studies in the current doctoral dissertation were performed with research samples of Flemish children.

1.2. Magnetic resonance imaging

Throughout this doctoral dissertation, neural data was collected by means of magnetic resonance imaging (MRI). MRI is a non-invasive imaging technique able to provide three dimensional detailed anatomical images without the use of radiation. An MRI scanner employs powerful magnets which produce a magnetic field that forces hydrogen atoms in the body to align with that field. When a radiofrequency current is then pulsed through the person in the scanner, the hydrogen atoms spin out of equilibrium, straining against the pull of the magnetic field. When that radiofrequency current is then turned off, the MRI sensors are able to detect the energy that the hydrogen atoms release when they realign with the original magnetic field. The time it takes for the hydrogen atoms to realign with this magnetic field (i.e., T1 relaxation), and the process by which the transverse components of magnetization decay or diphase (i.e., T2 relaxation), as well as the amount of energy released, depends on the tissue. This makes it possible to tell the difference between various types of tissues based on the collected three dimensional images (McRobbie, Moore, Graves, & Prince, 2006). For this doctoral project, MRI was used to investigate brain functionality, as well as various structural components of the brain, which is elaborated below. An example of MRI images of each acquisition method used in this doctoral dissertation is found in Figure 1.1.

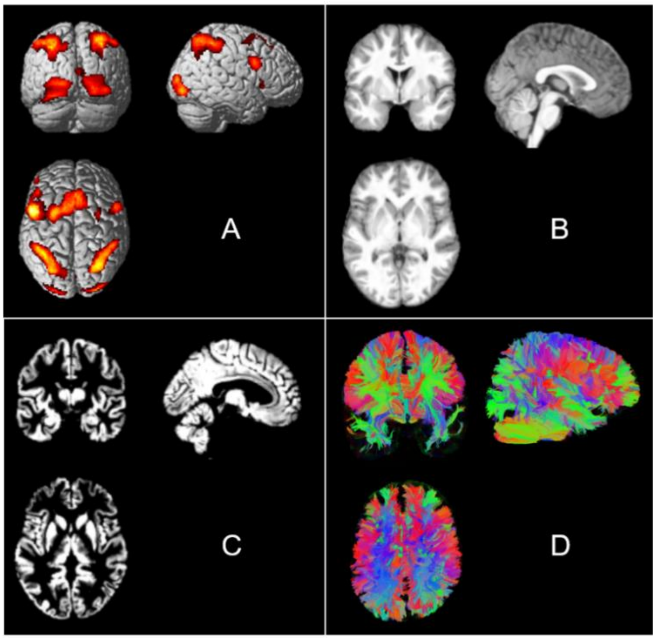

Figure 1.1. Coronal, sagittal, and axial visualizations of the MRI methods used throughout the current doctoral dissertation.

Note: A = fMRI activation maps, B = T1-weigthed structural images, C = extracted cortical surface, D = dMRI tractography.

1.2.1. Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) uses sequences sensitive to T2 relaxation to measure task-related brain activation based on the coupling of neuronal activation and cerebral blood flow (Arthurs & Boniface, 2002). In typical fMRI studies, a blood-oxygen-level dependent (BOLD) contrast is used to measure the hemodynamic response (i.e., the change in oxygenated blood flow), which occurs after brain cells get activated in order to perform a task, and is related to the energy used by those brain cells to perform that task. Accordingly, the BOLD contrast results from the change in magnetic field surrounding the red blood cells depending on the oxygen state of hemoglobin (Glover, 2011).

Studies using fMRI aim to induce different neural states in the brain, by manipulating visual, auditory or other stimuli during scanning. These tests result in activation maps that are a function of the probability that the brain states differ; the activation maps are obtained by comparing the signals recorded during different states. For data analysis, statistical testing for activation is mainly done via a general linear model, in which the effect of each condition (i.e., experimental conditions such as the various arithmetic tasks used for the current dissertation, or control conditions such as looking at a fixation point) on the neural responses is estimated per individual voxel, making statistical analysis possible (Glover, 2011; Huettel, Song, & McCarthy, 2014).

1.2.2. Structural magnetic resonance imaging

In contrast to fMRI, which can thus be regarded as the method providing dynamic physiological information, structural MRI provides static anatomical information of the human brain. The two types of structural MRI used for this doctoral dissertation are discussed below.

1.2.2.1. T1-weighted magnetic resonance imaging

T1-weighted imaging is one of the basic pulse sequences in MRI and demonstrates differences in the abovementioned T1 relaxation times of tissues. These T1 relaxation times thus measure how quickly the magnetic moments of the individual hydrogen atoms being measured recover to their ground state in the direction of the main static magnetic field of the MRI scanner. As not all tissues return back to equilibrium in the same amount of time, different neural structures can be dissociated from each other as they appear in a different shade of grey (e.g., fluids are black, muscle and grey matter are grey, fat and white matter are white; McRobbie et al., 2006).

The most commonly applied method to relate properties from T1-weighted images to cognitive measures is voxel-based morphometry, which uses T1-weighted volumetric MRI scans and makes it possible to perform statistical tests (e.g., a series of t-tests) across voxels to identify grey or white matter volume or density differences between groups. Whether or not volume or density is being investigated, depends on whether or not the segmentation output after nonlinear registration gets modulated, in order to compensate for local changes in volume caused by the alignment process (Ashburner & Friston, 2000). Furthermore, with voxel-based morphometry it is also possible to perform regression analyses across voxels for the assessment of neuroanatomical correlates of cognitive or behavioral skills within a typically developing population, instead of comparing clinical and control groups (Whitwell, 2009).

Although voxel-based morphometry has often been used to study associations of cortical volume and cognitive skills, it only takes cortical volume or density into account, disregarding other structural properties, such as surface shape. New techniques for structural data analysis, however, have arisen over the past years, allowing the study of structural brain differences to go beyond looking at volume alone (Yotter, Ziegler, Nenadic, Thompson, & Gaser, 2011). For example, by extracting the cortical surface from T1 images, surface-based morphometry makes additional metrics of cortical structure applicable, such as cortical complexity through fractal dimensionality, which studies surface shape by quantifying the spatial frequency of gyrification and fissuration of the brain surface, and has been linked to gender, age, but most importantly also to cognitive ability (Luders et al., 2004; King, Brown, Hwang, Jeon, & George, 2010; Im et al., 2006; Madan & Kensinger, 2016; Mustafa et al., 2012; Sandu et al., 2014).

1.2.2.2. Diffusion-weighted magnetic resonance imaging

White matter connections between cortical regions are also integral to efficient cognitive processing, as distant neural regions often cooperate for the completion of cognitive tasks (Johansen-Berg, 2010). Such white matter connections can be specifically examined with diffusion-weighted magnetic resonance imaging (dMRI), in which the MRI signal is sensitized to the random molecular motion, or diffusion, of groups of water molecules (e.g., all water molecules in a voxel) by the addition of diffusion encoding gradients in distinct directions to the standard magnetic pulse (Jones & Leemans, 2011). The diffusion thus represents the displacement of a group of water molecules, which is either isotropic (i.e., the water molecules diffuse freely in all directions) or anisotropic (i.e., diffusion is limited in the direction of an obstruction). This anisotropic diffusion occurs in neural tracts, where the water molecules diffuse more along the longitudinal direction of the tract rather than to the sides (Chilla, Tan, Xu, & Poh, 2015). In dMRI, the diffusion of these molecules is thus exploited to visualize the internal physiology.

The most frequently applied model to relate the dMRI signal to the underlying neurophysiology is Diffusion Tensor Imaging (DTI). Within DTI, it is possible to, for each voxel separately, infer properties such as fractional anisotropy or FA (i.e., the degree to which diffusion has a directional preference), mean diffusivity (i.e., the amount of molecular diffusion), axial diffusivity (i.e., the diffusion rate along the main axis of diffusion), and radial diffusivity (i.e., the diffusion rate in the transverse direction; Soares, Marques, Alves, & Sousa, 2013). The most frequently applied diffusion index to discuss white matter properties is FA, in which the estimated direction of diffusion per voxel is assumed to correspond to the dominant fiber orientation, and the estimated FA is assumed to correspond to the density, myelination and underlying architecture of the underlying axons, which can be related to performance on various cognitive abilities (Basser, Matiello, & LeBihan, 1994; Tournier, Calamante, & Connelly, 2007).

DTI, however, is subject to methodological limitations, as it can only estimate the direction of one fiber per imaging voxel, leading to inaccurate representations of the underlying neuroanatomy in regions with crossing white matter fibers (Assaf, Freidlin, Rohde, & Basser, 2004; Tournier et al., 2007). This is highly problematic, as the percentage of white matter voxels that contain crossing fibers in the human brain is estimated around 70-90% (Farquharson et al., 2013). A second methodological limitation lies in the interpretation of the FA index, which is ambiguous, as it provides a quantitative measure per voxel, determined by microstructural properties, such as the myelination of fibers, or size and density of cells, but also by macrostructural properties, such as the number of crossing fibers. The FA index thus reduces a lot of information to just one number, which in turn means that individual differences in FA could be due to a number of reasons.

These methodological limitations can be resolved through the use of more complex non-tensor models, such as spherical deconvolution, which has the asset that it can characterize the orientation of more than one fiber per voxel (Dell'Acqua et al., 2007, Tournier, Calamante, Gadian, & Connelly, 2004). Furthermore, the hindrance modulated orientational anisotropy (HMOA) index can be derived for quantitative spherical deconvolution analyses, which is defined as the absolute amplitude of each lobe of the fiber orientation distribution, and, in contrast to FA, is fiber-specific, and highly sensitive to changes in fiber diffusivity, such as myelination processes or axonal loss, and to differences in the microstructural organization of white matter, such as axonal diameter and fiber dispersion (Dell'Acqua, Simmons, Williams, & Catani, 2013). Accordingly, even in regions with fiber crossings, the HMOA index provides information about microscopic properties along each fiber orientation.

1.3. Neural correlates of arithmetic

In the next section, an overview will be given of the current literature on the functional and structural neural correlates of arithmetic in both adults and typically developing children.

1.3.1. Adults

At a functional level, arithmetic in the brain has been shown to be constituted of five clusters of brain regions (Figure 1.2; Menon, 2015). First, a cluster in the ventral temporal-occipital cortex is involved in decoding the visual form of the number at hand (Dehaene, Piazza, Pinel, & Cohen, 2003; Grotheer, Jeska, & Grill-Spector, 2018). Second, the intraparietal sulcus and superior parietal lobe are involved in the processing of number magnitude, cardinality, and numerical quantity, which are considered the building blocks from which arithmetic is composed (Menon, 2015). Together, these systems are thus able to build semantic representations of quantity from primary visuospatial events such as eye gazing and pointing (Ansari, 2008; Simon, Mangin, Cohen, Le Bihan, & Dehaene, 2002). Third, activation in the medial temporal cortex, anterior temporal lobe, and angular gyrus is linked to episodic and semantic

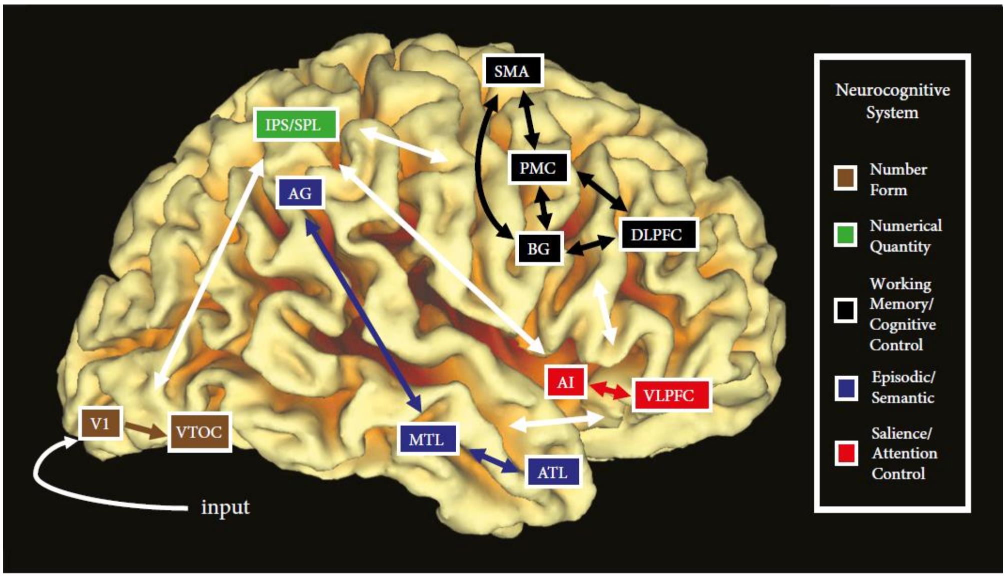

Figure 1.2. Schematic diagram of brain regions involved in arithmetic (Menon, 2015; pp. 503).

Abbreviations: V1 = primary visual cortex, VTOC = ventral temporal-occipital cortex, IPS = intraparietal sulcus, SPL = superior parietal lobe, AG = angular gyrus, MTL = medial temporal lobe, ATL = anterior temporal lobe, SMA = supplementary motor area, PMC = premotor cortex, DLPFC = dorsolateral prefrontal cortex, BG = basal ganglia, AI = anterior insula, VLPFC = ventrolateral prefrontal cortex.

long-term memory systems. Fourth, within arithmetic, systems for procedural memory, working memory, and cognitive control are also often engaged to, for example, manipulate multiple numbers and quantities. These systems are anchored in the dorsolateral prefrontal cortex, pre-motor cortex, supplementary motor area, and basal ganglia. Finally, the anterior insula and ventrolateral prefrontal cortex are important for attentional control processes, making it possible to maintain attention in service of goal-directed problem-solving (Menon, 2015).

When it comes to studying the neural basis of different arithmetic problem-solving strategies, fMRI studies first made assumptions on strategy use by investigating neural activation differences based on problem size. Increased activation for large in comparison to small problems, thus for problem assumed to be solved procedurally, has previously been observed in a fronto-parietal network including the intraparietal sulci, left inferior frontal gyrus, left precentral sulcus, right dorsolateral prefrontal cortex and bilateral cingulate gyri. Activation for small in comparison to large problems, thus problems assumed to be solved through fact retrieval, has been observed in the angular and supramarginal gyri, as well as the right inferior precentral gyrus, right superior temporal gyrus, and left insular cortex (Stanescu-Cosson et al., 2000). Such assumptions on strategy use were also made by studying differences between operations (e.g., Arsalidou & Taylor, 2011; Dehaene & Cohen, 1997; Lee, 2000; Prado et al., 2011; Zhou et al., 2007). For example, a study by Zhou et al. (2007) observed neural differences between addition and multiplication, where addition problems elicited more activation in the intraparietal sulcus, suggesting increased use of procedural strategies for addition, and multiplication elicited more activation in the precentral gyrus, supplementary motor areas, and posterior and anterior superior temporal gyrus, indicating a greater reliance on verbal processing and thus on fact retrieval. Similar results were observed for subtraction and multiplication in a study by Prado et al. (2011), where subtraction was found to be associated with greater activity in the intraparietal sulcus, again suggesting the use of procedural strategies, and multiplication elicited greater activation in regions involved in verbal processing, such as the middle temporal gyrus and the inferior frontal gyrus. Activation differences were also observed in a training study on complex multiplication problems by Delazer et al. (2003), where untrained problems mainly revealed increased activation in the left intraparietal sulcus and inferior frontal lobe, and in the left inferior frontal gyrus, and trained items showed activation differences in the left angular gyrus. None of these studies, however, directly measured the strategies used by the participants on a trial-by-trial basis, which is more suitable for capturing differences between arithmetic strategies, as not all items of a certain operation or problem size are consistently solved through the same arithmetic strategy (Siegler & Stern, 1998). Therefore, studies by Grabner et al. (2009) and Tschentscher and Hauk (2014) studied activation differences in arithmetic strategies through trial-by-trial self-reports, and observed that especially the angular gyri show stronger activation when retrieving, while the use of procedural manipulations leads to activation in the aforementioned fronto-parietal network, including the posterior superior parietal lobe and sensory-motor regions (Grabner et al., 2009; Tschentscher & Hauk, 2014). Noteworthy, however, is that the study by Tschentscher and Hauk (2014) also showed that, when taking arithmetic strategy use into account, no differences between operations could be observed (Tschentscher & Hauk, 2014), confirming that earlier findings on differences in brain activity between arithmetic operations were most likely due to differences in strategy use.

It is thus clear that the arithmetic brain network, as well as arithmetic performance, can be affected by various aspects such as problem size or strategy use. Furthermore, as elucidated above, arithmetic performance, and more specifically performance within multiplication fact retrieval, can also be affected by the more recently discovered interference effect. This interference effect has also been clearly studied at the neurofunctional level, which mainly indicates that activation in the angular gyri is, next to problem size, training, and strategy use, also modulated by interference, as higher activation was found for low interfering items in comparison to high interfering items (De Visscher, Berens, Keidel, Noël, & Noël, 2015; De Visscher et al., 2018), which all reflects an automated mapping between problem and solution, stored in long-term memory in the angular gyri.

Literature on the structural grey matter correlates of arithmetic and mathematical ability in adults is very scarce. A voxel-based morphometry study by Aydin et al. (2007), however, did indicate that cortical gray matter density in the left inferior frontal and bilateral inferior parietal lobes of experts in mathematics (i.e., academics working at departments of mathematics) were significantly increased compared to control subjects, again highlighting the importance for the inferior frontal and parietal lobes within mathematics. Overall, research on the structural neural correlates of arithmetic is thus focused more around white matter connectivity. Previously, FA and radial diffusivity in left parietal white matter have been shown to predict mathematical skill, indicating the left superior corona radiata, corticospinal tract, and superior longitudinal fasciculus as key pathways (Matejko, Price, Mazzocco, & Ansari, 2013). Furthermore, as the functional neuroimaging literature has clearly indicated distinct neural networks for different arithmetic problem-solving strategies (Grabner et al., 2009; Tschentscher & Hauk, 2014), it has also been hypothesized that different white matter networks are likely to support the processing for these different problem-solving strategies. Accordingly, two distinct pathways have been observed to be correlated with fact retrieval or procedural strategies. A first pathway, associated to more difficult problems, more likely to be solved through procedural manipulations, includes both dorsal (i.e., superior longitudinal fasciculus) and ventral (i.e., lateral and inferior to the external capsule) streams. A second pathway, related to problems more likely to be solved by fact retrieval, includes a ventral stream of fibers between fronto-parietal areas corresponding to the middle longitudinal fasciculus (Klein, Moeller, Glauche, Weiller, & Willmes, 2013). Finally, looking at white matter structure and brain function simultaneously, white matter integrity in the left superior corona radiata has also been correlated with arithmetic activation in the left angular gyrus, for problems likely to be solved through fact retrieval (Van Eimeren et al., 2010).

1.3.2. Typically developing children

The overview of the neural correlates of arithmetic above was all based on research in adults (Arsalidou & Taylor, 2011; Menon, 2015) and is not necessarily transferable to children (Ansari, 2010). At the functional level, overall number and arithmetic processing from childhood to adulthood is characterized by a decreasing engagement of the prefrontal cortex and an increasing engagement and functional specialization of the inferior and posterior parietal cortex (Ansari, Garcia, Lucas, Hamon, & Dhital, 2005; Cantlon, Libertus, Pinel, & Dehaene, 2009; Rivera, Reiss, Eckert, & Menon, 2005). Even the short interval from second to third grade is linked to significant task-related changes in brain activation, such as greater activity in both dorsal and ventral visual stream areas (Rosenberg-Lee, Barth, & Menon, 2011), pointing towards a large developmental trajectory in the arithmetic brain network, and highlighting the importance of studying samples with small age ranges as to not miss important neurodevelopmental changes that occur during key stages of academic learning (Menon, 2015). Furthermore, in contrast to adults, increased activation in the hippocampus is often observed in children (e.g., De Smedt et al., 2011; Qin et al., 2014), as the initial transition from counting to memory-based retrieval strategies, as well as longitudinal improvements in fact retrieval use, are paralleled by increased hippocampal engagement (Qin et al., 2014). Accordingly, children’s functional arithmetic brain network is similar but not identical to the adult arithmetic brain network and includes a large set of interconnected frontal (both ventro- and dorsolateral prefrontal cortex), parietal (intraparietal sulcus, angular gyrus, and supramarginal gyrus), occipito-temporal and medial temporal (including the hippocampus) areas. It is, however, crucial to know how exactly these neural correlates develop, in order to understand how instruction affects this development, and to assess and remediate abnormal developmental patterns at an early age (Geary, Bailey, & Hoard, 2009; Rykhlevskaia, Uddin, Kondos, & Menon, 2009).

The abovementioned shift from increased engagement of the intraparietal sulci (i.e., numerical processing) and prefrontal cortex (i.e., working memory) towards an increased reliance on left perisylvian language-related areas, such as the angular gyrus (Houdé, Rossi, Lubin, & Joliot, 2010; Kaufmann, Wood, Rubinsten, & Henik, 2011; Menon, 2015), has also been linked to the clear shift in strategy children use to solve arithmetic problems (i.e., inefficient procedural strategies gradually get replaced with the retrieval of domain relevant facts; Geary et al., 1992; Geary et al., 2004; Siegler, 1996). Previous neuroimaging studies, however, have never explicitly investigated the neural activation associated with different arithmetic problem-solving strategies in children, but have only made assumptions on strategy use based on reaction time, problem size, operation, or presentation (Cho, Ryali, Geary, & Menon, 2011; De Smedt et al., 2011; Peters, Polspoel, Op de Beeck, & De Smedt, 2016; Prado et al., 2014). For example, when comparing single-digit addition to subtraction, De Smedt et al. (2011) observed increased activation in the left hippocampus for addition, and in a fronto-parietal network for subtraction, relating it to retrieval and procedural strategy use, respectively. Prado et al. (2014), on the other hand, contrasted multiplication and subtraction and found grade-related increases of activation for multiplication, assumed to be solved through retrieval, in the left temporal cortex, and activation increases for subtraction, assumed to be solved through procedural strategy use, in the right parietal cortex. These studies, however, made the assumption that the items of the same operation would be solved with the same strategy at the same age. In another study, Cho et al. (2011) assessed children's strategy use in a single-digit addition task and compared participants that solved over 60% of the items with a retrieval or procedural strategy, respectively. Doing so, they observed that the retrievers more strongly activated the left ventrolateral prefrontal cortex, but again, this study did not analyze brain activity during calculation as a function of the strategy used during problem solving. Finally, Peters et al. (2016) investigated brain activation during a subtraction task in symbolic (i.e., Arabic digits or number words) and non-symbolic (i.e., arrays of dots) formats, and showed that the symbolic formats, assumed to be solved by a fact retrieval strategy, showed increased activity in the bilateral angular and supramarginal gyri. Overall, these studies thus point towards increased activation in the left temporal cortex, left hippocampus, left angular gyrus, and left supramarginal gyrus for items assumed to be retrieved, and increased activation in a fronto-parietal network for items assumed to be solved procedurally. As mentioned, however, this approach has been criticized for many years in developmental behavioral studies (e.g., Siegler, 1987; Siegler, 1996), as not all problems of a similar size or operation are always solved through the same strategy. This is especially problematic in the context of developmental research, as the strategies that children use to solve particular types of arithmetic problems change over time (i.e., with education and practice), as do the brain regions responsible for those strategies. For example, the longitudinal study by Qin et al. (2014) showed that the recruitment of hippocampal-dependent memory processes is important in the development of children's memory-based problem-solving strategies (i.e., fact retrieval), yet this hippocampal activation for retrieval strategies is not observed in adults, but rather replaced by activation in the angular gyri (Grabner et al., 2009; Tschentscher and Hauk, 2014). In *Chapter 2* of the current doctoral dissertation, we aimed to resolve these issues by using fMRI to investigate the neural differences between children's arithmetic strategies on a trial-by-trial basis.

Within arithmetic fact retrieval specifically, the previously mentioned effects of problem size and interference have been clearly studied at the behavioral level (e.g., De Brauwer et al., 2006; De Visscher & Noël, 2014b), yet the neural basis of the interference effect and how it differs from the problem size effect was never studied in children. This, however, would be highly interesting as this interference effect has been shown to determine a substantial part of arithmetic performance beyond the problem size effect, and has possible detrimental consequences for the storing of arithmetic facts (De Visscher and Noël, 2014a; De Visscher and Noël, 2014b). Accordingly, in *Chapter 3*, in order to increase our understanding of the development of the arithmetic brain network and its time course, and to potentially help chart the development of individual differences, we set out to investigate the neural basis of both the problem size and the interference effects in children's multiplication fact solving, similar to previous fMRI studies in adults (De Visscher et al., 2015; De Visscher et al., 2018).

The amount of studies looking at the structural grey matter correlates of children's arithmetic, on the other hand, is scarce (Arsalidou, Pawliw-Levac, Sadeghi, & Pascual-Leone, 2018; Peters and De Smedt, 2018). Using voxel-based morphometry, however, the existing studies investigated associations between grey matter volume and arithmetic within typically developing children, and mainly observed positive correlations between arithmetic fluency and grey matter volume in the left intraparietal sulcus (Li, Hu, Wang, Weng, & Chen, 2013; Price, Wilkey, Yeo, & Cutting, 2016), between arithmetic growth in primary school and grey matter volume in the posterior parietal cortex, ventral occipito-temporal cortex, and prefrontal cortex (Evans et al., 2015), and between third graders' learning gains of one-on-one tutoring sessions and the volume of the right hippocampus (Supekar et al., 2013). When comparing groups of children who differ in their level of arithmetic skill (Isaacs, Edmonds, Lucas, & Gadian, 2001; Ranpura et al., 2013; Rotzer et al., 2008; Rykhlevskaia et al., 2009), increased volume for the children that performed better was observed in the bilateral intraparietal sulci, left inferior frontal gyrus, bilateral middle frontal gyri, and bilateral fusiform gyri, concurring to the arithmetic brain network described above. A lot of these studies, however, were conducted in research samples with wide age ranges, which is problematic as, even though statistically controlled for, the study of samples with wide age ranges might lead to maturational confounds and to a possible over-interpretation of the observed results. Furthermore, these existing studies are rather limited as they only took cortical volume into account, without regarding other relevant structural properties, such as cortical shape, which has also been related to cognitive ability (e.g., Im et al., 2006; Sandu et al., 2014). In *Chapter 4* of the current doctoral dissertation, we aimed to investigate the structural neural correlates of children’s arithmetic fluency, but looked beyond cortical volume by also implementing cortical complexity analyses in typically developing 9- to 10-year-old children.

Research on structural connectivity in children’s arithmetic is also scarce and inconclusive, as many different white matter pathways have been found to be related to individual differences in arithmetic or other mathematical skills (Matejko and Ansari, 2015; Moeller, Willmes, & Klein, 2015). These pathways include (1) the arcuate fasciculus, generally associated to phonological, reading and language skills, (2) the superior longitudinal fasciculus, often confused for the arcuate fasciculus and important for regulating motor behavior, spatial attention, and language skills, (3) the inferior longitudinal fasciculus, mainly involved in visual processing, (4) the inferior fronto-occipital fasciculus, related to orthographic processing, (5) the uncinate fasciculus, associated with memory, (6) the corona radiata and corticospinal tract, important for efficient motor functions, and (7) the corpus callosum, connecting both cerebral hemispheres. (Li et al., 2013; Kucian et al., 2013; Navas-Sánchez et al., 2014; Rykhlevskaia et al., 2009; Tsang, Dougherty, Deutsch, Wandell, & Ben-Sachar, 2009; Van Beek, Ghesquière, Lagae, & De Smedt, 2013; Van Eimeren, Niogi, McCandliss, Holloway, & Ansari, 2008). These studies, however, applied classic DTI to study the various white matter tracts, which, as mentioned, is subject to methodological limitations (e.g., Assaf et al. 2004; Dell'Acqua et al. 2013; Farquharson et al. 2013; Tournier et al. 2007). Research using novel imaging techniques to tackle these limitations in the field of arithmetic, however, do not exist. Furthermore, the existing studies were again often conducted in research samples with wide age ranges (e.g., 7- to 10- or 10- to 15-years-old; Matejko & Ansari 2015), which possibly has missing neurodevelopmental processes that only occur at a single point in arithmetic development as a consequence. Therefore, in *Chapter 5* of this doctoral dissertation, the structural white matter correlates of 9- to 10-year-old children’s arithmetic fluency was investigated via spherical deconvolution.

1.4. Aims of the doctoral project

In this doctoral project, we aimed to contribute knowledge on both the functional and structural neural basis of typically developing children’s arithmetic fluency. This was done through two functional imaging studies, two structural imaging studies, and one study looking at the predictive value of various behavioral and structural brain imaging measures for children’s arithmetic fluency simultaneously.

First of all, as mentioned, the neural basis of arithmetic strategy use has not been properly studied in children, as only assumptions on strategy use have been made, based on reaction time, problem size, operation, or stimuli presentation (see De Smedt, 2016, for a discussion). Trial-by-trial self-reports are more appropriate to capture arithmetic strategy use, as they allow for the estimation of individual differences in the choice of strategy (Siegler & Stern, 1998). Neuroimaging studies in adults have already used such an approach and have shown that the strategy used to solve arithmetic items modulates activation in the arithmetic brain network (Grabner et al., 2009; Tschentscher & Hauk, 2014). In children, however, this has never been clearly studied. In *Chapter 2*, we therefore investigated the neural activation associated with both fact retrieval and procedural manipulations in subtraction and multiplication in typically developing children, by verifying the used strategy on a trial-by-trial basis, and contrasting both strategies and operations to see if, similar to adults, observed operation differences disappear when taking strategy into account.

Second, within the development and acquisition of arithmetic facts, especially in multiplication, performance can be influenced by various effects, such as the problem size (De Brauwer et al., 2006) and interference effect (De Visscher & Noël, 2013; De Visscher & Noël, 2014a; De Visscher & Noël, 2014b), as evidenced by behavioral research. In *Chapter 3*, we aimed to achieve a more detailed understanding of the regions in the neural network of children's arithmetic fact retrieval (e.g., regarding the interpretation of activation in the angular gyrus), by studying the functional neural basis of both of these effects in typically developing children, as was previously done in adults (De Visscher et al., 2015; De Visscher et al., 2018).

Third, only a small amount of studies have looked at the structural neural correlates of children's arithmetic, and those that did mainly implemented voxel-based morphometry, which only takes the volume of regions into account, without looking at other structural properties, and often used research samples with wide age ranges, which might lead to the over-interpretation of observed associations (e.g., Evans et al., 2015; Isaacs et al., 2001; Price et al., 2016). Therefore, in *Chapter 4*, the structural neural correlates of typically developing children's arithmetic fluency were investigated, not only through voxel-based morphometry, but, as to gain a more holistic view on these structural correlates, also through cortical complexity analyses (i.e., fractal dimensionality analyses) to also take the shape of cortical structures into account.

Fourth, within mathematical cognition, some studies exist on the structural white matter connections in children's arithmetic, but these studies implemented classic DTI (which has methodological limitations), often used samples with wide age ranges, and correlated structural neural markers to various broad mathematical skills (see Matejko & Ansari, 2015, and Moeller et al., 2015 for reviews). In *Chapter 5*, these previously observed arithmetic-related white matter pathways were studied and correlated to a specific arithmetic fluency task in 9- to 10-year-old children, implementing spherical deconvolution.

Finally, a lot of behavioral and neural research has thus already been done on children's arithmetic to, on the on hand, try and gain insights into how arithmetic fluency develops and into which cognitive factors might explain or predict individual differences in arithmetic, and, on the other hand, try and unravel the neural basis of these individual differences. However, in contrast to research on other cognitive skills such as reading (Hoeft et al., 2011), hardly any studies have studied the various relevant behavioral and neural correlates of arithmetic simultaneously, in order to examine the added value of one type of predictor over and above the other type. Therefore, in *Chapter 6*, we investigated the added value of the structural brain imaging measures collected in *Chapters 4 and 5* in predicting individual differences in children's arithmetic fluency on top of well-known behavioral predictors, based on previous behavioral research, and including symbolic numerical magnitude processing (e.g., Schneider et al., 2017), working memory (e.g., Peng, Namkung, Barnes, & Sun, 2016), and rapid automatized naming (Hecht, Torgesen, Wagner, & Rashotte, 2001).

For the entire doctoral project, behavioral and neuroimaging data of 50 children were collected. For all particpants of Chapter 2 (n = 26), behavioral data collection included the assessment of intelligence, arithmetic, reading, number comparison, working memory, rapid automatized naming, and motor reaction time, as well as a study-specific task on arithmetic strategy use. Neuroimaging data collection included fMRI (on arithmetic strategy use), T1-weighted MRI (to be used in Chapter 4), and dMRI (to be used in Chapter 5). Participants of Chapter 3 (n = 24) went through the same behavioral data collection, except for a study-specific task on interference and problem size instead of on arithmetic strategy use. Neuroimaging data collection again included fMRI (now on interference and problem size), T1-weighted MRI (to be used in Chapter 4), and dMRI (to be used in Chapter 5). For Chapters 4, 5 and 6, the collected behavioral data and structural neuroimaging data of Chapters 2 and 3 were thus combined to form the total research sample (n = 50).

In all, the current doctoral project attempted to complement the observed gaps in the current literature and intended to reevaluate previous studies by implementing novel methods of analyzing brain imaging measures, and by focusing on arithmetic fluency in research samples with a narrow age range to minimize maturational confounds. Accordingly, all studies were performed in 9- to 10-year-old children (i.e., 4th graders), which is a point in development where considerable arithmetic knowledge has already been automatized.

CHAPTER 2

Strategy over Operation:
Neural Activation in Subtraction and Multiplication during Fact Retrieval and Procedural Strategy Use in Children

Published as

Polspoel, B., Peters, L., Vandermosten, M., & De Smedt, B. (2017). Strategy over operation: Neural activation in subtraction and multiplication during fact retrieval and procedural strategy use in children. *Human Brain Mapping*, *38*(9), 4657-4670.

Abstract

Arithmetic development is characterized by strategy shifts between procedural strategy use and fact retrieval. The current study is the first to explicitly investigate children's neural activation associated with the use of these different strategies. Participants were 26 typically developing 4th graders (9- to 10-year-olds), who, in a behavioral session, were asked to verbally report on a trial-by-trial basis how they had solved 100 subtraction and multiplication items. These items were subsequently presented during functional magnetic resonance imaging (fMRI). An event-related design allowed us to analyze the brain responses during retrieval and procedural trials, based on the children's verbal reports. During procedural strategy use, and more specifically for the decomposition of operands strategy, activation increases were observed in the inferior and superior parietal lobes (intraparietal sulci), inferior to superior frontal gyri, bilateral areas in the occipital lobe, and insular cortex. For retrieval, in comparison to procedural strategy use, we observed increased activity in the bilateral angular and supramarginal gyri, left middle to inferior temporal gyrus, right superior temporal gyrus, and superior medial frontal gyrus. No neural differences were found between the two operations under study. These results are the first in children to provide direct evidence for alternate neural activation when different arithmetic strategies are used and further unravel that previously found effects of operation on brain activity reflect differences in arithmetic strategy use.

2.1. Introduction

To date, relatively little is known about the neural substrate of arithmetic in children, an academic skill of clear importance in everyday life (e.g., we maintain budgets, or work with proportions when cooking). On the other hand, accumulating evidence in adults is suggesting that a fronto-parietal network, which includes the superior and inferior parietal lobes, the inferior frontal gyri and the insular cortex, is consistently being activated during arithmetic (for a review, see Arsalidou & Taylor, 2011; Menon, 2015). Most of these studies, however, did not directly take strategy use – arithmetic problems can be solved through fact retrieval or by means of procedural manipulations – into account. This is important, as strategy use has been shown to modulate the adult arithmetic brain network (Grabner et al., 2009; Tschentscher & Hauk, 2014). Studies in children have never directly investigated the neural activity during these strategies. The current study is therefore the first to investigate neural activation during arithmetic while taking into account individual differences in children's arithmetic strategy use, which is crucial considering the large developmental changes in children's acquisition of arithmetic strategies (Siegler, 1996).

Functional magnetic resonance imaging (fMRI) research in adults has often implicated dorsal parts of the parietal cortex, including the intraparietal sulcus, as a critical hub for the representation and manipulation of numerical quantity (e.g., Ansari, 2008; Cohen Kadosh, Lammertyn, & Izard, 2008; Dehaene et al., 2003). These regions make up the magnitude code of the adult Triple Code Model as postulated by Dehaene and Cohen (1997). This model also proposed a visual code, located in bilateral inferior ventral occipito-temporal regions, in which numbers are represented as identified strings of digits. The Triple Code Model also postulated a verbal code, located in left-hemispheric temporo-parietal areas, in which numbers are phonologically represented, and which is implicated in accessing arithmetic facts (Dehaene & Cohen, 1997). Although many brain imaging studies in numbers and arithmetic have limited their focus to the parietal cortex, many areas outside the parietal cortex are also involved in arithmetic (Arsalidou & Taylor, 2011; Menon, 2015). Calculation places a demand on various cognitive systems (e.g., working memory or cognitive control), and thus multiple regions, such as the anterior insula and anterior cingulate cortex for directing attention, and the ventro- and dorsolateral prefrontal cortex for effortful maintenance and manipulation of information, respectively, are also typically activated during calculation (e.g., Arsalidou & Taylor, 2011; Menon, 2015).

Only a very small number of fMRI studies have investigated the functional properties of the arithmetic network in children (Menon, 2015; Peters & De Smedt, 2017, for a review). This network involves a large set of interconnected areas that include frontal (both ventro- and dorsolateral prefrontal cortex), parietal (intraparietal sulcus, angular gyrus, and supramarginal gyrus), occipito-temporal and medial temporal (including the hippocampus) areas. This network shows some similarities to the network observed in adults, but it is a clearly different network which is recruited by children, particularly during the development of arithmetic facts (Menon, 2015; Peters & De Smedt, 2017). For example, adults typically show activation increases in angular and supramarginal gyri during more easy problems, which are likely to be solved by fact retrieval, yet these changes in brain activity have not been consistently observed in children. Moreover, increased activity in the hippocampus has been observed in the early stages of learning arithmetic, more specifically in addition (e.g., De Smedt et al., 2011). It is important to take into account, however, that previous fMRI studies in children used samples with wide age ranges, which may have affected these reported findings. This is particularly relevant, as arithmetic development is characterized by a decreasing engagement of the prefrontal cortex and by an increasing engagement and functional specialization of the inferior and posterior parietal cortex (Kucian, von Aster, Loenneker, Dietrich, & Martin, 2008; Rivera et al., 2005). Even the short interval from second to third grade is linked to significant task-related changes in brain activation, such as greater activity in both dorsal stream parietal and ventral visual stream areas (Rosenberg-Lee et al., 2011), pointing towards a large developmental trajectory in the arithmetic brain network (Menon, 2015). In all, this suggests a need for studies that focus on one particular age range.

As mentioned, arithmetic problems can be solved through different strategies. Furthermore, a well validated fact through decades of behavioral research (e.g., Siegler, 1996), is that, over development, changes occur in the strategies that children use, yet this has never been explicitly investigated in children through fMRI. These strategies can be categorized as retrieval (i.e., remembering the solution to a certain problem) or procedure. Such a procedural strategy is used when the solution to a certain problem cannot be directly retrieved from memory, and procedural manipulations, such as counting or the decomposition of operands (e.g., $24 - 7 = 24 - 4 - 3 = 20 - 3 = 17$ or $3 \times 13 = (3 \times 10) + (3 \times 3) = 30 + 9 = 39$) are needed. When learning to solve arithmetic problems, children initially rely heavily on effortful and time consuming procedures, such as counting. Repeated use of counting, however, will lead to the formation of associations between a problem and its solution, which will in turn lead to the retrieval of the correct answer whenever that problem is presented (Siegler & Shrager, 1984).

Previous studies, particularly neuroimaging work, however, have often made assumptions on strategy use based on reaction time, problem size, or operation, yet it is crucial to emphasize that these approaches are limited (see De Smedt, 2016, for a discussion), as not all problems of a particular size or operation are solved with the same strategy at the same point in development (Siegler, 1987). For example, behavioral studies have shown that, even in adults, single-digit arithmetic items are sometimes solved by procedures, such as counting (LeFevre, Sadesky, & Bisanz, 1996). Ignoring the use of different strategies is especially problematic in the context of developmental research, as, with education and practice, the strategies that children use to solve particular types of problems change over time. Therefore, trial-by-trial self-reports, which (in children) have sufficient reliability and validity (Siegler & Stern, 1998), might be more appropriate to capture arithmetic strategy use, as it allows for the estimation of individual differences in the choice of strategy.

Adult fMRI studies have recently started to take arithmetic strategy use into account, and revealed that the choice of strategy modulates brain activity during arithmetic (Grabner et al., 2009; Tschentscher & Hauk, 2014). In an event-related fMRI study, Grabner et al. (2009) provided the first evidence for alternate neural activation when using different strategies. The study implemented trial-by-trial self-reports in adults immediately after scanning, by asking the participants how they had solved the items during scanning. Their results pointed out that adults show stronger activation of the left angular gyrus when retrieving, while procedural strategy use leads to activation in a more widespread fronto-parietal network. Extending these results, Tschentscher & Hauk (2014) also used strategy self-reports, during addition and multiplication in adults, and mainly found increased activation in the bilateral angular gyrus for fact retrieval. For procedural strategy use, increased activation was observed in the prefrontal cortices, motor areas, posterior superior parietal lobe, and intraparietal sulcus. The results of these studies are thus in line with the idea that the angular gyrus supports retrieval processes in adults, while more activation in the posterior superior parietal lobe and sensory-motor regions is linked to procedural strategy use. It remains to be determined if a similar pattern of findings can be found in children.

Importantly, Tschentscher & Hauk (2014) did not observe any effect of arithmetic operation on brain activity once arithmetic strategy was taken into account. This suggests that earlier findings on differences in brain activity between arithmetic operations (e.g., Arsalidou & Taylor, 2011; Dehaene & Cohen, 1997; Prado et al., 2011; Zhou et al., 2007) should be interpreted with great caution. In children, such operation effects have also been reported repeatedly (e.g., De Smedt et al., 2011; Prado et al., 2014). More specifically, De Smedt et al. (2011) observed increased activity in the left hippocampus for single-digit addition in comparison to subtraction, and in a fronto-parietal network for subtraction in comparison to addition, while Prado et al. (2014) found grade-related increases of activity for multiplication, but not for subtraction, in the left temporal cortex, and increases of activity for subtraction, but not for multiplication, in the right parietal cortex. These studies, however, did not take the participants' strategy use into account, and were only able to make implicit assumptions about strategies, as they assumed that the items of the same operation would be solved with the same strategy at the same age. Consequently, it is still unclear how strategy use modulates activation of the arithmetic brain network in children, and whether previously found operation effects in children might be due to differences in arithmetic strategy use.

Interestingly, Cho et al. (2011) assessed children's strategy use through verbal reports with a single-digit addition task prior to scanning, but categorized their participants into retrievers and counters if they had solved over 60% of the items with a retrieval or procedural strategy, respectively. Subsequently, they compared the brain activity of those two groups during an addition task and observed that the retrievers more strongly activated the left ventrolateral prefrontal cortex. Here as well, brain activity during calculation was not analyzed as a function of the strategy used during problem solving. Furthermore, the study only included one operation (i.e., addition), leaving it open whether these results were transferable to different operations. Research which uses trial-by-trial self-reports, and checks the neural activation patterns for different strategies in different operations, is thus yet to be done in children.

The present study is the first to investigate children's neural activation during calculation, as a function of strategy use, determined on a trial-by-trial basis. The study follows an approach similar to Grabner et al. (2009) and Tschentscher & Hauk (2014) in adults, who implemented trial-by-trial strategy assessment outside of the scanner in order to analyze strategy use for each item separately. Such a trial-by-trial approach is even more needed in children, given that children are more likely to implement a variety of strategies, and that, during development, changes in strategy use occur (Siegler, 1996).

We developed an arithmetic task designed to elicit retrieval (single-digit items) or procedural (double-digit items) strategies. The task included both subtraction and multiplication, allowing us to investigate potential operation effects. Strategy use was recorded on a trial-by-trial basis and the task was administered approximately three weeks prior to scanning. During fMRI acquisition, children were presented with a subset of problems of the strategy assessment task, implemented in a 2 × 2 full factorial design (strategy: retrieval vs. procedure × operation: subtraction vs. multiplication). We employed an event-related design that, for each child individually, allowed us to use the trial-by-trial strategy data, obtained prior to scanning, to categorize each trial during scanning into retrieval or procedure.

In light of the existing literature, we expected to find increases in activation in the hippocampus (based on the strategy assumptions of developmental literature; De Smedt et al., 2011; Qin et al., 2014) for fact retrieval trials. For procedural trials, we predicted an increase in brain activity in a more widespread fronto-parietal network (De Smedt et al., 2011) as has been observed in adults (Grabner et al., 2009; Tschentscher & Hauk, 2014).

As the current study used both subtraction and multiplication, we were also able to test differences between operations and possible interaction effects between strategy use and operation. This allowed us to directly verify whether the previously observed operation effects in children (De Smedt et al., 2011; Prado et al., 2014) reflect differences in strategy use. If the latter is the case, then operation effects will disappear when these strategies are taken into account, as has been observed in adults (Tschentscher & Hauk, 2014).

It is important to point out that, different from most of the existing developmental fMRI studies in the field of mathematical cognition, we have focused our study on children with a very narrow age range (i.e., only 4th graders). This is crucial, as merging data across wide age ranges could lead to missing important neurodevelopmental changes, given that substantial differences in brain activity can already be observed after one year of schooling (Rosenberg-Lee et al., 2011). By minimizing the variability in age, we reduced potential effects of different stages of development and of the received amount of mathematics instruction.

2.2. Methods

2.2.1. Participants

Participants were 26 typically developing Flemish 4th graders (ages 9 to 10), with no history of learning difficulties, or neurological or psychiatric disorders. All children were recruited via the elementary school they attended. Data of six children, however, were discarded, five of which due to excessive motion during functional scanning (see details below), one due to technical acquisition problems. We thus analyzed data of 20 children (M = 9.6, SD = 0.29; 13 boys, 7 girls; 1 left-handed). In return for participating, all children were given a financial compensation. Written informed consent was obtained from a parent or legal guardian of each participating child. The study was approved by the Medical Ethical Committee of the University of Leuven.

2.2.2. Procedure

All children took part in two test sessions. The first session, during which only behavioral data were collected, always preceded the second one by approximately three weeks (M = 21.92 days, SD = 6.13), and included both standardized and strategy assessment. The second session included the actual acquisition of MRI data.

2.2.2.1. Standardized assessment

Standardized assessment consisted of the evaluation of arithmetic, reading, and intellectual ability. Arithmetical competence was measured by the Tempo Test Arithmetic (TTA; de Vos, 1992); a standardized test of arithmetical fluency, similar to the Math Fluency subtest of the Woodcock-Johnson III tests of Achievement (Woodcock, McGrew, & Mather, 2003). The test exists of five columns of arithmetic items, increasing in difficulty (one column per operation and a column with mixed operations); each child gets one minute per column to provide as many correct answers as possible. Reading ability was assessed using a combination of the One-Minute Test (OMT; Brus & Voeten, 1979) and the Klepel (Van den Bos, Spelberg, Scheepstra, & De Vries, 1994), which measure the reading of words and pseudowords, respectively; both tests consist of 116 words. For the OMT, the children get one minute to correctly read aloud as many words as possible; for the Klepel, the time limit is set to two minutes, and the children read aloud pseudowords. Finally, an index of intellectual ability was measured by the WISC-III-NL Block Design and Vocabulary subtests, as measures of performance and verbal IQ respectively (Wechsler, 2005). Standardized scores were calculated for all tasks. Figure 2.1 displays box plots with the descriptive statistics of this cognitive assessment. These results show that the means of our sample were close to the population averages, and show proper variation – especially for the TTA – as is expected in the general population. It is important to note that even though the minimum score for the TTA was low, none of the participating children had been diagnosed with learning disabilities, or dyscalculia in particular.

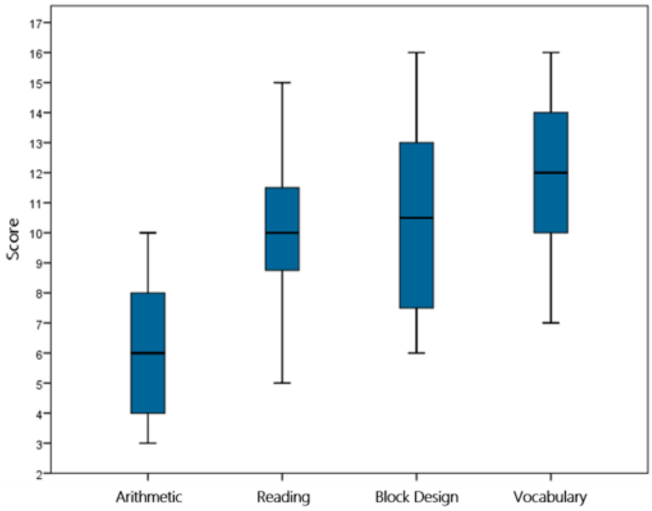

Figure 2.1. Box plots displaying the performance of the standardized assessment.

Note: The scores above are standardized scores. The scores on the arithmetic test are standardized as *M* = 5, *SD* = 2, with a maximum of 10. The scores on the other tests are standardized as *M* = 10, *SD* = 3, with a maximum of 20.

2.2.2.2. Strategy assessment

Strategy use was assessed by a task in which the children were read aloud 100 arithmetic problems, and were asked to verbally solve them. They had to report how they had solved each item on a trial-by-trial basis; children were allowed and encouraged to use any strategy they wanted. The 100 items were divided into 50 subtraction and 50 multiplication items, each of which in turn were divided into items that were a priori expected to elicit either a retrieval or procedural strategy. The problems were presented in a pseudo-randomized order (i.e., never more than five consecutive items of the same operation). For each item, the children's accuracy and used strategy was registered.

In subtraction, the retrieval items consisted of two single-digit operands (e.g., 8 – 3), which have been indicated to be mainly answered through fact retrieval by previous verbal report data in children of a similar age range and math curriculum (Vanbinst et al., 2012). The procedural items crossed the bridge of either 20 or 30 (i.e., the first operand varied from 21 to 28 or 31 to 38, the second operand varied from 4 to 9, solutions varied from 12 to 19 or 22 to 29; e.g., 25 - 8). These items were expected to be solved procedurally, as it is unlikely for them to be stored in memory; as the children reported in the verbal reports, multiple steps were needed to find the answer. In multiplication, the retrieval items existed of two single-digit operands (e.g., 4 × 3), which, by previous verbal report data in a similar sample of children, have also been indicated to be mainly answered through retrieval (Imbo & Vandierendonck, 2008). As all participating children came from Flemish schools, which have a high emphasis on fact retrieval for all single digit multiplication items, it was impossible to use a homogenous subset of single-digit problems to investigate procedural strategy use. Therefore, for the procedural items, one operand varied from 3 to 6 and the other from 12 to 16, leading to solutions between 35 and 100 (e.g., 3 × 14). As multiplication tables beyond 10 are not taught in our curriculum, it is, again, unlikely that, in our sample, these items were stored in memory; as the children reported, multiple steps were needed to find the answer to these problems, making it very likely that these items would be solved by using procedural strategies.

After solving each item, the children were asked how they had solved it; responses were categorized as retrieval (i.e., the participant knew the answer without any sign of overt calculations), procedural (i.e., the participant indicated to have used any form of procedural strategy, such as counting or the decomposition of operands; the type of procedural strategy was also registered) or undefined (i.e., the participant did not know how (s)he had solved the item or used an unclear strategy); this last category was rare as it only occurred in 1.04% of the items.

2.2.2.3. fMRI experimental design

Each participant was presented with a set of 80 of the 100 problems of the strategy assessment task (i.e., 20 items per operation, per expected strategy). Stimuli were presented with E-prime 2.0 (Psychological Software Tools, Pittsburgh, PA), via an NEC projector onto a screen behind the participants, made visible through a mirror attached to the head coil. All stimuli were presented in white (Arial, font size 60) on a black background. Problems were presented horizontally in Arabic digits, and after two seconds two possible answers (a correct and an incorrect one) were simultaneously presented (Figure 2.2). The children were asked to indicate the correct answer by pressing the left or right button on the response box for the left or right response alternative, respectively. For the subtraction items, incorrect answers were created by adding or subtracting 1 or 2 from the correct answer. For the multiplication items, incorrect answers were created by adding or subtracting the value of the smallest operand to or from the correct answer; the proposed false answers were thus always a table related product. When choosing the

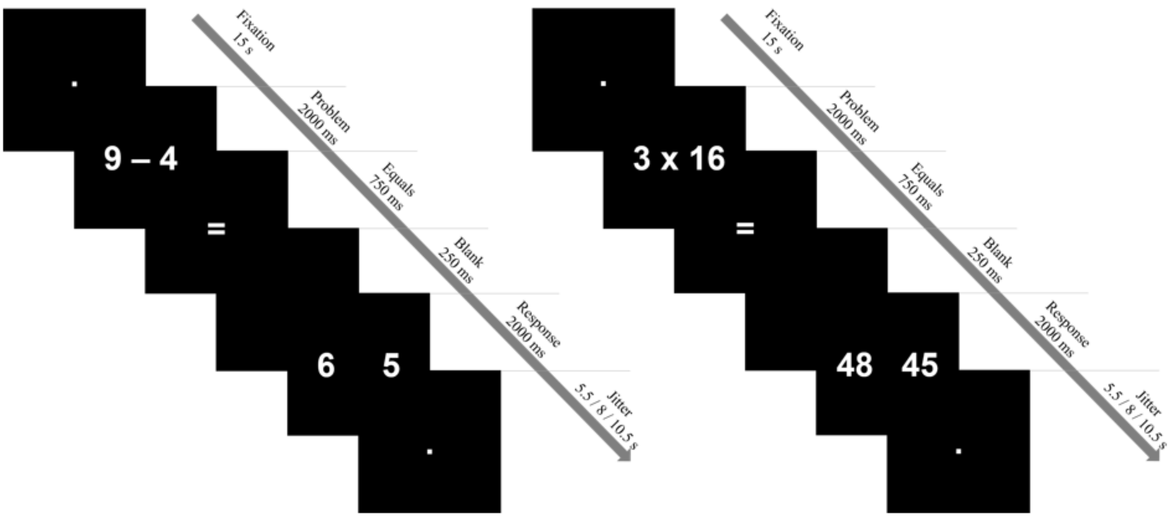

Figure 2.2. Schematic overview of an expected retrieval (left) and procedural (right) trial.

proposed false answers, however, we made sure the equations could not be solved more easily by applying a certain rule, such as the five rule (e.g., in the item 2 × 9, the proposed false answer was 16 instead of 20; in the item 6 × 14, the proposed false answer was 78 instead of 90). The position of the correct answers was balanced.

The task was presented across four functional runs in an event-related fMRI design (similar to De Smedt et al., 2011 and Grabner et al., 2009). Each run consisted of 15 s of fixation at the start of the run, 20 trials (5 items per operation, per expected strategy), and 15 s of fixation at the end of the run. Every trial included the presentation of a problem (2000 ms), followed by a centered equality sign and a blank screen (750 ms and 250 ms, respectively), followed by the presentation of the response alternatives (2000 ms). In between trials, a jittered inter-trial interval of 5.5, 8 or 10.5 s (averaged at 8 s) was randomly added to enable the deconvolution of the hemodynamic response functions (see Figure 2.2 for a schematic overview of a trial). The children were asked to answer as accurately and quickly as possible and were allowed to answer during both the presentation of the response alternatives and the inter-trial fixation period. Only the first 5000 ms of the trial were used for data analysis. The duration of each run was approximately 5 minutes.

2.2.3. MRI data acquisition and analysis

Functional and structural images were acquired by a Philips Ingenia 3.0T CX MRI scanner with a SENSE 32-channel head-coil, located at the Department of Radiology of the University Hospital in Leuven, Belgium. To minimize head motion, wash cloths were used to stabilize the children's heads. For the fMRI data, 52 slices were recorded in an ascending order, using a T2*-sequence (2.19 × 2.19 × 2.2 mm voxel size, 2.2 mm slice thickness, 0.3 mm interslice distance, 96 × 95 acquisition matrix, 90º flip angle) and covered the whole brain (field of view: 210 × 210 × 130 mm). Each run consisted of 94 measurements (TR = 3000 ms, TE = 29.8 ms). Anatomical images were acquired with a T1 weighted sequence (0.98 × 0.98 × 1.2 mm voxel size, 256 × 256 acquisition matrix, 8º flip angle, TE 4.6 ms, 250 × 250 × 218 mm field of view).

All preprocessing was conducted with the Statistical Parametric Mapping software package for Matlab (SPM12, Wellcome Department of Cognitive Neurology, London). Preprocessing included correcting the functional images for slice timing differences, motion correction by realignment to the first functional image, coregistration (alignment to the respective high-resolution anatomical image), normalization to the standard Montreal Neurological 152-brain average template, and spatial smoothing with a 10 mm FWHM Gaussian smoothing kernel.

We only included the correctly answered items into our general linear model. In runs that showed excessive motion (i.e., if the movement from one image to the next was greater than the voxel size of 2.2 mm), only the items before the time point of excessive movement were included. This was only the case if at least one item per condition remained in that run; if this was not the case, the entire run was discarded. Data of participants with less than two completely usable runs were discarded entirely. Following these motion criteria, and taking into account the technical acquisition problems for the data collection of one participant, we completely discarded the data of six participants. Of the remaining 20 participants, five runs (i.e., 6.25%) were discarded, and five runs were only partially added to the model.

A general linear model, modelling only the correctly solved items, was calculated per participant. The motion realignment parameters were included as regressors of no interest to control for variation as a result of movement artifacts. A whole-brain full factorial 2 × 2 ANOVA was performed on the imaging data, with strategy (retrieval vs. procedure) and operation (subtraction vs. multiplication) as within-subject factors. To provide more information on the direction of any found main effects, t-contrasts were calculated between all conditions. All whole-brain activation maps were corrected for multiple comparisons through a family wise error (FWE) correction with a $p < .05$ threshold.

2.3. Results

2.3.1. Behavioral results

Results of the strategy assessment task are displayed in Table 2.1. Overall accuracy on this task was very high. Furthermore, the verbal reports indicated that children used a retrieval strategy on most of the items designed to elicit fact retrieval; the same was true for procedural strategy use. It is important to note that the vast majority of the reported procedural items were solved through the decomposition of operands strategy. Other procedural strategies were rare; repeated addition, for example, was only reported in 0.34% of all trials, and a counting strategy was never reported. The consistency of retrieving

Table 2.1

Performance on strategy assessment task

Condition	Accuracy (% correct)		Frequency (%)	
	M	*SD*	*M*	*SD*
Retrieval				
Multiplication	99.33	1.37	90.83	11.49
Subtraction	100	0	100	0
Procedure				
Multiplication	98.50	3.28	99.75	1.12
Subtraction	99.50	2.24	91.50	12.68

Note: The differentiation between retrieval and procedure is based on the self-reports of the participants.

single digit items and the absence of counting for procedural items was expected as, in the Flemish educational system, there is a high emphasis on either fact retrieval or on using the decomposition of operands strategy, while counting, from a very early point on in first grade, is discouraged or even prohibited (De Smedt, 2016, for a discussion). This exclusive focus on the decomposition strategy in teaching arithmetic also resulted in large similarities in the way procedural subtraction and multiplication items were solved. Children always reported a similar decomposition strategy for both operations: e.g., 25 – 8 = 25 – 5 – 3 = 17 for subtraction, and 4 × 13 = (4 × 10) + (4 × 3) = 40 + 12 = 52 for multiplication. As this decomposition strategy was used throughout almost all procedural trials, the current study can only discuss this particular strategy and cannot make any claims regarding other procedural strategies, such as counting or repeated addition.

Behavioral data on the arithmetic task in the scanner are displayed in Figures 2.3 and 2.4. We performed a 2 × 2 repeated measures ANOVA with strategy (retrieval vs. procedure; i.e., retrieval vs. decomposition) and operation (multiplication vs. subtraction) as within-subject factors for both accuracy and reaction time. Note that the reaction times were measured starting from the onset of the presentation of the response alternatives, and of not the problem itself.

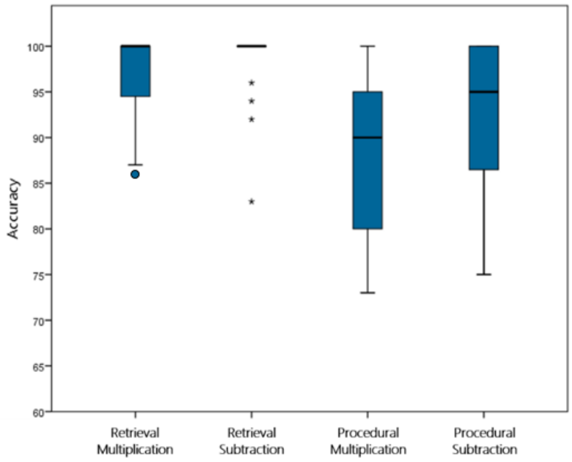

Figure 2.3. Box plots displaying the accuracy per category on the arithmetic task during fMRI.

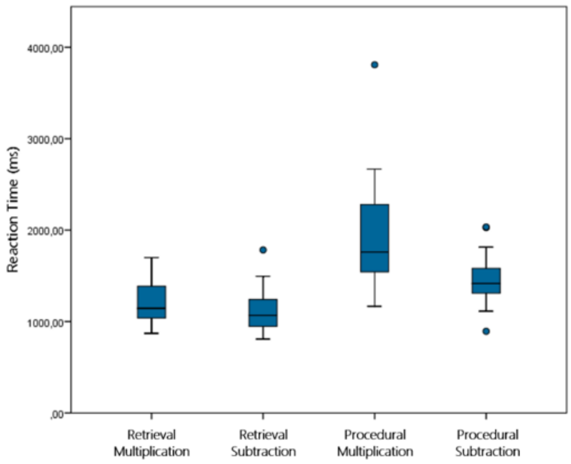

Figure 2.4. Box plots displaying the reaction time per category on the arithmetic task during fMRI.

The analyses of accuracy indicated main effects of both strategy ($F(1,19) = 24.59$, $p < .001$, $\eta p^2 = .56$) and operation ($F(1,19) = 5.24$, $p = .034$, $\eta p^2 = .22$), showing more accurate performance for retrieval than for the decomposition strategy, and for subtraction than for multiplication. There was no strategy × operation interaction ($F(1,19) = 0.68$, $p = .42$, $\eta p^2 = .04$).

With regard to reaction time, similar results were observed, as main effects for strategy ($F(1,19) = 49.07$, $p < .001$, $\eta p^2 = .72$) and operation ($F(1,19) = 16.94$, $p = .001$, $\eta p^2 = .47$) were found, indicating faster responses for retrieval compared to the decomposition strategy, and for subtraction compared to multiplication. A significant strategy × operation interaction was also found ($F(1,19) = 7.69$, $p = .012$, $\eta p^2 = .29$), revealing a larger difference between the use of a decomposition strategy and retrieval in multiplication compared to subtraction.

2.3.2. Imaging results

Neural differences between both strategies were found, as our whole-brain analysis revealed a main effect of strategy on brain activation (an overview of all significantly activated clusters can be found in Table 2.2; a visualization of this main effect is displayed in Figure 2.5). Retrieval strategy use was associated with stronger activation in the bilateral angular and supramarginal gyri, the left middle to inferior temporal gyrus, the right superior temporal gyrus, and the bilateral middle orbital and superior medial frontal gyrus. The stronger activation found for the retrieval vs. decomposition contrast, however, does not reflect an actual increase in activation, but a lesser amount of deactivation compared to baseline. This was determined by extracting the beta values of each activation cluster for each contrast separately, for which a negative value would imply lower activation in comparison to baseline. The use of a decomposition strategy more strongly activated a large bilateral, mainly fronto-parietal, network, which includes the inferior and superior parietal lobes, including the intraparietal sulci, inferior to superior frontal gyri, but also bilateral areas in the occipital lobe, and insular cortex.

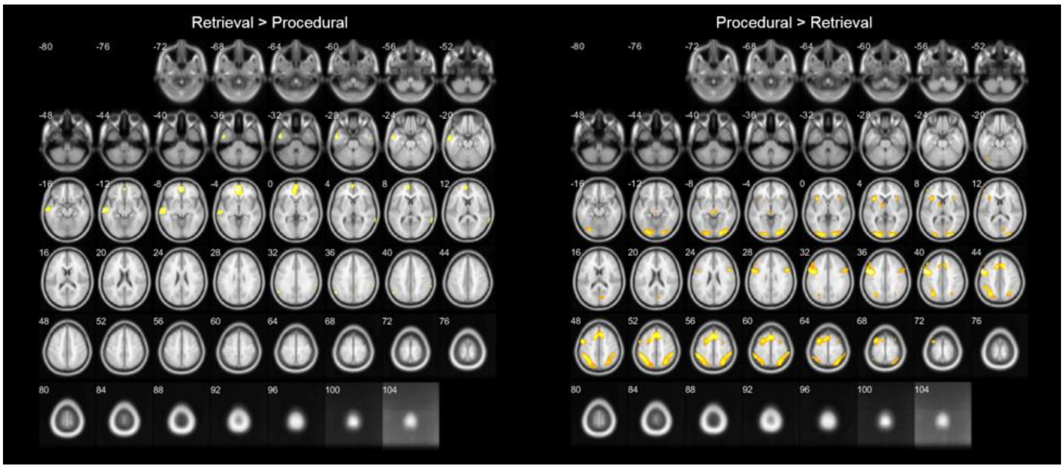

Figure 2.5. Transverse slices of differences in brain activation between self-reported retrieval and procedural strategy use ($p < .05$, FWE corrected).

Table 2.2

Regions, peak coordinates, cluster sizes (k), and t-values of the significantly activated clusters (p < .05, FWE corrected) for the main effect of strategy

Cluster	**Peak coordinates**			**k**	***t***
	x	**y**	**z**		
Retrieval > Procedural					
Bil frontal pole (middle orbital / superior medial frontal G / ACC)	-6	50	-4	943	7.40
L middle / inferior middle temporal G	-52	-4	-32	894	6.41
L inferior parietal lobe (angular / supramarginal G)	-54	-58	36	104	5.86
R superior temporal G to inferior parietal lobe (angular G)	70	-48	8	76	6.33
R inferior parietal lobe (supramarginal G)	66	-44	42	37	5.06
Procedural > Retrieval					
Bil superior / medial frontal G	-12	16	50	2687	9.09
L superior to inferior parietal lobe (intraparietal sulcus)	-44	-46	52	2273	8.93
R superior to inferior parietal lobe (intraparietal sulcus)	32	-62	54	1917	8.60
L inferior frontal G (Broca's region) / precentral G	-50	6	38	1802	10.22
R occipital lobe (V1 / occipital G)	24	-90	-6	1236	8.60
L occipital lobe (V1 / occipital G)	-22	-96	4	1234	8.60
R inferior frontal G (Broca's region)	52	10	26	433	7.69
L insula	-34	20	4	294	7.24
R insula	36	20	8	161	6.34
R middle to superior frontal G	32	6	66	111	5.31
L inferior frontal G	-50	46	10	69	5.92

Note: Only clusters of 20 voxels or more are reported.

Abbreviations: L = left hemisphere; R = right hemisphere; Bil = bilateral; G = gyrus; ACC = anterior cingulate cortex; V1 = primary visual cortex.

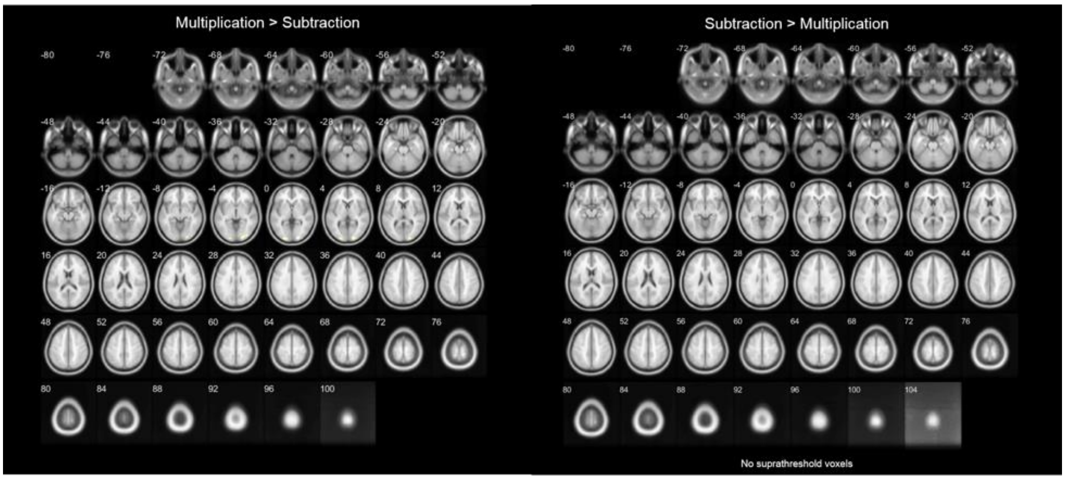

Figure 2.6. Transverse slices of differences in brain activation between multiplication and subtraction ($p < .05$, FWE corrected).

Table 2.3

Regions, peak coordinates, cluster sizes (k), and t-values of the significantly activated clusters (p < .05, FWE corrected) from the main effect of operation.

Cluster	**Peak coordinates**			**k**	***t***
	x	**y**	**z**		
Multiplication > Subtraction					
R occipital lobe (V1 / inferior occipital G / calcarine G)	18	-96	4	243	5.80
L occipital lobe (V1 / middle occipital G)	-22	-96	2	128	5.78
Subtraction > Multiplication					
/	/	/	/	/	/

Note: Only clusters of 20 voxels or more are reported.

Abbreviations: L = left hemisphere; R = right hemisphere; G = gyrus; V1 = primary visual cortex.

Turning to the main effect of operation (Table 2.3; Figure 2.6), we have only observed differences in the bilateral primary visual cortex, with higher activity during multiplication than for subtraction. The strategy × operation interaction revealed no significantly activated clusters, indicating that the effect of strategy was not different for both operations.

2.3.3. Additional control analyses

To verify that the observed activation differences were not merely explained by task difficulty effects, we performed an additional control analysis (see Table 2.4 and Figure 2.7), in which we compared problems that had a similar level of difficulty, but varied in terms of their strategies. This was done by splitting the trials from both operations into easy and hard items, based on the size of both operands, and comparing the hard retrieval and easy decomposition items (i.e., items of comparable size and difficulty; e.g., large retrieval: 3 × 8 vs. small decomposition: 3 × 12). As these analyses were intended as control analyses, and as taking task difficulty into account led to less trials per contrast, and hence a decrease in statistical power, these analyses were performed without a correction for multiple comparisons ($p < .001$).

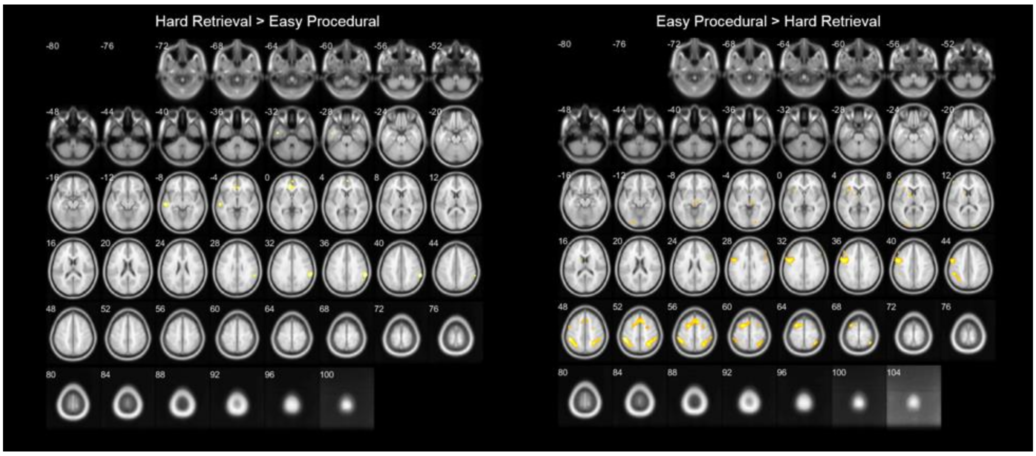

Figure 2.7. Transverse slices of differences in brain activation between hard retrieval and easy procedural items ($p < .001$, uncorrected).

Table 2.4

Regions, peak coordinates, cluster sizes (k), and *t*-values of the significantly activated clusters ($p < .001$, uncorrected) from the control analysis comparing hard retrieval items to easy procedural items.

Cluster	**Peak coordinates**			**k**	***t***
	x	**y**	**z**		
Large retrieval > Small procedural					
R supramarginal G	64	-40	38	295	5.19
Bil ACC	-6	34	0	198	5.02
L middle temporal G	-56	-22	-8	161	4.94
L inferior temporal G	-48	-4	-30	36	4.33
Bil frontal pole (superior medial frontal G)	4	56	2	28	4.07
R inferior parietal lobe	54	-56	38	21	3.81
Small procedural > Large retrieval					
L superior frontal to medial frontal G	-20	4	60	1321	6.31
L precentral G	-48	4	34	1106	6.26
L superior to inferior parietal lobe (intraparietal sulcus)	-48	-40	50	1023	7.20
R superior to inferior parietal lobe (intraparietal sulcus)	28	-56	50	1010	6.03
R middle frontal G / precentral G	24	-4	48	127	4.52
L inferior frontal G	-52	46	12	103	5.33
L middle frontal G	-50	32	40	75	4.08
L insula	-34	18	4	74	4.47
R inferior frontal G	52	10	26	70	4.44
L occipital lobe (middle occipital G)	-28	-94	6	47	4.41
L occipital lobe (inferior occipital G / lingual G)	-22	-90	-12	37	4.22
R cuneus	20	-98	10	36	5.01
R inferior frontal G	58	30	30	30	4.36
R occipital lobe (lingual gyrus / V1)	22	-88	-6	29	4.23

Note: Only clusters of 20 voxels or more are reported.

Abbreviations: L = left hemisphere; R = right hemisphere; G = gyrus; ACC = anterior cingulate cortex; V1 = primary visual cortex.

First, for each subject, we compared the difference in accuracy and reaction time between retrieval and decomposition trials vs. the difference in accuracy and reaction time between hard retrieval and easy decomposition items. These differences were significantly smaller in the latter (accuracy: $t(19) = 3.474$, $p = .003$; reaction time: $t(19) = -3.781$, $p = 0.001$), suggesting that the effect of task difficulty is significantly smaller in the hard retrieval vs. small decomposition contrast compared to the overall retrieval vs. decomposition contrast. If the outcome of the retrieval vs. decomposition contrast merely reflected an effect of task difficulty on brain activity, we expected that the contrast with a significantly reduced task difficulty effect (i.e., between hard retrieval vs. easy decomposition items) would show no differences or different activation clusters in comparison to the more general retrieval- decomposition contrasts. This is, however, not what we observed. The neural activation differences between hard retrieval and easy decomposition were very similar to those of the general retrieval-decomposition contrasts (Tables 2.2 and 2.4, Figures 2.5 and 2.7). More specifically, when contrasting the hard retrieval with the easy decomposition items, increased activation was found in the right supramarginal gyrus, left middle to inferior temporal gyrus and the bilateral frontal pole. The easy decomposition vs. hard retrieval contrast, on the other hand, revealed increased activation in a wide fronto-parietal network, including the bilateral inferior to superior parietal lobes and inferior frontal gyri, but also bilateral occipital areas. The fact that these networks were very similar to those of the general retrieval-decomposition contrasts, further supports the idea that the abovementioned results are not merely due to task difficulty effects.

2.4. Discussion

To date it is still unclear how children's arithmetic brain network is modulated by the strategies used to solve different problems, as earlier studies were only able to make implicit assumptions on strategy use, based on, for example, operation (Prado et al., 2014; De Smedt et al., 2011). This approach has been criticized for many years in developmental behavioral studies (e.g., Siegler, 1987; Siegler, 1996), as not all problems of a particular operation are solved with the same strategy. This is especially problematic in the context of developmental research, as the strategies that children use to solve particular types of problems change over time (i.e., with education and practice). Trial-by-trial self-reports offer a valid and reliable way of capturing these differences in (children's) strategy use (Siegler & Stern, 1998), and adult brain imaging studies have already successfully applied this approach to investigate brain activity during different solution strategies (Grabner et al., 2009; Tschentscher & Hauk, 2014). To the best of our knowledge, no such approach had been used in children. Against the background of the previously reported neural activation differences during arithmetic between adults and children, the present study set out to investigate the neural differences in children's arithmetic strategy use.

The present study was thus the first to explicitly investigate the neural activation underlying different arithmetic strategies during subtraction and multiplication in children of a narrow age range (i.e., 4th grade). Our data show a clear effect of strategy on brain activity, which is similar in both subtraction and multiplication. These data suggest that previously found effects of operation (subtraction vs. multiplication) on brain activity reflect differences in strategy use, rather than differences in operations.

During retrieval use, we observed increased activation in the supramarginal and angular gyri, middle temporal gyri and frontal pole. These results concur with previous adult studies that used a similar methodology (Grabner et al., 2009; Tschentscher & Hauk, 2014), as temporo-parietal regions (more specifically the angular gyri) have been shown to be related to fact retrieval. Similar to the data of Tschentscher & Hauk (2014), we found bilateral activation in these areas, which is in contrast to earlier observations by Grabner et al. (2009), in which the activity in these areas was left-lateralized.

The current results further extend previous fMRI studies in children, which could only make implicit assumptions on children's strategy use (Cho et al., 2011; De Smedt et al., 2011; Peters et al., 2016; Prado et al., 2014). For example, we have found stronger activation in the middle temporal gyrus during retrieval, which echoes earlier findings by Prado et al. (2014), who showed that activity in this area increases with age during multiplication, potentially due to increased use of retrieval strategies. The current study goes beyond the findings of Prado et al. (2014), as we directly correlated brain activity with participant's strategy use and observed the same effect of strategy use for both multiplication and subtraction. The current data consequently confirm that the earlier reported operation differences reflect the increased use of fact retrieval strategies, and that this activity is independent of the operation that is being performed. Our results also coincide with an fMRI study in children by Peters et al. (2016), who manipulated the presentation format to study differences between retrieval and procedural strategies. Specifically, these authors investigated brain activation during a subtraction task in symbolic (i.e., Arabic digits or number words) and non-symbolic (i.e., arrays of dots) formats. The symbolic formats, assumed to be solved by fact retrieval, showed increased activity in the bilateral angular and supramarginal gyri, as was also found in the current study.

In contrast to our expectations and to previous developmental fMRI studies (i.e., De Smedt et al., 2011; Qin et al., 2014), our analyses did not reveal specific increases in the hippocampus during fact retrieval. However, when applying a less stringent correction for multiple comparisons (i.e., a False Discovery Rate (FDR) correction with a $p < .05$ threshold), we observed an activated cluster in the left hippocampus (x = -24, y = -16 z = -18, k = 208, t = 4.18), which is in line with the earlier developmental reports. This observation is also in line with a recent adult study on the interference effect in multiplication problems by De Visscher et al. (2015) that also found greater activation in the left hippocampus for fact retrieval, implying that the hippocampus might not only play a role in fact retrieval during the early stages of arithmetic development.

The current study also tried to examine neural activation during procedural strategy use, but as the children in our sample almost exclusively implemented a decomposition of operands strategy for all procedural items, only claims can be made on this particular strategy. Furthermore, this decomposition of operands strategy cannot be seen as a procedural strategy in the same way as, for example, counting can, for it – to some extent – also involves fact retrieval. However, it is crucial to emphasize that this decomposition of operands, and consequently the small degree of fact retrieval involved, is not random, as is the case during mere fact retrieval. The decomposition strategy follows a fixed sequence (i.e., in subtraction it starts with subtracting to tens, and in multiplication it includes a multiplication by ten), and, to find the correct solution, both solutions to the newly formed items still need to be subtracted from or added to one another, thus clearly making it a multistep procedural strategy. For the decomposition of operands strategy, we observed increased activation in a fronto-parietal network, which includes the bilateral inferior to superior parietal lobes (including the intraparietal sulci), and the inferior to superior frontal gyri, but also bilateral areas in the occipital lobe, and insular cortex. These results converge with previous studies in adults (Grabner et al., 2009; Tschentscher & Hauk, 2014), although the adult data were more left-lateralized, in contrast to the bilateral network we have found.

These results on the decomposition strategy concur with studies in children who manipulated operation (De Smedt et al., 2011; Prado et al., 2014) or presentation format (Peters et al., 2016) to investigate strategy use. The effect of decomposition on brain activity was also the same across both operations, again indicating that it is the strategy and not the operation that elicits changes in brain activity. The observed increases in fronto-parietal activation for the decomposition strategy might point to an increasing demand on working memory and attentional resources, as reflected by increases in frontal activation, especially in the insula and inferior to middle frontal gyrus (Duncan & Owen, 2000; Menon, 2015), as well as a larger involvement of quantity-based processes, reflected by the increased parietal activation, specifically in the intraparietal sulci (De Smedt et al., 2011). The current study, however, cannot disentangle these different processes; future studies should, therefore, adopt a carefully selected localizer approach in order to test this.

Besides the decomposition of operands strategy, other types of procedural strategies exist, including repeated addition or counting. The choice of the used (procedural) strategy, however, is highly dependent of the math curriculum under study. The participants in the current study all came from schools with a high emphasis on fact retrieval for single digit problems and on the use of the decomposition of operands strategy for larger problems (coupled with a limited attention or even prohibition of counting), as is in accordance with the mandatory guidelines of the Flemish education system. Consequently, and as expected, strategies other than the decomposition of one of the operands (e.g., counting or repeated addition) were hardly used. It is noteworthy that different types of procedural strategies (e.g., counting vs. decomposition) might elicit alternate neural activation patters, but unfortunately this could not be tested in the current sample as children were very homogenous in their choice of procedural strategies (i.e., decomposition). In all, it is important to acknowledge potential educational differences between countries. A fact that is often overlooked in educational neuroscience studies that deal with culturally transmitted skills, is that the way these skills are taught in school, will affect children's performance on an educationally relevant task and will consequently affect brain activity (De Smedt & Grabner, 2015, for a discussion). Future brain imaging studies should therefore take this educational context into account and might consider to investigate cross-curricular differences.

Next, we should also consider to what extend the current differences are driven by task difficulty effects. To investigate this possibility, we ran a control analysis in which we compared problems with a similar level of difficulty, but for which different strategies were needed. Consequently, we split all items into easy and hard items, based on the size of both operands and contrasted the hard retrieval and easy decomposition items, which showed a significantly reduced task difficulty effect (for both accuracy and reaction time) in comparison to the general retrieval-decomposition contrasts. The results of this control analysis showed similar neural activation networks between the hard retrieval vs. easy decomposition contrasts and the general retrieval vs. decomposition contrasts. For the hard retrieval vs. easy decomposition contrast, we observed increased activation in the right supramarginal gyrus, left middle to inferior temporal gyrus and the bilateral frontal pole , which is very similar to the retrieval vs. procedural contrast. The easy decomposition vs. hard retrieval contrast, on the other hand, displayed increased activation in a fronto-parietal network, similar to the procedural vs. retrieval contrast. The similarities between these findings suggest that the differences between retrieval and decomposition trials are unlikely to be merely driven by task difficulty effects.

Further, it needs to be noted that the stronger activation found for the retrieval condition reflected less deactivation and not increased activation in comparison to baseline (see also, De Smedt et al., 2011; Peters et al., 2016). The regions showing this deactivation are to some extent part of the default mode network (Raichle et al., 2001; Supekar et al., 2010), which decreases in activation as the cognitive demand of a task increases and vice versa. However, the less deactivated regions that were found for the retrieval condition do not fully coincide with the areas active in this resting brain state, as, for example, no activation in the retrosplenial cortex was found (Vann, Aggleton, & Maguire, 2009). The regions found to be activated more strongly for the decomposition condition, on the other hand, seem to coincide with those of the so-called multiple-demand network (Fedorenko, Duncan, & Kanwisher, 2013), which, in adults, shows increases of activation for any kind of cognitive demand, independent of the content of the task. These differences between fact retrieval and decomposition, and differences in regions that seem to be part of the default mode and multiple-demand network respectively, might be explained by the inevitable association between strategy use and the task load of the items at hand. As evidenced by research of Siegler (1984, 1996) fact retrieval is an easy, accurate and fast strategy for solving arithmetic problems, which has a smaller cognitive demand than decomposition strategy use. This was also apparent in our data, as the retrieved items were solved more quickly and more accurately, while the decomposition items showed the opposite pattern.

Concurring with the adult data of Tschentscher & Hauk (2014), we did not find any differences in activation between operations. Although such operation differences have been previously observed in children (e.g., De Smedt et al., 2011; Prado et al., 2014), those studies did not directly assess strategy use. The current data show for the first time that such operation differences in children, just as in adults, are explained by the strategy that is used. In other words, it is the strategy and not the operation itself that determines brain activity. One small main effect of operation was observed, however, but only for regions in the primary visual cortex, that were increasingly activated for multiplication in comparison to subtraction. This is due to the inevitable, yet subtle, differences in visual presentation between the subtraction and multiplication items: The response alternatives for multiplication were unavoidably larger in subtraction items, hence more visual information was displayed during multiplication.

In contrast to the approach of the current study, which is based on the distinction between retrieval and procedural strategies, recent studies have suggested that, instead of retrieval, children may use automatized procedural strategies, which over time contrasts the dominant view of an evolution from counting to retrieval strategies, but implies a shift from slow to quick counting procedures (Thevenot et al., 2016). Although this notion cannot strictly be excluded, the current data still point to alternate neural activation between both strategy conditions, and more importantly, provide evidence against the idea that the arithmetic brain network is modulated by the operation of items (Tschentscher & Hauk, 2014). Furthermore, the effects found by Thevenot et al. (2016) were only described in small addition problems, and not in subtraction or multiplication. As these small addition problems are consistently solved faster than, for example, subtraction problems, it is uncertain if these automatized counting procedures would also occur in subtraction, let alone in multiplication.

Future developmental imaging studies on mathematical cognition should thus avoid thinking in terms of operations, but instead take strategy use into account. Our results, coupled with those of Tschentscher and Hauk (2014) clearly indicate that arithmetic strategy rather than operation modulates brain activity. Furthermore, behavioral research has implicated large developmental aspects in strategy use, especially in the frequency and efficiency of those strategies (Imbo & Vandierendonck, 2008; Siegler, 1996; Siegler et al., 1996; Vanbinst, Ceulemans, Ghesquière, & De Smedt, 2015a). Consequently, future brain imaging studies should longitudinally study how the neural networks found in our group of 4th graders for both retrieval and procedural strategies develop, from an early-arithmetic stage (e.g., 1st or 2nd graders) to a more advanced arithmetic stage. Moreover, as difficulties in arithmetic strategy use are considered the hallmark of children with dyscalculia, who experience persistent deficits in acquiring basic mathematical competencies (American Psychiatric Association, 2013), and as fact retrieval deficits have also been observed in children with dyslexia (Evans, Flowers, Napoliello, Olulade, & Eden, 2014), future brain imaging studies on strategy use in these atypical populations are also needed.

One limitation of the current study lies in the differences in format of the arithmetic task during strategy assessment outside the scanner and during MRI-acquisition in the scanner (i.e., a production task with auditory input and verbal output in the strategy assessment session, and a delayed verification task with visual input and manual output during MRI-acquisition). This was done to find a balance between ecologically valid strategy assessment (allowing for precise measurement of strategies) and the practical limitations of the scanning environment for the scanning task. In view of the high consistency in the implementation of strategies during the behavioral task, coupled with the use of delayed verification, which limits the possibility of parity checking, five-checking and other estimation strategies, we contend that children employed the same strategy in the scanner as in the behavioral session. The validity of this verbal protocol is also supported further by the fact that we found a significant main effect for both accuracy and reaction time for the in-scanner task, pointing to more accurate and faster responses during the retrieval condition, which concurs with previous behavioral findings (e.g., Siegler, 1984; Siegler, 1996). It is also supported by the fact that significantly larger hard-easy performance differences can be found for the decomposition strategy in comparison to fact retrieval, for both accuracy and reaction time (accuracy: $t(19) = 2.598$, $p = .018$; reaction time: $t(19) = -3.748$, $p = .001$).

We would like to highlight again that the acquisition and use of arithmetic strategies does not occur in isolation, but depends on the extent to which math curricula emphasize the importance of fact retrieval and automatization, and on the particular strategies these curricula focus on; behavioral studies have clearly shown cross-cultural differences in retrieval use in adults depending on the emphasis of the math curriculum on fact retrieval and automatization (e.g., Campbell & Xue, 2001). The children of the current study came from Flemish elementary schools with a curriculum that puts a high emphasis on automatization processes and fact retrieval on the one hand, and on the decomposition of operands as an effective procedural strategy on the other hand, leading to a limited generalization of the current findings to other cultures. Future studies might therefore explore how, for example, differences in the emphasis on fact retrieval or certain procedural strategies in math curricula correlate with strategy-related brain activity. Such studies have the potential to provide a fruitful contribution to the emerging field of educational neuroscience.

Finally, the current study focused on fourth graders, which were capable of both retrieving the answers to multiplication items and solving more difficult items procedurally. As mentioned, we have chosen to focus on a narrow age range, as merging data across wide age ranges, even though statistically controlled for, might lead to misleading conclusions. Accordingly, we would like to emphasize the need for similar studies in children of different ages (e.g., sixth graders or children in secondary school), and for studies with a longitudinal follow-up throughout development, as such studies are destined to provide meaningful insights in the development of these strategies.

CHAPTER 3

The Neural Substrates of the Problem Size and Interference Effect in Children's Multiplication: An fMRI Study

Published as

Polspoel, B., De Visscher, A., Vandermosten, M., Vogel, S. E., Grabner, R. H., & De Smedt, B. (2019). The neural substrates of the problem size and interference effect in children's multiplication: An fMRI study. *Brain Research*, *1714*, 147-157.

Abstract

Within children's multiplication fact retrieval, performance can be influenced by various effects, such as the well-known problem size effect (i.e., smaller problems are solved faster and more accurately) and the more recent interference effect (i.e., the quality of memory representations of problems depends on previously learned problems; the more similar a problem is to a previously learned one, the more proactive interference impacts on storing in long-term-memory). This interference effect has been observed in behavioral studies, and determines a substantial part of performance beyond problem size. Unlike the problem size effect, the neural basis of the interference effect in children has not been studied. To better understand the underpinning mechanisms behind children's arithmetic fact retrieval, we aimed to investigate the neural basis of both effects in typically developing children. Twenty-four healthy 9- to 10-year-olds took part in a behavioral and fMRI scanning session, during which multiplication items had to be solved. Data were analyzed by manipulating problem size and interference level in a 2 × 2 factorial design. Concurring with previous studies, our results reveal clear behavioral effects of problem size and interference, with larger and high interfering items being solved significantly slower. At the neural level, a clear problem size effect was observed in a fronto-parietal and temporal network. The interference effect, however, was not detected; no clear neural distinctions were observed between low and high interfering items.

3.1. Introduction

Over the past few years, an increase in developmental neuroscientific research on numerical cognition and arithmetic, both functionally and structurally, has occurred, aimed at identifying which brain regions are involved in children's arithmetic (Peters & De Smedt, 2018). Accordingly, accumulating evidence suggests that children's arithmetic brain network involves a large set of interconnected areas, including frontal (i.e., both ventro- and dorsolateral prefrontal cortex), parietal (i.e., intraparietal sulcus, angular gyrus, and supramarginal gyrus), occipito-temporal, and medial temporal (i.e., hippocampus) areas. Though this network shows some similarities to the network observed in adults (Menon, 2015), children's arithmetic network is different, as, in children, less involvement of the superior parietal lobe is observed, along with larger involvement of the hippocampus, and of occipito-temporal, and prefrontal medial regions (Menon, 2015; Peters & De Smedt, 2018). This arithmetic network is modulated by strategy use, as different neural regions get activated when children solve problems through a fact retrieval strategy instead of through procedural manipulations (Polspoel, Peters, Vandermosten, & De Smedt, 2017). As such, it is important to note that within arithmetic fact retrieval, performance (both accuracy and reaction time) can be influenced by various effects. Two of these effects regard the well-studied problem size effect (i.e., better performance is often observed for small in comparison to large problems; e.g., Berteletti et al., 2014; De Smedt et al., 2011; Prado et al., 2013; Prado et al., 2014), and the more recently found interference effect (i.e., the quality of memory representations of multiplication problems depends on previously learned problems; the more similar a problem is to a previously learned one, the more proactive interference will impact on storing in long-term-memory, leading to poorer performance; De Visscher et al., 2015; De Visscher et al., 2018). This interference effect has been observed in children's behavioral studies, both in typical and atypical development, and determines a substantial part of performance over and above the problem size effect. Unlike the problem size effect, the interference effect in children's multiplication has never been investigated at the neural level. To achieve a more detailed understanding of the exact function of the regions in this arithmetic fact retrieval network, the current study aimed to investigate the neural basis of both problem size and interference effects in typically developing children.

A clearly studied aspect of arithmetic fact retrieval thus lies in the notion that better performance (i.e., more accurate and faster performance), is often observed for small problems in comparison to large problems (De Brauwer et al., 2006). This problem size effect is most often explained by the idea that smaller problems are more likely to be retrieved in comparison to larger problems (Zbrodoff & Logan, 2005). However, the problem size effect still exists within retrieved items only. Various other explanations for these performance differences exist as well, such as the frequency theory (i.e., smaller problems are solved faster and more accurately because they appear more frequently; Ashcraft & Christy, 1995), the distribution of associations model (i.e., each problem is associated with all previously computed answers, both correct and wrong, making the amount of errors increase as problem size increases; Siegler, 1988), or the network interfering theory (i.e., magnitude representations follow a psychophysical scale that is more compressed as magnitude increases, making the representations of large answers more similar to one another than representations of small answers; Campbell, 1995).

At the neural level, studies show that the activation of the arithmetic brain network is modulated by the problem size effect (e.g., Stanescu–Cosson et al., 2000; Molko et al., 2003). As such, the angular gyri show increased activation for small retrieved items, in both children (Polspoel et al., 2017) and adults (Grabner et al., 2009; Stanescu-Cosson et al., 2000; Tschentscher & Hauk, 2014). The associated activation for larger problems, on the other hand, lies more in fronto-parietal and temporal brain regions, including the superior parietal lobe and intraparietal sulcus, inferior frontal gyrus, and fusiform gyrus (Grabner et al., 2007; Grabner et al., 2009; Polspoel et al., 2017; Tschentscher & Hauk, 2014). Furthermore, in children, the hippocampus has been suggested to be involved in the learning phase of arithmetic facts, as it also shows greater activation for small problems in comparison to large problems in 10-12-year-old children (De Smedt et al., 2011).

The interference effect, on the other hand, constitutes a more recent hypothesis on performance differences within multiplication fact retrieval (De Visscher & Noël, 2013; De Visscher & Noël, 2014a; De Visscher & Noël, 2014b). The main idea behind the hypothesis lies in the notion that arithmetic facts are learned through associations of operands and answers which all use the same 10 elements (i.e., the digits 0 to 9), and are acquired in a specific order. Items that have been memorized, can then interfere with the memorization process of new items, if the items share element combinations and are thus similar. This interference theory, based on the feature overlap theory (Nairne, 1990), thus implies that the quality of memory representations of multiplication problems depends on previously learned problems, and that, the more similar a problem is to a previously learned one (i.e., through similar elements used in a new combination), the more interference will occur during the storage phase, and hence will reduce the probability of retrieval (De Visscher & Noël, 2014b). Based on this theoretical framework, De Visscher and Noël (2014b) calculated an interference parameter for all 36 different multiplication problems (considering that a unique representation underpins each problem and its commutative pair; e.g., 3 × 4 and 4 × 3) from table 2 up to table 9. This was done by calculating the frequency of associations of two digits in all problems (in both operands and product). The parameter thus represents the weight of proactive interference for each problem. For example, the first multiplication problems children learn are 2 × 2 = 4 and 2 × 3 = 6, both items without any interference. However, the next item children learn, 2 × 4 = 8, shares the digits 2 and 4 with 2 × 2 = 4, and is thus subject to interference in memory; it receives an interference level of 1. This, however, does not mean that problems encountered later are always more interfering, but means that the more similar a certain problem is to previously learned problems, the larger the impact of interference will be. For example, 7 × 7 = 49, has a lower interference weight than 4 × 8 = 32, even though it is encountered later in arithmetic development.

At a behavioral level, the effect of the sensitivity-to-interference in memory has been studied in a single case study (De Visscher and Noël, 2013) in which it was proposed that a deficit in arithmetic facts storage could stem from hypersensitivity-to-interference in memory. Through elaborate cognitive (e.g., intelligence, attention, memory, and reading tasks) and mathematical (e.g., dot enumeration, numerical stroop, and arithmetic facts tasks) assessment, the interference effect was first established in a certain profile of dyscalculia, with a restricted deficit in arithmetic facts knowledge. Furthermore, in a comparison of fourth graders with low arithmetic fluency (i.e., low arithmetic fact learners) to controls (De Visscher and Noël 2014a), group differences were also observed. Children with poor arithmetic fluency displayed hypersensitivity-to-interference in memory and showed worse performance (i.e., more incorrect responses) on a custom multiplication task (i.e., a computerized task including one of each possible commutative pair of operands from 2 to 9) in comparison to the control group.

The interference effect has not only been studied in people with math difficulties, but also in typically developing children and adults. In a comparative study of 3rd graders, 5th graders and adults, De Visscher & Noël (2014b) showed that the interference parameter explains a large part of the reaction time variance in all three age groups; the time needed to solve multiplication items increased, as the level of interference increased. The interference effect even determined a substantial part of participants' reaction times beyond the problem size effect, emphasizing that both effects constitute two separate characteristics of performance. Furthermore, individuals' sensitivity to the interference parameter (i.e., the slope of a regression with reaction time as the dependent variable and interference level of the items as the independent variable) substantially predicted multiplication performance in both children and adults.

Even though this interference effect has thus been clearly established in children at a behavioral level, determining a substantial part of performance beyond the problem size effect, and having possible detrimental consequences for the storing of arithmetic facts (De Visscher and Noël, 2014a; 2014b; De Visscher, Noël, & De Smedt, 2016), the neural substrates of the effect have not yet been studied in children. A few studies have examined the neural basis of the effect in adults (De Visscher et al., 2015; De Visscher et al., 2018). In the study by De Visscher et al. (2015), which contrasted the neurocognitive correlates of the problem size effect and the interference effect in multiplication fact retrieval in adults, the problem size effect was found to modulate the bilateral intra-parietal sulci. The left angular gyrus, on the other hand, was found to be specifically modulated by the interference effect, as higher activation was found for low interfering items in comparison to high interfering items. This result specifies the role of the angular gyrus within multiplication fact retrieval and suggests a sensitivity of the angular gyrus to the level of interference of retrieved multiplication problems, which might reflect an automated mapping between problem and solution, stored in long-term memory. Furthermore, various other brain regions (i.e., bilateral insula lobes, bilateral supplementary motor area, middle cingulate gyri, and bilateral inferior frontal gyri) were modulated by both problem size and interference effect (De Visscher et al., 2015).

A more recent study on the neural basis of the interference effect in adults (De Visscher et al., 2018) also observed increased activation for low interfering problems in the angular gyrus, albeit in the right hemisphere, instead of the left. The study also pointed towards increased activation for low interfering problems in another cluster in the right inferior parietal lobe and bilaterally in the middle orbital gyri. This study also went a step further than contrasting low and high interfering items, and took individual differences in arithmetic fluency into account, and observed a neural interference effect in the left inferior frontal gyrus, which showed a negative relation with individual differences in arithmetic fluency, indicating a higher neural interference effect for low performers in comparison to high performers.

As clear behavioral evidence for both the problem size and interference effect has thus been found in children, with the interference effect explaining a large part of reaction time variance for retrieved arithmetic items over and above the problem size effect in typically developing children (De Visscher and Noël 2014b), the current study set out to investigate the neural basis of both the problem size and the interference effects in children’s multiplication fact solving. In doing so, we will follow a similar fMRI design of De Visscher et al. (2015) and De Visscher et al. (2018) that was used in adults. More specifically, we will manipulate both problem size and interference level in a 2 × 2 full factorial, event-related design. This will allow us to investigate which brain regions become increasingly activated for small and large, and low and high interfering items, and whether the interference effect in children modulates the activation in different brain regions in comparison to the problem size effect. Based on previous studies on the problem size effect (e.g., De Smedt et al., 2011; Polspoel et al., 2017; Prado et al., 2013; Prado et al., 2014), we mainly expect to find increased activation in a fronto-parietal network (e.g., superior parietal lobe, inferior fontal gyrus) for large items in comparison to small items. For the interference effect, a modulation of children’s arithmetic fact retrieval network is mainly expected in the angular gyrus (as was previously observed in adults). As such, this study will help to understand the development of the arithmetic network and its time course, and potentially help chart the development of individual differences.

3.2. Methods

3.2.1. Participants

Participants for this study were 24 typically developing Flemish 9- to 10-year-olds; all participants were in the 4th grade of primary school. This age group was selected as 4th graders already have sufficient capabilities in arithmetic fact retrieval; the small age range was selected in order to minimize maturational confounds. None of the participants had a history of learning difficulties, or neurological or psychiatric disorders. Due to excessive in-scanner motion (for more details, see below), data of four children had to be discarded. Ultimately, data of 20 children (M = 9.79, SD = 0.29; range = 9.33-10.76; 7 males; 4 left-handed) were analyzed. All children were recruited via the elementary school they attended, in the surrounding area of the university, and were given a financial compensation for their participation. Written informed consent was obtained from a parent or legal guardian of each participating child. This study was approved by the Medical Ethical Committee of the University of Leuven (S59167).

3.2.2. Procedure

All participants took part in two test sessions (i.e., a behavioral and scanning session). Behavioral testing always preceded fMRI scanning by approximately 2 weeks (M = 17.2 days, SD = 5.8 days). Each behavioral session included standardized assessment of arithmetic, reading, and intellectual ability, as well as a custom multiplication task to measure each child's sensitivity-to-interference during multiplication. The second session included the actual acquisition of the fMRI data. This fMRI acquisition session also partly contained diffusion MRI data collection for another study, which is reported in Polspoel, Vandermosten, & De Smedt (2018).

3.2.2.1. Standardized assessment

In order to characterize the research sample, the standardized assessment session contained the evaluation of arithmetic, reading, and intellectual ability. Arithmetic fluency was assessed by the Tempo Test Arithmetic (TTA; de Vos, 1992), which is a standardized test of arithmetical fluency. Similar to the Math Fluency subtest of the Woodcock-Johnson III tests of Achievement (Woodcock et al., 2003), the TTA comprises five columns of arithmetic items, one column per operation and a mixed column, each increasing in difficulty. Participants get one minute per column to provide as many correct answers as possible. Reading ability was assessed by the One-Minute Test (OMT; Brus & Voeten, 1979) and the Klepel (van den Bos et al., 1994), which measure the reading of words and pseudowords, respectively. The OMT contains 116 words for which the participants get one minute to correctly read aloud as many words as possible; for the Klepel, the time limit is set to two minutes, and the children have to read aloud pseudowords. Finally, intellectual ability was assessed through the WISC-III-NL Block Design and Vocabulary subtests, as proxies of performance and verbal IQ, respectively (Wechsler, 2005). Standardized scores were calculated for all tasks.

3.2.2.2. Custom multiplication task

The custom multiplication task to measure sensitivity to problem size and interference, which was part of the behavioral session, consisted of the same 24 items used in De Visscher et al. (2015) and De Visscher et al. (2018), which were subsequently also used during fMRI scanning (see Table 3.1 for an overview of the items and their respective problem size and level of interference). Using the E-Prime 2

Table 3.1

Overview of the items used for the interference task by level of interference and problem size

	Small problem size			Large problem size		
	Item	IL	PS	Item	IL	PS
Low interfering	2 × 6 =	3	12	6 × 6 =	4	36
	5 × 5 =	3	25	6 × 5 =	6	30
	2 × 7 =	4	14	5 × 9 =	6	45
	4 × 4 =	5	16	9 × 9 =	6	81
	2 × 8 =	7	16	5 × 7 =	7	35
	9 × 2 =	7	18	7 × 7 =	7	49
High interfering	3 × 6 =	8	18	3 × 9 =	9	27
	5 × 4 =	8	20	9 × 4 =	9	36
	4 × 3 =	10	12	8 × 5 =	9	40
	4 × 6 =	12	24	7 × 8 =	9	56
	3 × 7 =	13	21	6 × 7 =	22	42
	8 × 3 =	13	24	4 × 8 =	25	32

Note: IL = Interference Level; PS = Problem Size. Problem size was defined as the product of both operands. Small items had a product of 25 or smaller. Interference level was calculated as in De Visscher & Noël (2014b), where items receive a level of interference each time they have two digits in either operands or solution in common with a previously learned item. Low interfering items had an interference level of 7 or lower. Similar to previous research (De Visscher et al., 2015; De Visscher et al., 2018), both interference level and problems size were considered categorical variables (i.e., low-high and small-large).

software (Psychology Software Tools), participants were presented with multiplication items in white on a black background. All 24 items were presented once. Each trial started with 2 seconds of fixation, followed by the item, which was displayed until an answer was provided. Participants were asked to verbally respond as quickly as possible. Reaction times were measured through voice key, after which the experimenter indicated if (a) the voice key registration was successful, (b) whether or not the provided answer was correct, and, after asking the child after each item, (c) whether or not the item was solved through fact retrieval or through any kind of procedural manipulation. After these checks, the next trial started.

Participants' individual sensitivity to problem size and to interference was calculated similarly as in De Visscher and Noël (2014b). A multiple regression was calculated for each participant, with reaction time as the dependent variable and the problem size and interference level of each item as the independent variables. The slopes of the independent variables were then used as a measure of individual differences in sensitivity to problem size or interference.

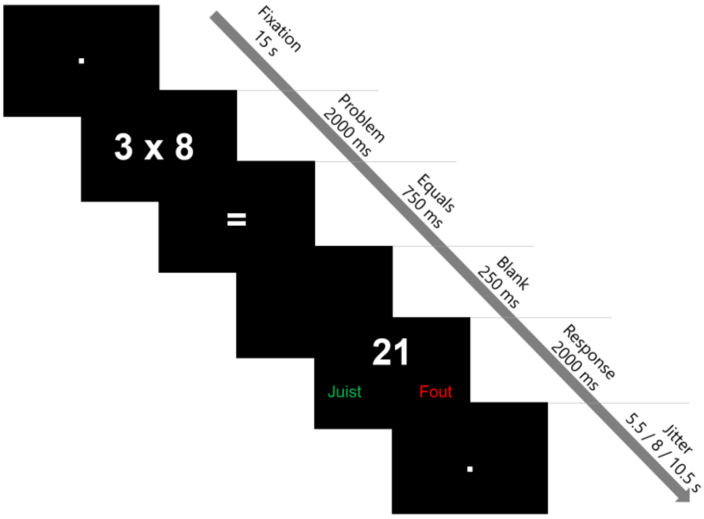

Figure 3.1. Example of a trial of the in-scanner multiplication task.

3.2.2.3. fMRI experimental design

In an event-related fMRI design, the children performed a multiplication verification task with the same items as the behavioral interference task (Figure 3.1). Stimuli were presented with E-prime 2.0 (Psychological Software Tools), via an NEC projector onto a screen behind the participants, made visible through a mirror attached to the head coil. All stimuli were presented in white (Arial, font size 60) on a black background. A trial started with the presentation of one of the 24 multiplication problems for 2s. After an equality sign and a black screen were presented (1s in total), a potential answer for the item was presented for 2s, during which the participants had to decide if the answer was correct or not, by pressing the left and right response button, respectively. All proposed answers that were incorrect, were always a table-related product, created by adding or subtracting the smallest operand to or from the correct answer, all of which was balanced within and across runs. To assure that the participants pressed the correct button for the response they wanted to give, two assisting labels were also displayed on the response screen (i.e., a green and red “juist” and ”fout” label – Dutch for “right” and ”wrong” – were displayed on the left and right bottom side of the screen, respectively). In between trials, a jittered inter-trial interval of 5.5, 8 or 10.5s (averaged at 8s) was randomly added to enable the deconvolution of the hemodynamic response functions. Only the first 5000ms of the correctly answered trials were used for data analysis. On average, only 5.2 out of 96 items per participant had to be discarded (range 0 – 9) due to incorrect answers.

3.2.3. MRI data acquisition and analysis

Functional and structural images were acquired by a Philips Ingenia 3.0T CX MRI scanner with a SENSE 32-channel head-coil, located at the Department of Radiology of the University Hospital in Leuven, Belgium. Wash cloths were used to stabilize the children’s heads and consequently minimize head motion. For the fMRI data, 52 slices were recorded in ascending order, using a T2*-sequence (2.19

× 2.19 × 2.2 mm voxel size, 2.2 mm slice thickness, 0.3 mm interslice distance, 96 × 95 acquisition matrix, 90º flip angle) and covered the whole brain (field of view: 210 × 210 × 130 mm). Each run consisted of 111 measurements (TR = 3000 ms, TE = 29.8 ms), and lasted approximately six minutes. Anatomical images were acquired with a T1 weighted sequence (0.98 × 0.98 × 1.2 mm voxel size, 256 × 256 acquisition matrix, 8º flip angle, TE 4.6 ms, 250 × 250 × 218 mm field of view), which lasted approximately eight minutes.

All preprocessing was conducted with the Statistical Parametric Mapping software package for Matlab (SPM12, Wellcome Department of Cognitive Neurology, London). Preprocessing included correcting the functional images for slice timing differences, motion correction by realignment to the first functional image, coregistration (alignment to the respective high-resolution anatomical image), normalization to the standard Montreal Neurological 152-brain average template, and spatial smoothing with a 10 mm FWHM Gaussian smoothing kernel.

Only correctly answered items were included for data analysis. If a participant displayed excessive motion during a run (i.e., movement greater than the voxel size of 2.2 mm from one image to the next), only the items before the time point of excessive movement were included. If a run did not contain at least one item for each condition (i.e., small or large problem size and low or high interference), the entire run was discarded (Vogel, Matejko, & Ansari, 2016). This, however, only occurred in 8.75% of the data. Data of participants with less than two completely usable runs were also entirely discarded. Following these motion criteria, the eventual analyses were performed on 20 participants.

To analyze the data, a general linear model, modelling only the correctly solved items, was calculated per participant. The motion realignment parameters were included as regressors of no interest to control for variation as a result of movement artifacts. A whole-brain full factorial 2 × 2 ANOVA was performed on the imaging data, with problem size (small vs. large) and interference level (low vs. high) as within-subject factors.

3.3. Results

3.3.1. Behavioral results

Descriptive statistics of the standardized assessment can be found in Table 3.2. These results show that the means of our sample were close to population averages and show proper variation. An important note is that, even though some of the minima were low (especially for arithmetic), none of the participating children had been diagnosed with any kind of learning or intellectual disability.

On the multiplication task outside of the scanner, participants had an average accuracy of 96.67% (SD = 4.19%; Range = 87.5 to 100%) and an average reaction time of 3295ms (SD = 1226ms; Range = 1500 to 5940ms). The average amount of items that were retrieved was 87% (SD = 16). To study the effects

Table 3.2

Descriptive statistics of the standardized assessment

	Mean	**SD**	**Minimum**	**Maximum**
Arithmetic – Total	4.70	3.15	1	10
Arithmetic – Multiplication	4.90	2.83	1	10
Reading	9.84	2.76	6	16.50
Block Design	11.35	3.44	6	19
Vocabulary	11.42	1.61	8	14

Note: The scores of arithmetic, reading, block design, and vocabulary are standardized scores. The scores on arithmetic are standardized as M = 5, SD = 2, with a maximum of 10. The scores on reading, block design, and vocabulary are standardized as M = 10, SD = 3, with a maximum of 19.

of problem size and interference level on both accuracy and reaction time, both frequentist and Bayesian statistics were performed at item and participant level. Bayesian statistics have the advantage of being able to quantify the evidence that the data provide for one hypothesis over another (Andraszewicz et al. 2015). Accordingly, Bayes factors (BF_{10}) of 1, 1-3, 3-10, 10-30, 30-100, or > 100 respectively point towards no, anecdotal, substantial, strong, very strong, or decisive evidence for the alternative hypothesis (Jeffreys, 1961).

To check for continuous associations between the variables, group level linear regressions were first calculated on both accuracy and reaction time. This was done by calculating the average reaction time for each item across participants, and by then calculating a regression model on reaction time with both interference level and problem size as dependent variables. Statistics for these analyses can be found in Table 3.3. For accuracy, neither problem size nor interference level were found to be significant predictors. For reaction time, on the other hand, problem size was not found to be a significant predictor, while interference level was.

Next, across participants, two-way ANOVAs were performed with problem size and interference level as independent variables, and accuracy and reaction time as dependent variables, to check for general group differences between the conditions. For accuracy, no significant main or interaction effects were found (interaction effect: $F(1, 76) = 3.021$, $p = .086$, $BF_{10} = 0.995$; main effect of problem size: $F(1, 76) = 0.062$, $p = .805$, $BF_{10} = 0.239$; main effect of interference: $F(1, 76) = 0.551$, $p = .460$, $BF_{10} = 0.294$). For reaction time, on the other hand, two significant main effects were observed (main effect of problem size: $F(1, 76) = 5.503$, $p = .022$, $BF_{10} = 1.792$; main effect of interference: $F(1, 76) = 13.916$, $p < .001$, $BF_{10} = 57.44$), but no significant interaction effect was found for interference level and problem size ($F(1, 76) = 0.014$, $p = .906$, $BF_{10} = 0.319$).

Table 3.3

Multiple regression analyses across items, with interference level and problem size as independent variables, and accuracy and reaction time as dependent variables

Measure	Zero-order correlation	Partial correlation	*B*	*SE B*	β	*t*	*p*	BF_{10}	R^2
Accuracy									
PS	-.173	-.166	0.000	0.001	-.166	-0.772	.449	0.478	.038
IL	-.104	-.092	-0.001	0.002	-.090	-0.421	.678	0.407	
Reaction time									
PS	.287	.300	20.557	14.266	.237	1.441	.164	0.755	.434
IL	.615	.619	157.532	43.637	.595	3.610	.002	23.383	

Note: IL = Interference Level; PS = Problem Size. Problem size was defined as the product of both operands.

3.3.2. Imaging results

3.3.2.1. In-scanner performance

A two-way ANOVA with problem size and interference level as independent variables, and accuracy as a dependent variable was also calculated for the in-scanner task. Here, no significant interaction effects (interaction effect: $F(1, 76) = 0.009$, $p = .925$, $BF_{10} = 0.681$), or main effect of problem size ($F(1, 76) = 1.105$, $p = .296$, $BF_{10} = 0.369$) were observed, but the analysis did reveal a main effect of interference ($F(1, 76) = 4.703$, $p = .033$, $BF_{10} = 1.780$). However, as the Bayes Factor for this main effect of interference is between 1 and 3, indicating anecdotal evidence at best (Jeffreys, 1961), this result must be interpreted with caution. An ANOVA with reaction time as a dependent variable, however, was not possible for the in-scanner task, as, due to our experimental fMRI design with delayed verification, participants were only able to answer 3 seconds after the initial presentation of the problem, leading to inaccurate reaction times. For example, it is possible for a participant to solve a certain item in the first 3 seconds, before being able to actually answer the item, thus not providing an accurate reaction for that item. Because of this issue, only statistical analyses on the reaction time data of the custom multiplication task outside of the scanner are discussed.

To assure that the fMRI signal for the same amount of items per category would be analyzed, we also calculated a 2 × 2 ANOVA on the number of items of each category (i.e., small or large problem size vs. low or high interference level). In this analysis, no significant main or interaction effects were observed, indicating no differences in the amount of items analyzed per condition (interaction effect: $F(1, 76) = 0.006$, $p = .937$, $BF_{10} = 0.072$; main effect of problem size: $F(1, 76) = 0.250$, $p = .618$, $BF_{10} = 0.260$; main effect of interference: $F(1, 76) = 0.433$, $p = .512$, $BF_{10} = 0.282$). Furthermore, the Bayes

Factors (BF_{10}) for this analysis were consistently below 0.3, indicating at least substantial evidence for the null hypothesis of no differences between conditions.

3.3.2.2. fMRI data analysis

The fMRI data were first analyzed through a 2 × 2 ANOVA with problem size (small vs. large) and interference level (low vs. high) as within-subject factors, only modeling the correctly solved items. Due to the stringency of the ANOVA analysis, however, no significantly activated clusters for all main or interaction effects were found. Because of this, we calculated all separate t-contrasts (i.e., small vs. large problem size, large vs. small problem size, low vs. high interference, and high vs. low interference), with an FDR ($p < .05$) correction for multiple comparisons. The results of these analyses can be found in Table 3.4. Significantly activated clusters, however, were only observed for the problem size effect (large vs. small contrast), in fronto-parietal regions, but also in the fusiform gyri. A visual representation of these results can be found in Figure 3.2.

Table 3.4

Regions, peak coordinates, cluster sizes (k), and t-values of the significantly activated clusters (p < .05, FDR corrected) for all separate t-contrasts

Cluster	**Peak coordinates**			**k**	***t***
	x	**y**	**z**		
Problem Size					
Small > Large					
No suprathreshold voxels found					
Large > Small					
R Fusiform gyrus	42	-60	-18	487	7.18
L Fusiform gyrus	-34	-78	-18	429	5.95
R Calcarine gyrus	18	-96	4	223	5.33
R Fusiform gyrus	22	-86	-20	193	5.54
L Superior parietal lobe	-6	-84	48	171	5.72
L Middle/superior frontal gyrus / precentral gyrus	-28	0	56	143	4.63
R Inferior occipital gyrus	46	-82	-6	109	4.82
Interference					
Low > High					
No suprathreshold voxels found					
High > Low					
No suprathreshold voxels found					

Note: Only clusters of 20 voxels or more are reported.

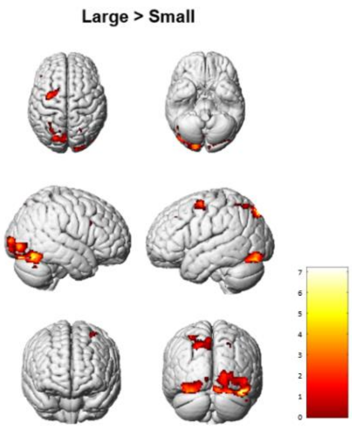

Figure 3.2. Visual representation of differences in brain activation of the t-contrasts (p < .05, FDR corrected) large vs small problem size.

As, when controlling for multiple comparisons, no significantly activated clusters were observed for any of the contrasts for interference, the contrasts were also calculated without controlling for multiple comparisons (p < .001), as a too stringent correction for multiple comparisons could have resulted in a lack of power to detect an effect. These results can be found in Table 3.5. A visual representation of these results can be found in Figure 3.3. In these separate t-contrasts without controlling for multiple comparisons, differences were observed between both conditions, with increased activation in the right middle temporal gyrus, left insula, and right angular gyrus for low interfering items in comparison to high interfering items. However, as these results were only observed without any form of controlling for multiple comparisons, they must be interpreted with caution.

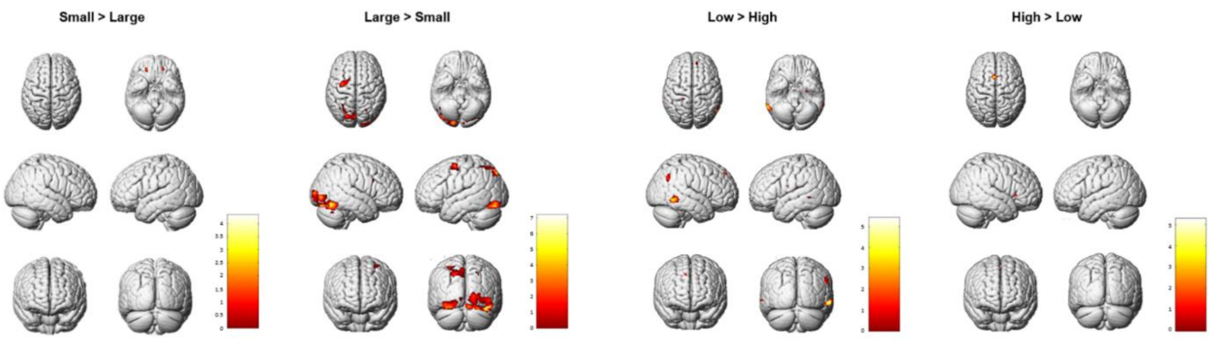

Figure 3.3. Visual representation of differences in brain activation of t-contrasts (p < .001, uncorrected) between small and large items (left) and low and high interfering items (right).

Table 3.5

Regions, peak coordinates, cluster sizes (k), and t-values of the significantly activated clusters (p < .001, uncorrected) for all separate t-contrasts

Cluster	**Peak coordinates**			**k**	***t***
	x	**y**	**z**		
Problem Size					
Small > Large					
R Limbic lobe / ACC	20	36	-6	26	4.24
L Limbic lobe / ACC	-18	42	-4	22	4.33
Large > Small					
R Fusiform gyrus / calcarine gyrus / temporal/occipital lobe	42	-60	-18	1724	7.18
L Fusiform gyrus	-36	-78	-18	727	5.95
L Superior occipital/parietal lobe	-6	-84	48	320	4.31
L Middle/superior frontal gyrus / precentral gyrus	-28	0	56	313	3.74
L Superior parietal lobe	-28	-66	56	75	3.38
R Superior parietal lobe / superior occipital gyrus	24	-64	48	23	3.43
Interference					
Low > High					
R Inferior/middle temporal gyrus / fusiform gyrus	64	-54	-8	172	4.62
L Insula	-38	0	20	110	5.41
R Angular gyrus / supramarginal gyrus	58	-64	38	63	4.23
High > Low					
Cingulate gyrus	8	-22	26	102	5.25
L Posterior medial frontal (BA6)	-6	14	56	83	4.49

Note: All results above were not controlled for multiple comparisons ($p < .001$). Only clusters of 20 voxels or more are reported.

To further explore this clear presence of a neural problem size effect, but possible absence of a neural interference effect, average beta weights were extracted for regions-of-interest (ROIs) expected to be activated by either effect. ROIs were created and beta weights were extracted with the Marsbar toolbox for Matlab (Brett, Anton, Valabregue, & Poline, 2002). Next, for both contrasts of interest (i.e., large vs. small problem size, and high vs. low interference level), one sample t-tests were performed on the beta weights with both frequentist and Bayesian statistics. The results of this analysis can be found in Table 3.6. Results point towards a clear effect of problem size in the left superior parietal lobe and right fusiform gyrus, indicating increased activation for large in comparison to small problems. A significant result, was also observed for the left fusiform gyrus, but the Bayes Factor (BF_{10}) for this t-test was only 1.785, pointing towards anecdotal evidence at best (Jeffreys, 1961). For the high vs. low interference level contrast, significant results were observed for the left angular and middle temporal gyrus, with increased activation observed for low in comparison to high interfering items. The Bayes Factor (BF_{10})

Table 3.6

One sample t-tests for the beta weights of expected ROIs for both high vs. low interference level, and large vs. small problem size contrasts

	Large vs. Small Problem Size			High vs. Low Interference		
	t	***p***	**BF_{10}**	***t***	***p***	**BF_{10}**
L Angular Gyrus	0.534	.599	0.264	-2.396	.027	2.264
R Angular Gyrus	1.447	.164	0.572	-1.494	.152	0.605
L Supramarginal Gyrus	0.986	.337	0.357	-2.031	.057	1.263
R Supramarginal Gyrus	0.248	.807	0.239	-1.779	.091	0.876
L Superior Parietal Lobe	3.194	.005	9.536	-.0194	.848	0.236
R Superior Parietal Lobe	1.990	.061	1.188	-0.484	.634	0.258
L Inferior Frontal Gyrus	1.230	.234	0.449	-0.794	.437	0.308
R Inferior Frontal Gyrus	1.620	.122	0.708	0.089	.930	0.233
L Middle Frontal Gyrus	1.565	.134	0.660	-0.872	.394	0.326
R Middle Frontal Gyrus	1.864	.078	0.988	-0.618	.544	0.276
L Middle Temporal Gyrus	0.175	.863	0.236	-2.168	.043	1.562
R Middle Temporal Gyrus	0.746	.465	0.298	-1.885	.075	1.019
L Fusiform Gyrus	2.251	.036	1.785	-0.706	.489	0.290
R Fusiform Gyrus	2.708	.014	3.890	-0.733	.473	0.295
L Insula	0.758	.458	0.300	0.035	.973	0.232
R Insula	0.831	.416	0.316	0.551	.588	0.266
Anterior Cingulate Cortex	-0.038	.970	0.233	0.141	.890	0.234
Medial Frontal Gyrus	0.904	.377	0.334	-0.648	.525	0.280

Note: L = Left; R = Right.

for these t-tests, however, were only 2.264 and 1.562, respectively, again indicating anecdotal evidenceat best (Jeffreys, 1961). For most of the other ROIs, the Bayes Factors (BF_{10}) were below 1, pointing towards, albeit not necessarily substantial, evidence for the null hypothesis of no differences between both conditions.

3.3.3. Explorative statistical pattern recognition analyses

The inconclusive results of the analyses for interference effect above may have emerged from the implementation of univariate methods to measure the overall activation differences between conditions. Such an approach, however, limits our understanding of the information encoded by neural populations in a certain region. Therefore, it has been suggested that applications of multivariate pattern analysis to fMRI data, such as statistical pattern recognition analysis, allow for the detection of differences between conditions with higher sensitivity than conventional univariate analyses. This is done by focusing on the analysis and comparison of distributed patterns of activity within a certain region, and might thus offer solutions to the limitations of univariate analyses (Ansari, 2008; Bulthé, De Smedt, & Op de Beeck, 2014; Cohen Kadosh & Walsh, 2009; Haxby, 2012). Consequently, we decided to further explore our data through statistical pattern recognition analyses, using PRoNTo (Pattern Recognition for Neuroimaging Toolbox; Schrouff et al., 2013) in SPM12. Statistical pattern recognition is a field within the area of machine learning, concerned with the automatic discovery of regularities in data using computer algorithms. Consequently, these regularities can be used to, for example, classify data into different categories (Bishop, 2006), such as low and high interference, as in the current study.

Using this type of multivariate analysis, we only looked at the interference effect and calculated classification models in 20 different ROIs. Every ROI was defined using predefined ROIs in standard MNI space, using the anatomical WFU PickAtlas Toolbox in SPM. ROIs were first selected on a large spatial scale (i.e., all separate brain lobes), after which ROIs were selected on an intermediate spatial scale (Bulthé et al., 2014), based on the results of our univariate analyses and based on the existing adults studies (e.g., angular gyri, inferior frontal gyri, and insular cortices; De Visscher et al., 2015; De Visscher et al., 2018). To start the analyses, data were put into the system clarifying which data points of each subject were from low or high interfering items, after which a feature set was made per ROI. Subsequently, classification models were specified and estimated, based on a leave-one-subject-out (LOSO) cross-validation scheme. Within this cross-validation structure, the voxel-wise mean was subtracted from each data vector, and samples were constructed by computing the average of all scans for each subject and condition. The LOSO cross-validation then leaves one subject out of the model and tries to classify the data of that subject, based on the data that remained in the model, repeating this process across all subjects. However, as in functional neuroimaging, the assumption that data are independently and identically distributed is often not met, which leads to confidence intervals not always being appropriate, multiple permutation tests (i.e., 1000 for each ROI) were calculated for each ROI. If consistent differences between and similarities within conditions (i.e. low vs. high interference) exist, the model will then be able to classify the data across subjects with an above-chance accuracy. Finally, the model then provides a balanced accuracy (i.e., taking the number of samples in each class into account) of the classification model, as well as an associated p-value of that accuracy.

A full overview of the ROIs (i.e., first looking at all ROIs on a large scale, then on an intermediate scale) and their respective balanced accuracies can be found in Figure 3.4. Performing these analyses, only a small significant classification above chance was observed when looking at the left parietal lobe in its entirety, at a balanced classification accuracy of 52.96% (p = .01). For all other ROIs, no significant classifications above chance were observed across subjects.

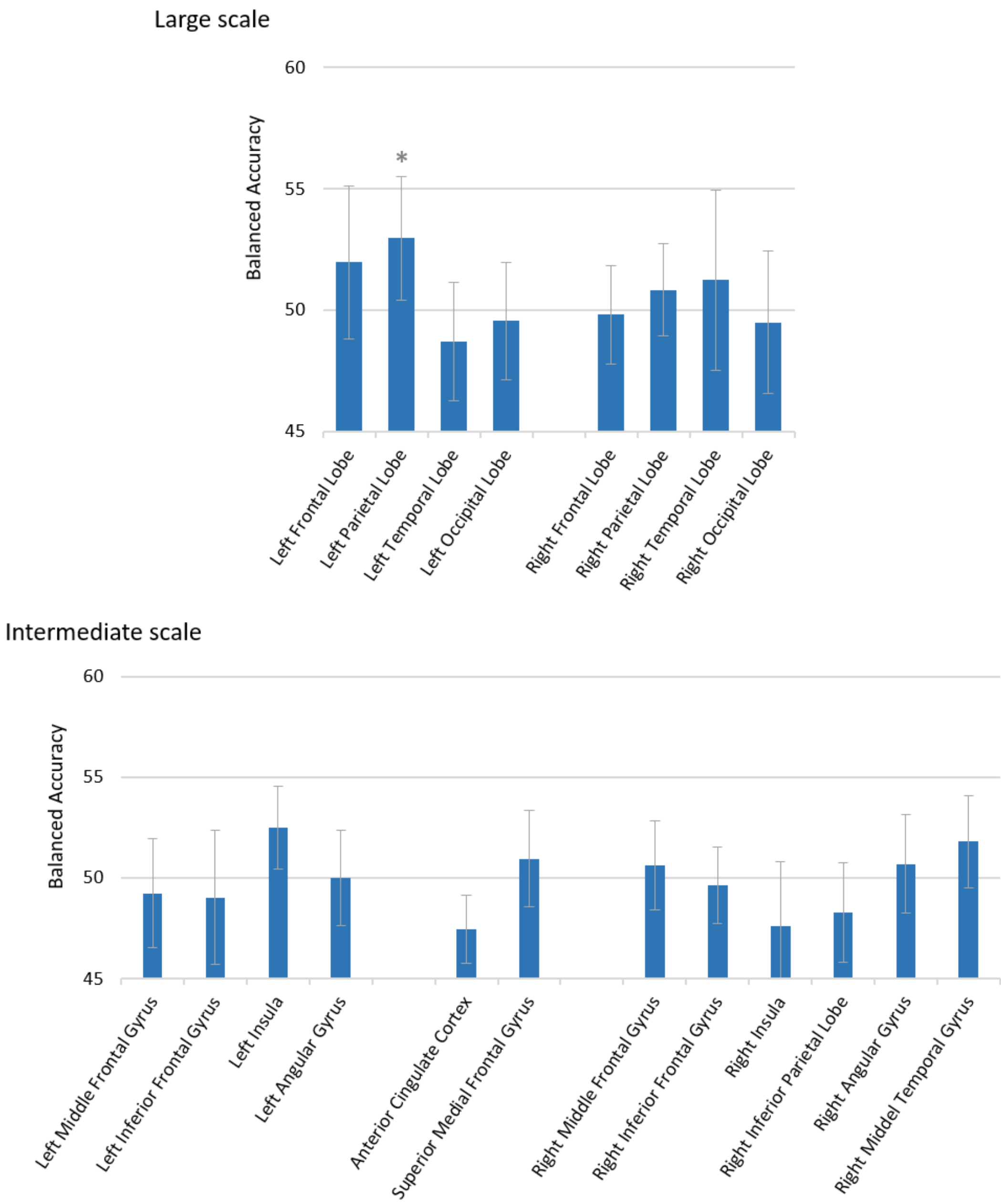

Figure 3.4. Overview of the ROIs and respective balanced accuracies for the statistical pattern recognition analyses, after 1000 permutations for each ROI. On top are the ROIs on a large spatial scale, at the bottom on an intermediate spatial scale.

3.4. Discussion

In order to get more detailed insights into children's neural arithmetic fact retrieval network, and to determine the specific functions of the relevant neural regions more precisely, the present study aimed at investigating the neural basis of both the problem size and interference effect in typically developing 9- to 10-year-olds (i.e., 4th graders), by implementing a similar design as previous adult studies (De Visscher et al., 2015; De Visscher et al., 2018). Following this interference paradigm, our behavioral results show clear effects of both problem size and interference level for multiplication items, both across items and across participants. These results corroborate previous research findings in similar age groups (with similar □ values in the regression models; De Visscher and Noël, 2014b; De Visscher, Noël, & De Smedt, 2016), by showing that both larger and high interfering problems take significantly longer to solve, and that this behavioral effect for reaction time is stronger for the interference level. For accuracy, also in accordance with previous studies in typically developing children, no clear effects of problem size or interference level could be found. A significant effect on the accuracy of the in-scanner task was observed for interference level, yet, as the Bayes Factor for this main effect of interference was between 1 and 3, indicating anecdotal evidence at best, this result must be interpreted with caution.

At the neural level, the t-contrasts, corrected for multiple comparisons, displayed clear neural differences for problem size. Increased activation for large items was found in bilateral fusiform gyri, the left superior parietal lobe, and the left precentral gyrus. These results were also confirmed by our ROI analysis. Furthermore, these results (i.e., the increased activation of fronto-parietal and temporal regions for larger items) agree with previous studies on the problem size effect (e.g., Stanescu–Cosson et al., 2000; Molko et al., 2003), but show weaker effects, as the current study only looked at multiplication within fact retrieval, and not at other, procedural strategies. These results also concur greatly to the results of the interference studies in adults (De Visscher et al., 2015; De Visscher et al., 2018) and with previous developmental imaging work on the problem size effect (De Smedt et al., 2011). We also observed a problem size effect in the visual areas of the occipital cortex. These differences could be due to small but consistent differences in the presentation of the answers for small and large items. For example, the displayed false answer to the small items could be single digit (e.g., $3 \times 4 = 9$), while this could not be the case for the large items. These observed differences could also be due to cognitively induced visual processing mechanisms. For example, it is possible that participants may have used more visual resources (e.g., looking at the problem multiple times) to process more difficult items, where easier items would not need reevaluation of the presented stimuli.

For the interference effect, on the other hand, results of the current study are inconclusive, as no activation differences between low and high interfering problems were found in the full factorial model, or in the whole-brain separate t-contrasts when correcting for multiple comparisons. Accordingly, these results are in contrast to previous adult studies (De Visscher et al., 2015; De Visscher et al., 2018), which, in a full factorial model, mainly observed increased activation in the angular gyri for low interfering items, and in a bilateral fronto-parietal network, including superior medial gyri, middle frontal gyri, inferior and superior frontal gyri, intraparietal sulci, and insular cortices for high interfering problems.

As, against our expectations, no neural differences between both conditions were observed in our full factorial model, or in the separate t-contrasts when controlling for multiple comparisons, we decided to perform additional explorative analyses (i.e., separate t-contrasts at $p < .001$ uncorrected, ROI analyses, and statistical pattern recognition analyses). In the separate t-contrasts without controlling for multiple comparisons, differences were observed between both conditions, with increased activation in the right middle temporal gyrus, left insula, and right angular gyrus for low interfering items in comparison to high interfering items. The results of our one sample t-tests with the beta weights of the high vs. low interference level contrast in theoretically relevant ROIs, on the other hand point towards increased activation in the left angular and middle temporal gyrus, but only provide anecdotal evidence at best (Jeffreys, 1961). These findings of increased activation in the angular gyrus might concur with the adult studies of De Visscher et al. (2015) and De Visscher et al. (2018), in which similar activation was found for low interfering items. However, as the results of the current study were only observed without controlling for multiple comparisons, or only provide anecdotal evidence at best in case of our ROI analysis, this comparison between De Visscher et al. (2015) and De Visscher et al. (2018), and our study must be interpreted with great caution. Furthermore, for the other ROIs, the Bayes Factors (BF_{10}) were consistently below 1, which points towards, albeit not necessarily substantial, evidence for the null hypothesis of no differences between both conditions.

Additionally, when conducting the explorative statistical pattern recognition analyses, the model was only able to classify subjects' neural responses to low and high interfering items when looking at the left parietal lobe in its entirety, and the effect was very small; only a barely above-chance classification accuracy of 52.96% was observed. In the angular gyri, however, no significant classification results were found. In all, this secondary set of analyses confirms the notion that no clear evidence of a neural interference effect was to be found in this sample.

In contrast to our hypotheses, and to the previous studies in adults (De Visscher et al., 2015; De Visscher et al., 2018), we only observed a neural effect of problem size, as no clear neural distinctions between low and high interfering items were observed at our predefined threshold. This is particularly unexpected as the behavioral data revealed clear effects of both problem size and interference. This inconsistency across children and adults, and across behavioral and neural results could be due to a number of reasons. First of all, studies in adults (De Visscher et al., 2015; De Visscher et al., 2018) especially point towards increased activation in the angular gyri for low interfering items in comparison to high interfering items. The fact that these results were not observed in children at the FDR level, leads to the prediction that, as (mathematical) development progresses, more neural distinctions are made between multiplication items, based on the degree to which they are similar to previously learned items (i.e., based on their level of interference). Accordingly, it is possible that, as the fourth graders of the current sample, in contrast to adults, are still frequently in contact with these multiplication items, clear neural distinctions are not yet made at this stage of development. In line with this notion, we suggest that similar research needs to be done across different age groups, such as 2nd graders, who are at the beginning of learning multiplication tables, or adolescents, who are no longer being trained to memorize these multiplication tables, but have reached higher levels of mathematical ability. Similarly, longitudinal research following children across arithmetic development is certain to deliver critical insights into how these effects of problem size and interference develop.

Furthermore, it is important to emphasize that stimuli for both interference conditions were highly similar (i.e., both conditions contained multiplication items of both small and large problem sizes, all of which were regularly solved through fact retrieval), and that behavioral differences between both conditions were only observed for reaction time. Consequently, it is possible that neural differences between both conditions should not be investigated at a spatial level, looking at certain regions showing increased activation for one condition in comparison to the other, but more at a temporal level, focusing on the speed to which certain brain regions respond to either condition. From this point of view, the use of fMRI might not have been optimal for studying the neural differences between high and low interfering items in children. Data collection methods with a greater temporal resolution (e.g., electroencephalography) should be implemented.

Another possible venture for future research could lie in the analysis of made errors, as, if the interference effect were to affect performance differently, this would become apparent in erroneous trials. This type of analysis, however, was not possible in the current study due to the high accuracy on the in-scanner task. An analysis of errors could be possible when adjusting the research design to elicit as many errors as possible, by for example lowering the allotted time to answer, thus increasing pressure and possibly maximizing the amount of errors made. Further research investigating this principle is necessary.

Finally, the interference effect has also been studied in atypical populations. For example, in a case study by De Visscher and Noël (2013), a hypersensitivity to the interference parameter was established in a certain profile of dyscalculia, with a restricted deficit in arithmetic facts knowledge. Consequently, it is also possible that clear neural correlates of the interference effect in children might only be found in children with low arithmetic performance, or even more specifically, in atypical populations, such as children with developmental dyscalculia, and therefore not in the current research sample. Future fMRI studies should explore this possibility.

In conclusion, our results confirm the observations of previous behavioral studies that both the problem size and level of interference of multiplication items that are retrieved from memory affect the speed with which children solve those items. Next to these behavioral effects, our results also point towards clear neural distinctions between small and large multiplication items. However, no clear neural differences could be observed when contrasting low and high interfering items (at the FDR level), suggesting that, even though strong behavioral effects were found, the neural basis of this effect is not as strong as was previously observed in adults (De Visscher et al., 2015; De Visscher et al., 2018), and is not as strong as the problem size effect, but might develop over time. Future research is needed to evaluate this.

CHAPTER 4

The Association of Grey Matter Volume and Cortical Complexity with Individual Differences in Children's Arithmetic Fluency

Abstract

Only a small amount of studies have looked at the structural neural correlates of children's arithmetic. Furthermore, these studies mainly implemented voxel-based morphometry (VBM), which only takes into account the volume of regions, without looking at other structural properties. The current study aimed to contribute knowledge on which brain regions are important for children's arithmetic at a structural level, by not only implementing VBM, but also cortical complexity analyses, based on the fractal dimension index. This complexity measure describes a characteristic of surface shape. Data of 43 typically developing 9-10 year-olds were analyzed. All children were asked to take part in two test sessions: behavioral data collection and MRI data acquisition. For data analysis, mean values for volume and cortical complexity were estimated within regions of interest (ROIs) and extracted for further analysis. The selected ROIs were based on regions found to be related to children's mathematical abilities in previous research. Results point towards associations between arithmetic fluency and the volume of the right fusiform gyrus, as well as the cortical complexity of the left postcentral gyrus, right insular sulcus, and left lateral orbital sulcus. Remarkably, no significant associations were observed between the children's arithmetic fluency and the volume or cortical complexity of typically arithmetic-associated parietal regions, such as the superior parietal lobe, intraparietal sulcus, or angular gyrus. Accordingly, the current study highlights the importance of structural characteristics of brain regions other than typically arithmetic-associated parietal regions for children's arithmetic fluency.

4.1. Introduction

Arithmetic, or the ability to add, subtract, multiply or divide numbers, is an essential skill for further mathematical development (Kilpatrick, Swafford, & Findell, 2001), with a ubiquitous role in daily life, especially for children. Over the past few years, many neuroimaging studies have aimed to unravel the neural basis of children's arithmetic, but have mainly focused on functional neural aspects (Arsalidou et al., 2018, for a meta-analysis; Peters and De Smedt, 2018, for a review). The amount of studies looking at the structural neural correlates of children's arithmetic is scarce (e.g., Evans et al., 2015; Isaacs et al., 2001; Price et al., 2016). The current study aims to contribute knowledge on which brain regions are important for children's arithmetic at a structural level, by not only implementing voxel-based morphometry (VBM), but also cortical complexity analyses, based on the fractal dimension index (Yotter et al., 2011), and focuses on children with a small age range (9-10 year-olds, all 4th graders) to minimize maturational confounds.

In children, accumulating evidence of functional magnetic resonance imaging (fMRI) research points towards an arithmetic brain network involving a large set of interconnected areas, including frontal (i.e., both ventro- and dorsolateral prefrontal cortex), parietal (i.e., inferior and superior parietal lobes), occipito-temporal, and medial-temporal (i.e., hippocampus) regions. Though this network shows some similarities to the network observed in adults (Menon, 2015), children's arithmetic network is different, as overall less involvement of the superior parietal lobe is observed, along with larger involvement of the hippocampus, occipito-temporal, and prefrontal medial regions (Arsalidou et al., 2018, for a meta-analysis; Peters and De Smedt, 2018, for a review).

Only a small amount of studies have looked at the structural neural correlates of children's arithmetic. The few studies that did study these structural correlates (see Peters and De Smedt, 2018 for an overview) mainly implemented VBM, which typically uses T1-weighted volumetric MRI scans and performs statistical tests across voxels to identify volume differences between groups. Accordingly, a series of t-tests can be performed at every voxel to identify differences in patterns of regional anatomy between groups of subjects (Whitwell, 2009). Using this method of data analysis, some structural imaging studies have compared groups of children who differed in their level of arithmetic skill (Isaacs et al., 2001; Ranpura et al., 2013; Rotzer et al., 2008; Rykhlevskaia et al., 2009). The first study reporting such data was Isaacs et al. (2001), who performed VBM analyses on 24 adolescents born preterm at 30 weeks gestation or less, but without any neurological disabilities, comparing those with and without difficulties in arithmetic calculation. Doing so, the study observed reduced grey matter volume in the left intraparietal sulcus for the children with a deficit in calculation ability. Subsequent studies then similarly aimed to investigate differences between children with dyscalculia, which is a specific neurodevelopmental learning disorder characterized by difficulties in calculation which cannot be explained by intellectual disabilities, uncorrected sensory problems, mental or neurological disorders or inadequate instruction (American Psychiatric Association, 2013), and typically developing children. For example, a study by Rotzer et al. (2008) compared 12 9-year-old children with developmental dyscalculia to 12 age-matched controls, and observed reduced grey matter volume for the children with dyscalculia in the right posterior parietal cortex (including the right intraparietal sulcus), the anterior cingulum, the left inferior frontal gyrus, and the bilateral middle frontal gyrus (Rotzer et al., 2008). A later study by Ryklehvskaia et al. (2009) investigated 24 7- to 9-year-old children with developmental dyscalculia, and compared them to a group of typically developing children matched on age, gender, intelligence, reading abilities and working memory capacity. This study also revealed reduced grey matter volume for the children with dyscalculia in the superior parietal lobule (again including the right intraparietal sulcus), but also in the bilateral fusiform gyrus, parahippocampal gyrus and right anterior temporal cortex. Finally, a study by Ranpura et al. (2013) investigated volumetric differences between 11 8- to 14-year-old children with dyscalculia, and typically developing children, the results of which again pointed towards reduced grey matter volume in the right parietal cortex for children with dyscalculia, as well as volumetric reductions in the right occipital, fusiform and parahippocampal gyri. In all, these studies consistently showed that poor arithmetic performance is accompanied by less volume in (mostly right-hemispheric) parietal grey matter, but also highlight the importance of regions outside of the parietal cortex for children's arithmetic.

The group comparison method of the studies above, however, is limited, as it only allows for categorical comparisons of, for example, clinical or atypical groups of children to typically developing peers. Within VBM, however, it is possible to perform regression analyses across voxels to assess neuroanatomical correlates of cognitive or behavioral skills, thus applying a more dimensional approach (Whitwell, 2009). Doing so, the structural correlates of arithmetic can also be examined within a typically developing population. Surprisingly few studies, however, have examined this association between grey matter and arithmetic in typically developing children. A study by Li et al. (2013) revealed that, in 59 9- to 11-year-old Chinese children, individual differences in scores on the arithmetic subtest of the WISC-RC were significantly and positively correlated with the grey matter volume in the left intraparietal sulcus. Using a longitudinal design, Evans et al. (2015) investigated whether grey matter volume in early childhood (43 children, 7- to 9-year-olds) was predictive of outcomes in numerical abilities (based on the Numerical Operations subscale of the WIAT-II) six years later. The study reported that grey matter volumes of various parts of the arithmetic network (i.e., posterior parietal cortex, ventral occipito-temporal cortex, and prefrontal cortex) predicted the growth in arithmetic across primary school. Price et al. (2016) also investigated the relation between grey matter volume and math competence (based on the Woodcock–Johnson III Tests of Achievement) over one year in a sample of 50 6- to 8-year-olds. Their results displayed that grey matter volume in the left intraparietal sulcus at the end of the 1st grade is related to math competence at the end of the 2nd grade. Grey matter volume of the intraparietal sulcus, however, did not change over that year, but was still correlated with math competence at the end of 2nd grade. Finally, Supekar et al. (2013) had a group of 3rd graders (24 children, 8-9 years-old) follow an 8 week math tutoring program that focused on efficient counting and fact retrieval, in between two structural scanning sessions. Accordingly, they observed that the volume of the right hippocampus predicted the learning gains of the one-on-one tutoring sessions, with larger hippocampal volumes before the intervention predicting larger intervention gains, confirming the role of the hippocampus in arithmetic fact retrieval.

All of the studies above confirm that the grey matter volume of different neural regions (e.g., superior parietal lobe, intraparietal sulcus, inferior and middle frontal gyrus, fusiform gyrus, hippocampus) is associated with individual differences in arithmetic performance. These studies, however, have often merged data across wide age ranges (e.g., 8-14 years old), which might lead to the over-interpretation of associations between differences in volume and differences in mathematical development, as results might still be affected by maturational confounds. To minimize such maturational confounds, samples with small age ranges are necessary to clearly define brain regions of the arithmetic network for which grey matter volume is correlated to arithmetic fluency at a certain point in development.

New techniques for structural data analysis have also arisen over the past years, allowing the study of structural brain differences to go beyond looking at volume alone. For example, surface-based morphometry (SBM) has numerous advantages over the use of volumetric data, as it has been shown that the implementation of brain surface meshes for spatial registration increases the accuracy of brain registration compared to mere volume-based registration (Desai, Liebenthal, Possing, Waldron, & Binder, 2005). Accordingly, additional metrics of cortical structure are applicable. One such metric is cortical complexity, which quantifies the spatial frequency of gyrification and fissuration of the brain surface (Luders et al., 2004), and is most commonly measured through the use of a gyrification index, defined as the ratio of the inner surface size to the outer surface size of an outer hull. However, gyrification analyses have certain shortcomings, such as that the gyrification metric depends on how the outer hull is defined, on the normalization of the brain to reduce the effect of brain size, and on noise in the surface reconstruction, which could artificially inflate the surface area without corresponding to the actual anatomy. These shortcomings can be resolved by quantifying cortical complexity through the fractal dimensionality index (Yotter et al., 2011).

The fractal dimensionality index was originally designed to quantify the structure of fractals (Kiselev, Hahn, & Auer, 2003), but can describe a characteristic of the surface shape (see Yotter et al., 2011 for an in depth description of how brain complexity is measured), without relying on the definition of an outer hull (Lopes & Betrouni, 2009). In this sense, cortical complexity does not directly measure the intuitive meaning of the word complexity, such as the surface being more detailed. A fractal is a structure that is self-similar across a range of scales, making the complexity analysis correspond to how space-filling the fractal surface is. As such, regions with high fractal dimensionality values generally appear to be more periodically spaced (e.g., like a sine wave with regular peaks and troughs). This may be because the more periodically spaced structures also tend to fill more space over the range of scales examined for derivation of complexity values (Yotter et al., 2011).

Fractal dimensionality analyses have successfully been implemented in studies comparing patient groups, such as individuals with Alzheimer's disease (e.g., Ruiz de Miras et al., 2017) or Williams syndrome (e.g., Thompson et al., 2005), to controls, demonstrating a decline in fractal dimensionality in these group comparisons. Moreover, being sensitive to other differences in grey matter structure that are not indexed by volume or cortical thickness, cortical complexity has also been used to study age and gender related differences in brain structure (Luders et al., 2004; Madan & Kensinger, 2016), and, most notably, to study differences in cognitive function (King, et al., 2010; Im et al., 2006; Mustafa et al., 2012; Sandu et al., 2014). For example, Im et al. (2006) observed positive correlations between whole-brain fractal dimensionality and both IQ and years of education. Noteworthy, however, was that the correlations with education were slightly stronger than those with IQ, indicating a possible influence of education-related development on cortical complexity. King et al. (2010) even found that fractal dimensionality correlated more strongly with scores on a cognitive battery compared to measures of cortical thickness and gyrification. A study by Mustafa et al. (2012) observed that seniors with greater whole-brain white matter complexity had higher fluid intelligence scores and less evidence of age-related cognitive decline. Finally, Sandu et al. (2014) observed that decreases in late life cortical complexity are associated with declines in information processing speed, auditory-verbal learning, and reasoning. Cortical complexity analyses have not yet been used in mathematical cognition, but the studies above support the use of cortical complexity through fractal dimensionality as a sensitive metric for capturing relations between brain structure and cognitive function. Their implementation within the field of mathematical cognition could thus provide interesting insights into children's arithmetic brain network.

Against this background, the first aim of the current study was to use VBM to study how structural differences in brain anatomy relate to differences in typically developing 9- to 10-year-old children's arithmetic fluency. Secondly, as general associations have previously been observed between cognitive function and cortical complexity, we also aimed to examine if specific associations between arithmetic fluency and the cortical complexity of regions in children's arithmetic brain network (e.g., in the parietal cortex, including the superior and inferior parietal lobes, and the intraparietal sulcus) can be found. While doing these analyses, we will implement a region of interest (ROI) approach based on the neural regions that were previously found to be both structurally and/or functionally related to children's mathematical abilities (Arsalidou et al., 2018; Peters and De Smedt, 2018). To the best of our knowledge, this study is the first to go beyond VBM analyses by investigating such associations through fractal dimensionality.

4.2. Methods

4.2.1. Participants

For the current study, data of 50 typically developing Flemish 4th graders were collected, yet data of 3 children were discarded due to technical acquisition problems (*n* = 1), excessive motion (*n* = 1), or problems during standardized testing (*n* = 1), and data of 4 more children were discarded after data quality control (see below for more details). Of the remaining 43 participants (ages 9 to 10; M = 9.68 years, SD = 0.34) 23 were boys, 20 were girls, and 8 children were left-handed. No participants had a history of learning difficulties, or neurological or psychiatric disorders. All participants were recruited via the elementary school they attended, in the surrounding area of the university, and were given a financial compensation for their participation. Written informed consent was obtained from a parent or legal guardian of each participating child. The study was approved by the Medical Ethical Committee of the University of Leuven (S59167).

Each child was asked to take part in two test sessions. During the first session, behavioral data were collected through various standardized measures. The second session included the acquisition of the MRI data, and always followed the first session by two to three weeks (M = 19.54 days, SD = 6.34). This MRI acquisition session also contained the collection of diffusion (Polspoel et al., 2018) and functional data (Polspoel et al., 2017) that have been reported elsewhere. Only the T1-data and their association with the general standardized tests are reported in the current study.

4.2.2. Standardized assessment

The standardized assessment session consisted of the evaluation of arithmetic (i.e., our main variable of interest), reading (i.e., to check the specificity of our results to arithmetic), and intelligence and motor reaction time (i.e., as control variables). To measure children's arithmetic competence, the Tempo Test Arithmetic (TTA; de Vos, 1992) was used. This standardized arithmetic test, which is similar to the Math Fluency subtest of the Woodcock-Johnson III tests of Achievement (Woodcock et al., 2003), is very sensitive to individual differences in arithmetic fluency. The TTA contains five columns of arithmetic items: one column per operation and a fifth column with mixed operations. Each column starts with single digit items and increases in difficulty. Participants get one minute per column to write down as many correct answers as possible. As the current sample included a small age range, all participating children were part of the same norm group, and thus the raw scores (i.e., the sum of the amount of correctly answered items in each column) were used for statistical analysis.

Reading was assessed through tests that are similar in their conceptualization as the TTA, as they are timed tests as well: the One Minute Test (OMT; Brus and Voeten, 1979) and the Klepel (van den Bos et al., 1994), which respectively measure the reading of words and pseudowords. Both the OMT and Klepel consist of 116 words, but for the OMT, the children get one minute to correctly read aloud as many words as possible, while for the Klepel, the time limit is set to two minutes, and the children read aloud pseudowords. Again, raw scores (i.e., the amount of correctly read words) were used for statistical analysis.

Next, the WISC-III-NL Block Design and Vocabulary subtests were used to measure intellectual ability, as measures of performance and verbal IQ, respectively (Wechsler, 2005). For intellectual ability, norm scores were used for the statistical analyses. Finally, to measure motor reaction time, participants had to indicate which of two figures (always a circle, triangle, square, star, or heart; one on the left, one on the right) presented on a computer screen was filled in white by, as quickly as possible, pressing the corresponding key. Accuracy and reaction time were recorded for each trial, yet, as ceiling levels were reached for accuracy, only reaction time was used for the subsequent analyses (De Smedt & Boets, 2010).

4.2.3. MRI data acquisition

MRI scanning was performed with a Philips Ingenia 3.0T CX MRI scanner with a SENSE 32-channel head-coil, located at the Department of Radiology of the University Hospital in Leuven, Belgium. Wash cloths were used to stabilize the children's heads and consequently minimize head motion. The anatomical T1 images were acquired with the following sequence: 0.98 × 0.98 × 1.2 mm voxel size, 256 × 256 acquisition matrix, 8º flip angle, TE 4.6 ms, 250 × 250 × 218 mm field of view (approximately 8 minutes of scanning time). As part of data collection for different studies (e.g., Polspoel et al., 2017; Polspoel et al., 2018), the scanning session also included four functional MRI runs of 5 minutes, and a diffusion MRI run of 12 minutes, leading to a total scanning time of approximately 40 minutes.

All preprocessing was done with the Computational Anatomy Toolbox (CAT12) within the Statistical Parametric Mapping software package for Matlab (SPM12, Wellcome Department of Cognitive Neurology, London), following the standard processing pipeline within the CAT12 software. First, preprocessing included segmentation of the anatomical images; both grey matter and surface estimations were calculated. Next, data quality and sample homogeneity was tested through the Mahalanobis distance. This is a combination of weighted overall image quality, which is a measure of noise and spatial resolution before preprocessing, and mean correlation, which is a measure of the homogeneity of the data and thus the quality after preprocessing. Data of four subjects were discarded, as their Mahalanobis distance was larger than two standard deviations of the sample average. Finally, spatial smoothing was performed with 8 mm (VBM) and 20 mm (cortical complexity) FWHM Gaussian smoothing kernels.

4.2.4. Selection of ROIs

ROIs for statistical analyses were selected from the available atlases in the CAT12 software package (i.e., the Hammers and lpba40 atlases for VBM; the aparc.a2009s atlas for cortical complexity). The selected ROIs were based on regions found to be related to children's mathematical abilities in previous, both structural and functional, research (Arsalidou et al., 2018; Peters and De Smedt, 2018). After preprocessing, mean values for volume and cortical complexity were estimated within each ROI and extracted for further analysis. An overview of the selected ROIs can be found in Table 4.1.

4.2.5. Statistical analyses

For statistical analyses, the JASP software package (JASP Team, 2017) was used to calculate Pearson correlations and their corresponding Bayes Factors between the results of the TTA and the extracted mean values of the ROIs. The Bayesian approach has the advantage to quantify the evidence that data provide for one hypothesis over another (Andraszewicz et al., 2015). Accordingly, Bayes factors (BF_{10}) of 1, 1-3, 3-10, 10-30, 30-100, or > 100 respectively point towards no, anecdotal, substantial, strong, very strong, or decisive evidence for the hypothesis of an association between two variables (Jeffreys, 1961). The results of the frequentist approach to statistical testing are also reported. For these analyses, the Bonferroni method of controlling for multiple comparisons (p = Target Alpha Level / number of ROIs; $p = .05/42 = .001$ for VBM and $p = .05/78 = .0006$ for cortical complexity) was implemented. Partial correlations were also calculated with IQ and motor reaction time simultaneously added as control variables. To test the specificity of the results, the significant correlations were also calculated with our reading measure as to check whether any observed associations with arithmetic are also observed with another symbolic academic skill (i.e., reading), measured in a similar, i.e., timed, way.

Table 4.1

Overview of selected ROIs for the VBM and cortical complexity analyses

	Frontal	**Parietal**	**Temporal**	**Occipital**
Voxel-based morphometry	Orbito Front Gyr	Postcentral Gyr	Ant Med Temp Lobe	Fusiform Gyr
	Inf Front Gyr	Inf Par Lobe	Fusiform Gyr	Lingual Gyr
	Mid Front Gyr	Sup Par Lobe	Inf Mid Temp Gyr	Lat Occ Lobe
	Sup Front Gyr	Angular Gyr	Sup Temp Gyr	
	Ant Cingulate Gyr	Supramarginal Gyr	Post Temp Lobe	
	Precentral Gyr	Insular Cortex	Insular Cortex	
	Insular Cortex		Hippocampus	
Cortical complexity	Orbital Sulc/Gyr	Postcentral Sulc/Gyr	Inf Temp Sulc/Gyr	Ant Occ Sulc
	Inf Front Sulc/Gyr	Inf Par Lobe	Mid Temp Gyr	Inf Occ Gyr & Sulc
	Mid Front Sulc/Gyr	Sup Par Lobe	Sup Temp Sulc/Gyr	Par-Occ Sulc
	Sup Front Sulc/Gyr	Angular Gyr	Med Occ-Temp Sulc	Med Occ-Temp Sulc
	Ant Cingulate Gyr	Supramarginal Gyr	Lat Occ-Temp Sulc	Lat Occ-Temp Sulc
	Precentral Sulc/Gyr	Intraparietal Sulc	Insular Sulcus/Cortex	Lingual Sulc/Gyr
	Insular Sulc/Cortex	Precuneus	Parahippocampal Gyr	Fusiform Gyr
		Par-Occ Sulcus	Fusiform Gyr	
		Insular Sulc/Cortex		

Note: Sup = Superior; Inf = Inferior; Ant = Anterior; Post = Posterior; Mid = Middle; Med = Medial; Lat = Lateral; Front = Frontal; Par = Parietal; Temp = Temporal; Occ = Occipital; Gyr = Gyrus; Sulc = Sulcus. All ROIs were looked at bilaterally.

4.3. Results

4.3.1. Behavioral results

Figure 4.1 displays box plots with the descriptive statistics of arithmetic, reading, intellectual ability (all in standardized scores), and motor reaction time. These box plots demonstrate that our sample shows proper variation and has means close to the expected population averages. As the results of the different columns of the TTA highly correlated with each other, ranging from r = .616 to r = .818, only the raw total score was used for analysis (M = 102.12, SD = 19.01, Min = 73, Max = 160). Important to note is that, even though the minimum score for some of the tasks was low, none of the participating children had been diagnosed with any type of learning disorder or intellectual disability.

4.3.2. Voxel-based morphometry

Results of the VBM analyses only point towards statistical significance ($p < .05$) and substantial evidence ($BF_{10} > 3$) for an association between arithmetic fluency and the right fusiform gyrus (r = .376; BF_{10} = 3.760; p = .013). A visualization on transverse slices and a scatterplot of this correlation can be found in Figure 4.2. Due to the high amount of ROIs under study, the results using a frequentist approach to statistics did not survive after controlling for multiple comparisons. However, Bayesian statistics are affected less by this multiple comparison problem (Dienes, 2011), and still point towards substantial evidence for an association between arithmetic and the volume of the right fusiform gyrus, although these results must be interpreted with caution.

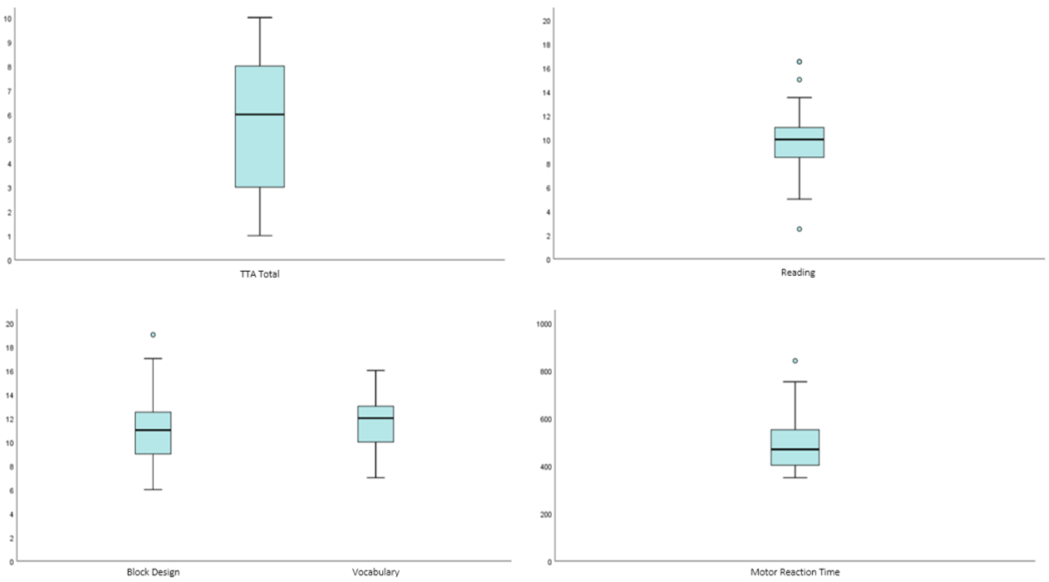

Figure 4.1. Box plots displaying performance on arithmetic, reading, intellectual ability, and motor reaction time.

Note: The scores for arithmetic, intellectual ability, and reading are standardized scores. The scores on the arithmetic test are standardized as M = 5, SD = 2, with a maximum of 10. The scores on the intelligence and reading tests are standardized as M = 10, SD = 3, with a maximum of 19. The scores for motor reaction time are raw scores displaying the average reaction time.

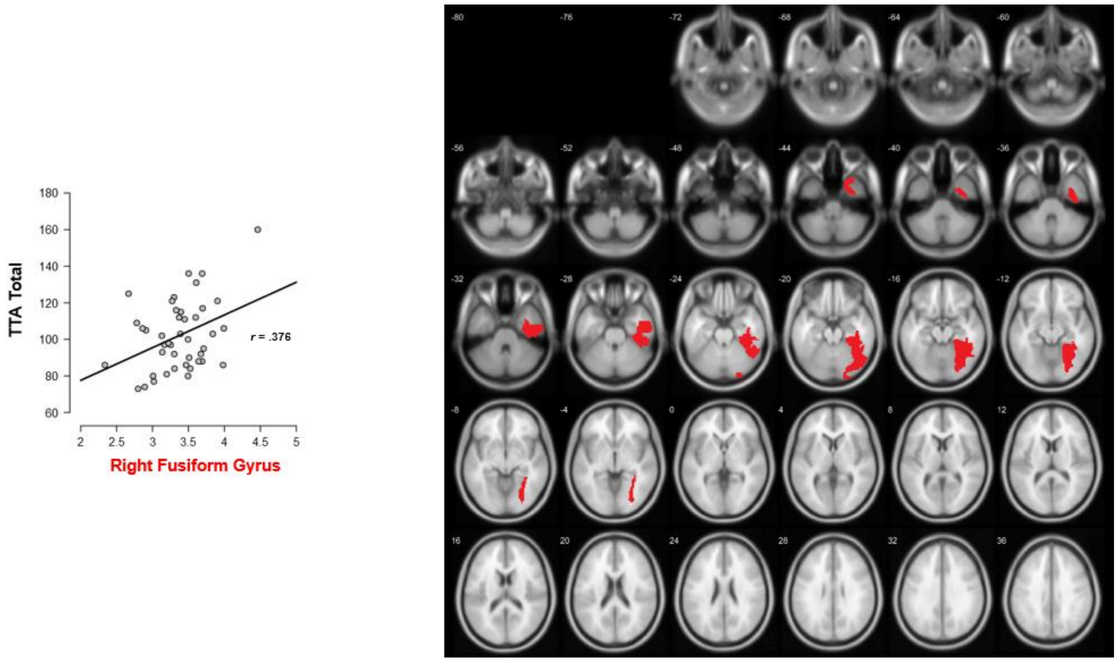

Figure 4.2. Transverse slices and scatterplot with fit line of the correlation between the total score of the TTA and the volume of the right fusiform gyrus.

Remarkably, no significant associations were observed between the children's arithmetic fluency and the volume of previously observed, typically arithmetic-associated parietal regions (i.e., superior parietal lobe, angular gyrus and supramarginal gyrus). Furthermore, the Bayes factors (BF_{10}) for the correlations of the TTA with the volume of these regions were consistently below 1, pointing towards, albeit not necessarily substantial, more evidence for the null hypothesis of no association between these parietal regions and arithmetic fluency. These results can be found in Table 4.2.

Table 4.2

Correlations between arithmetic fluency and typically arithmetic-associated parietal regions for the VBM analyses

		L SPL	R SPL	L AG	R AG	L SMG	R SMG
TTA Total	r	.046	.191	.171	.098	.099	.113
	BF_{10}	0.198	0.395	0.340	0.229	0.231	0.244
	p	.767	.219	.272	.533	.526	.472

Note: SPL = Superior Parietal Lobe; AG = Angular Gyrus; SMG = Supramarginal Gyrus. The used atlases only consider VBM possible on gyri, hence the absence of the intraparietal sulcus in this analysis, yet all typically arithmetic-associated areas that surround the intraparietal sulcus (i.e., postcentral gyrus, superior parietal lobe, angular gyrus, and supramarginal gyrus) were included. The critical p-values for Bonferonni correction for multiple comparisons is $p < .001$.

Finally, to assess the specificity of our results, correlations were calculated with our reading measure, yet no substantial evidence or statistical significance (r = .231, BF_{10} = 0.555, p = .136) was observed for the correlations between reading and the right fusiform gyrus.

4.3.3. Cortical complexity

For the cortical complexity analyses, statistical significance (p < .05) and substantial evidence (BF_{10} > 3) was observed for associations with the left postcentral gyrus (r = .539; BF_{10} = 155.773; p < .001), right insular sulcus (r = .425; BF_{10} = 9.388; p = .005), and left orbital sulcus (r = .382; BF_{10} = 4.183; p = .011). Visualizations on transverse slices and scatterplots of these correlations are displayed in Figure 4.3. These observed correlations remained significant when simultaneously correcting for intellectual ability and motor reaction time. Due to our stringent control for multiple comparisons, only the results for the left postcentral gyrus survived after controlling for multiple comparisons.

For the cortical complexity analyses as well, no substantial evidence was observed for associations between the children's arithmetic fluency and typically arithmetic-associated parietal regions (i.e., the intraparietal sulcus, superior parietal lobe, angular gyrus and supramarginal gyrus). The Bayes factors (BF_{10}) for these correlations were consistently below 1 (except for the left angular gyrus), again pointing towards more evidence for the null hypothesis of no association between the variables. These results can be found in Table 4.3.

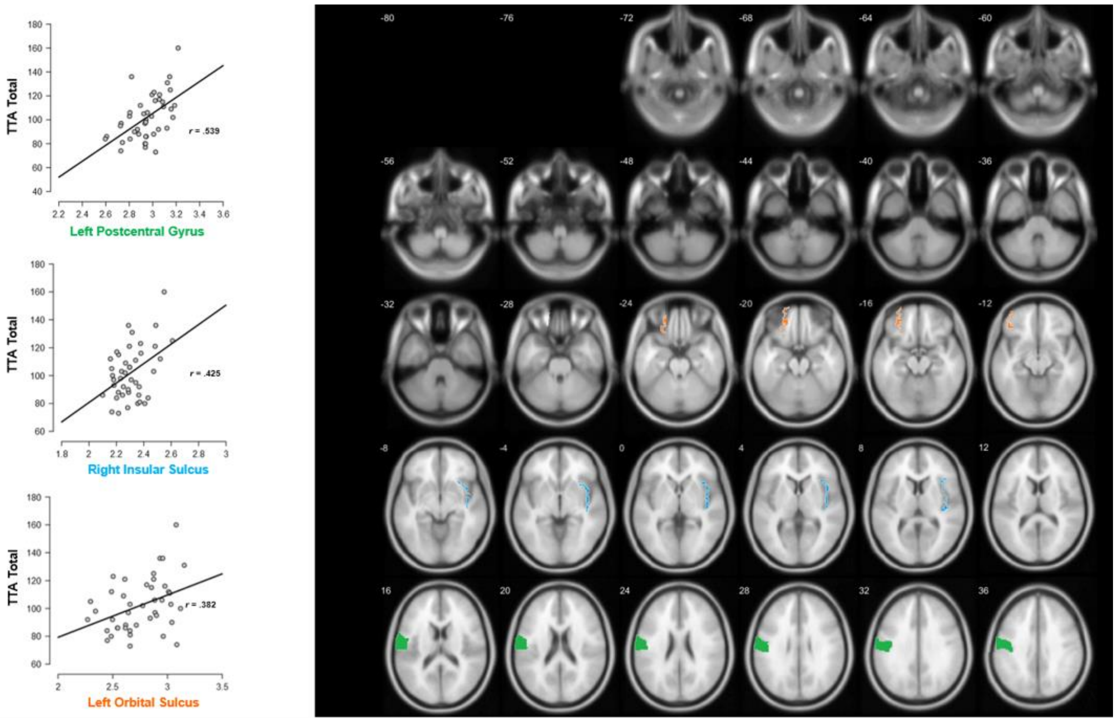

Figure 4.3. Transverse slices and scatterplots with fit lines of the correlations between the total score of the TTA and the cortical complexity of the left postcentral gyrus, right insular sulcus, and left orbital sulcus.

Table 4.3

Correlations between arithmetic fluency and typically arithmetic-associated parietal regions for the cortical complexity analyses

		L SPL	R SPL	L AG	R AG	L SPL	R SPL	L IPS	R IPS
TTA Total	*r*	-.086	-.168	.291	-.120	-.077	.229	.136	.070
	BF_{10}	0.220	0.334	1.073	0.253	0.214	0.544	0.274	0.209
	p	.581	.280	.059	.442	.140	.581	.384	.655

Note: SPL = Superior Parietal Lobe; AG = Angular Gyrus; SMG = Supramarginal Gyrus; IPS = Intraparietal Sulcus. The critical *p*-values for Bonferonni correction for multiple comparisons is $p < .0006$.

When calculating the correlations with our reading measure, strong evidence was observed for an association between reading and the left lateral orbital sulcus ($r = .471$, $BF_{10} = 26$, $p = .001$), suggesting that the associations with the left orbital sulcus are not specific to arithmetic. No substantial evidence ($BF_{10} > 3$) was found for an association with any of the other ROIs found to be associated with arithmetic fluency (left postcentral gyrus: $r = .306$, $BF_{10} = 1.309$, $p = .046$; right insular sulcus: $r = .224$, $BF_{10} = 0.520$, $p = .149$).

4.4. Discussion

Previous studies on the structural neural correlates of children's arithmetic have used VBM to highlight an association between arithmetic and grey matter volume in the superior parietal lobe, including the intraparietal sulcus, as well as other brain regions outside of the parietal cortex, such as the inferior and middle frontal gyrus, the fusiform gyrus and the hippocampus (Evans et al., 2015; Isaacs et al., 2001; Li et al., 2013; Price et al., 2016; Ranpura et al., 2013; Rotzer et al., 2008; Rykhlevskaia et al., 2009; Supekar et al., 2013). These existing studies, however, have often used research samples with wide age ranges, not fully eliminating possible maturational confounds, and were limited to VBM, which only takes the volume of brain structures into account, disregarding other structural properties. The current study studied children of a narrow age range and aimed to move beyond VBM analyses and implemented SBM analyses to correlate with 9- to 10-year-old children's arithmetic fluency, focusing on cortical complexity through fractal dimensionality. This cortical complexity looks at differences in the shape rather than the size of cortical structures, and is a sensitive metric for capturing relations between brain structure and cognitive function (King et al., 2010; Im et al., 2006; Madan and Kensinger, 2016; Mustafa et al., 2012; Sandu et al., 2014). Consequently, as general associations have previously been observed between cognitive function and cortical complexity, the current study aimed to examine if specific associations between arithmetic and the cortical complexity of regions in children's arithmetic brain network were to be found. Taken together, the current study thus aimed to provide a more comprehensive depiction of the structural neural correlates of children's arithmetic.

The results of the VBM analyses point towards positive correlations between arithmetic and the right fusiform gyrus. These results, however, must be interpreted with caution as, even though the Bayesian analysis indicated substantial evidence for an association of the fusiform gyrus with children's arithmetic, the frequentist results did not remain significant after controlling for multiple comparisons. Nevertheless, the finding aligns with previous research, as the volume of the fusiform gyrus had already been related to children's arithmetic in a study by Rykhlevskaia et al. (2009), who observed reduced volume of the fusiform gyrus for children with developmental dyscalculia in comparison to their typically developing peers. Furthermore, previous diffusion MRI research (e.g., Polspoel et al., 2018; Rykhlevskaia et al., 2009) has indicated the right ILF, which connects the occipital lobe to the anterior part of temporal lobe, including the fusiform gyrus, as a crucial white matter tract for children's arithmetic fluency, again emphasizing the importance of the fusiform gyrus within children's arithmetic (Pinheiro-Chagas, Daitch, Parvizi, & Dehaene, 2018). In fMRI research in adults, ventral visual stream areas, including the fusiform gyrus, are also found to be consistently co-activated with the intraparietal sulcus across a wide range of numerical tasks (Arsalidou and Taylor, 2011), with functional responses which increase with arithmetic complexity (Keller & Menon, 2009; Menon, 2015; Pinheiro-Chagas et al., 2018; Rickard et al., 2000; Rosenberg-Lee, Tsang, & Menon, 2009; Wu et al., 2009; Zago et al., 2001). Being part of the inferior temporal cortex, the fusiform gyrus probably plays an important role in the encoding of complex visual stimuli (i.e., orthographic processing, or the recognition and discrimination of number-letter strings; Allison, Puce, Spencer, & McCarthy, 1999; Binder, Medler, Westbury, Liebenthal, & Buchanan, 2006; Milner & Goodale, 2008). More recently, however, it has been suggested that the inferior temporal cortex may have a role of early identification of problem difficulty, beyond mere digit recognition (Pinheiro-Chagas et al., 2018). Accordingly, even though the VBM results of the current study must be interpreted with caution, this result is in line with previous research and emphasizes the importance of the right fusiform gyrus in children's arithmetic.

Next, the results of our fractal dimension analyses mainly point towards positive correlations between children's arithmetic and cortical complexity in the left postcentral gyrus. The postcentral gyrus is adjacent to the superior parietal lobe and lies in continuity with the intraparietal sulcus, whose roles within the representation and manipulation of numerical quantity and arithmetic in general have been clearly established (Menon, 2015). Consequently, cortical complexity in the postcentral gyrus might become important as the region acts as an extension of, and might affect its adjacent arithmetic-related regions. Previously, activation in the postcentral gyrus itself was mainly observed during grasping tasks (Simon et al., 2002), but postcentral activations have also been linked to the use of arithmetic strategies, such as subvocalization and finger counting, as the region corresponds to somatotopic regions responsible for lips, mouth, fingers, and hands (Kesler, Menon, & Reiss, 2006). Furthermore, the postcentral gyrus has also been discussed as being important for number tasks in children (Arsalidou et al., 2018), as, in both adults and children, activations in the right inferior parietal cortex (including the intraparietal sulcus and the postcentral gyrus) has been related to nonsymbolic numerical and spatial processing (Kaufmann et al., 2008). The current structural imaging study can only but speculate on the function of this postcentral region; further research investigating the region's exact functional role within arithmetic is necessary.

Our results also displayed positive correlations between the cortical complexity of the right insular cortex (more specifically the right insular sulcus) and children's arithmetic. Due to the stringency of our correction, this result did not remain statistically significant after controlling for multiple comparisons, yet a Bayes Factor (BF_{10}) close to 10 was observed, suggesting strong evidence for an association between both variables. Previously, structural associations between the right insula and children's arithmetic have been observed in a study by Han et al. (2013), who used deformation-based morphometry (DBM; based on the application of non-linear registration procedures to spatially normalize one brain to another one, where deformations then provide information about the type and localization of structural differences between the brains, which can be used for data analysis; Gaser, Nenadic, Buchsbaum, Hazlett, & Buchsbaum, 2001) to study anatomical variations between the brains of third graders with and without mathematical difficulties. The insula has also often been observed in studies on numerical cognition (Arsalidou and Taylor, 2011), but its exact function in arithmetic is still unclear. Over all, the insula is known for directing attentional resources and decision-making (Arsalidou and Taylor, 2011; Menon, 2015; Supekar & Menon, 2012), but has also been implicated to be important for emotional processing (Damasio et al., 2000) and speech-motor function (Fox et al., 2001). As such, the insula may be involved in intrinsically motivated behaviors (Arsalidou et al., 2018). The right insula in particular has also been identified as a key ROI in specific phobias, including mathematics anxiety (Lyons & Beilock, 2012). More specifically, these authors observed that the higher one's math anxiety is, the more increases in activity in regions such as the bilateral dorso-posterior insula (which are associated with visceral threat detection and the experience of pain itself) can be observed (Lyons & Beilock, 2012). As a result, disturbances in any one of the cognitive processes above could lead to disruptions in the normal course of procedures, and consequently interrupt the processes of solving calculation problems (Han et al., 2013).

The fractal dimensionality analyses also indicate the left orbital sulcus as being related to children's arithmetic, yet, due to this result not surviving controlling for multiple comparisons and the Bayes Factor only pointing towards substantial evidence at best, it must again be interpreted with caution. This result, however, does agree with previous research, as structural differences in the left orbitofrontal cortex were also observed in the study by Han et al. (2013) between children with and without mathematical difficulties. Similar to the insular cortex, the orbitofrontal cortex is associated with attention, decision-making, and executive function. Consequently, the function of this region is most likely not arithmetic-specific. Our data align with this because significant correlations were also observed between the fractal dimensionality of the left lateral orbital sulcus and our reading measure.

Surprisingly, no substantial evidence was observed for associations between the volume or cortical complexity of the previously reported number and arithmetic brain regions, i.e., the superior and inferior parietal lobes, with the intraparietal sulcus. Even more, the Bayes Factors (BF_{10}) for these correlational analyses were consistently below 1, pointing towards, albeit not necessarily substantial, evidence for the null hypothesis of no association between the volume or cortical complexity of these regions and the children's arithmetic fluency. This discrepancy with the available literature could be due to the wide age ranges used in previous studies, or to the implementation of different tasks for arithmetic assessment across studies. Together with these points, it is also possible that the importance of the intraparietal sulcus (at least at a structural level) mainly becomes apparent when comparing children with extremely low arithmetic fluency or developmental dyscalculia to typically developing peers, which was not the case for the present sample. Given these results, the current study does emphasize the importance of neural regions outside of the parietal cortex for children's arithmetic fluency.

Using the fractal dimensionality analyses within children's arithmetic, the current study also shines light on the implementation of structural MRI research that takes into account the shape of structures, to better capture individual differences in the organization of cortical regions. Furthermore, the current study was conducted with a research sample of only 9- to 10-year-olds, to minimize maturational confounds. Such narrow-aged studies are critical as merging data across wide age ranges, even though statistically controlled for, might lead to over-interpretations of associations between differences in grey mater volume/structure and differences in mathematical development. Accordingly, we feel it is crucial to emphasize the need for similar studies in children of different ages, such as first or second graders, who are at the beginning of their arithmetic development, or children in secondary school, who have reached a more advanced level of arithmetic. As previous research (Im et al., 2006) also observed stronger associations of cortical complexity to years of education than to IQ, indicating a possible influence of education-related development on cortical complexity, and keeping in mind (educational-based) neural plasticity, studies with a longitudinal follow-up throughout educational development are also deemed necessary to understand the direction of observed associations between cortical complexity and cognitive function, as well as to pinpoint when in development these associations start to emerge.

Finally, we would like to stress the necessity of similar research (i.e., research that moves beyond looking at volume for studying the structural correlates of arithmetic), not just across age groups, but in atypical populations, such as math-gifted children, or children with developmental dyscalculia, as this could deliver additional insights into the neural development of their mathematical skills.

CHAPTER 5

Relating Individual Differences in White Matter Pathways to Children's Arithmetic Fluency: A Spherical Deconvolution Study

Published as

Polspoel, B., Vandermosten, M., & De Smedt, B. (2018). Relating individual differences in white matter pathways to children's arithmetic fluency: A spherical deconvolution study. *Brain Structure and Function*, *224*(1), 337-350.

Abstract

Connectivity between brain regions is integral to efficient complex cognitive processing, making the study of white matter pathways in clarifying the neural mechanisms of individual differences in arithmetic abilities critical. This white matter connectivity underlying arithmetic has only been investigated through classic diffusion tensor imaging (DTI), which, due to methodological limitations, might lead to an oversimplification of the underlying anatomy. More complex non-tensor models, such as spherical deconvolution, however, allow a much more fine grained delineation of the underlying brain anatomy. Against this background, the current study is the first to use spherical deconvolution to investigate white matter tracts and their relation to individual differences in arithmetic fluency in typically developing children. Participants were 48 typically developing 9- to 10-year-olds, who were all in grade 4, and who underwent structural dMRI scanning. Theoretically relevant white matter tracts were manually delineated with a region of interest approach, after which the hindrance modulated orientational anisotropy (HMOA) index, which provides information on the structural integrity of the tract at hand, was derived for each tract. These HMOA indices were correlated with measures of arithmetic fluency, using frequentist and Bayesian approaches. Our results point towards an association between the HMOA of the right inferior longitudinal fasciculus (ILF) and individual differences in arithmetic fluency. This might reflect the efficiency with which children process Arabic numerals. Other previously found associations between white matter and individual differences in arithmetic fluency were not observed.

5.1. Introduction

Many studies have investigated the neural basis of arithmetic. Accumulating evidence in adults points towards a fronto-parietal network, including superior and inferior parietal lobes, inferior frontal gyri and the insular cortex, as being consistently activated during arithmetic (for a review, see Arsalidou & Taylor, 2011; Menon, 2015). In children, this arithmetic network involves a large set of bilateral areas, including frontal (both ventro- and dorsolateral prefrontal cortex), parietal (intraparietal sulcus, angular gyrus, and supramarginal gyrus), occipito-temporal and medial temporal (including the hippocampus) areas (Peters & De Smedt, 2018, for a review; Arsalidou et al., 2018, for a meta-analysis). Furthermore, arithmetic development is characterized by a decreasing engagement of the prefrontal cortex and by an increasing engagement and functional specialization of the inferior and posterior parietal cortex (Kucian et al., 2008; Rivera et al., 2005). This shift has been interpreted as reflecting a change in strategy use, from demanding procedural manipulations to fact retrieval; a hypothesis that has recently been confirmed at the neural level (Polspoel et al., 2017). Finally, within arithmetic, large individual differences exist at a behavioral level, which have also been established at the neural level, in both adults (Grabner et al., 2007) and children (De Smedt et al., 2011).

The functional regions of the abovementioned arithmetic brain network, however, are not adjacent, but spatially distant from one another, which makes it crucial to study the structural white matter connections between these regions. This connectivity between brain regions is integral to efficient cognitive processing (Johansen-Berg, 2010). Understanding the role of white matter pathways in arithmetic may thus further clarify the neural mechanisms of individual differences in arithmetic abilities (Matejko & Ansari, 2015).

These white matter connections can be examined with diffusion-weighted Magnetic Resonance Imaging (dMRI), an imaging technique that sensitizes the MRI signal to the diffusion (i.e., random molecular motion) of water molecules by adding diffusion encoding gradients in distinct directions to a standard MR pulse sequence (Jones & Leemans, 2011). As yet, the simplest and most frequently applied model to relate the dMRI signal to the actual underlying neuroanatomy, is Diffusion Tensor Imaging (DTI). The classic DTI model estimates the degree to which diffusion is not spherical but increased in a certain direction (fractional anisotropy or FA) per voxel. The estimated direction of diffusion per voxel is then assumed to correspond to the dominant fiber orientation, and the estimated fractional anisotropy is assumed to correspond to the density, myelination and underlying architecture of the underlying axons (Basser et al., 1994; Tournier et al., 2007).

To date, structural connectivity within arithmetic has only been investigated through classic DTI, which is subject to methodological limitations. For example, DTI can only estimate the direction of one fiber per imaging voxel, which leads to an oversimplification of the underlying anatomy in regions with multiple crossing white matter fibers (Assaf et al., 2004; Tournier et al., 2007). As many of these crossing fibers are situated in and around the parietal lobe, results of the available DTI studies must be interpreted with caution. The current study is the first to tackle these issues in the field of arithmetic, by implementing a more novel and complex non-tensor model (i.e., spherical deconvolution), which, among other things, has the asset that it can characterize the orientation of more than one fiber per voxel, and consequently provides the possibility of making more accurate statements about the associations between white matter tracts and arithmetic fluency (Dell'Acqua et al., 2007; Tournier et al., 2004).

Despite the availability of these techniques, structural connectivity research in children's arithmetic is scarce and inconclusive, as many different white matter pathways have been found to be related to individual differences in arithmetic or other mathematical skills (see Matejko & Ansari, 2015, and Moeller et al., 2015 for reviews). For example, studies on children with atypical mathematical abilities (i.e., math-gifted children or children with developmental dyscalculia) point towards higher FA in several temporo-parietal regions for math-gifted children (ages 12 to 15), including the uncinate fasciculus (UF), superior longitudinal fasciculus (SLF), and inferior longitudinal fasciculus (ILF; Navas-Sánchez et al., 2014), or a significantly lower probability of connectivity to the right inferior temporal gyrus (i.e., the right ILF and inferior fronto-occipital fasciculus; IFOF) in children with mathematical difficulties (ages 7 to 9; Rykhlevskaia et al., 2009). A study by Kucian et al. (2014) also compared a group of children with dyscalculia to a group of age matched controls (ages 8 to 11), and found lower FA in parietal and insular white matter clusters. Even though these studies on children with atypical mathematical abilities provide insights into which white matter pathways are related to arithmetic performance, making generalizations to typically developing children about the neurocognitive processes at hand might be difficult.

Looking at typically developing children, a DTI study on individual differences in children's arithmetic by Van Eimeren et al. (2008) found a correlation between children's (ages 7 to 9) scores on the numerical operations subtest of the Wechsler individual achievement test (i.e., a test of written calculation including simple arithmetic problems of all operations) and FA in the left ILF and the left corona radiata (CR). The authors speculated that the ILF could be related to participants' processing efficiency of Arabic numerals, as inferior temporal brain regions, to which the ILF connects, were previously found to be important for visual representations of numerical symbols and calculation problems (Dehaene et al., 2003; Shum et al., 2013). The relation of this temporo-parietal connections to arithmetic fluency might also reflect verbal/memory representations of numbers, as they connect the fusiform regions with temporo-parietal white matter (Catani & Mesulam, 2008). To a certain extent, these findings coincide with those of Rykhlevskaia et al. (2009), who found a correlation with the ILF in the right hemisphere, yet Van Eimeren et al. (2008) observed a correlation in the left hemisphere. These differences in hemisphere could be due to the atypical numerical processing in children with mathematical difficulties in Rykhlevskaia et al. (2009), yet, further research would need to clarify this. The left CR, on the other hand, had already been related to individual differences in reading skills (Ben-Shachar, Dougherty, &

Wandell, 2007), which led Van Eimeren et al. (2008) to suggest that its involvement in mathematical processing could become co-opted for exact, verbal mathematical skills. However, as these projection fibers connect the thalamus with the sensorimotor cortex (Catani & Thiebaut de Schotten, 2008), the notion of it having a role in perceptual and motor functions is more straightforward (Schmahmann & Pandya, 2008).

A study by Tsang et al. (2009) used DTI tractography in 10 to 15 year-old children to investigate the relationship between fronto-parietal white matter and mathematical performance (i.e., the SLF and arcuate fasciculus (AF) connect frontal and parietal regions that are typically activated during arithmetic). This was done by studying simple arithmetic facts, exact two-digit addition, and approximate two-digit addition. The study revealed an association between approximate addition and FA in the left anterior SLF, but not in the AF. However, remarks can be made on how both tracts were delineated. The AF needs to be divided into three different segments, which all connect different regions with one another and thus might have different functions (i.e., a direct, anterior and posterior segment, as proposed by Catani, Jones, & ffytche, 2005), and the SLF in its entirety contains three different parts as well (i.e., SLF I, SLF II, and SLF III; Thiebaut de Schotten et al., 2011). These distinctions were not made in Tsang et al. (2009), as the AF was considered a fronto-temporal part of the SLF. What was called the AF and SLF in the study by Tsang et al. (2009), however, seems to coincide with the direct and anterior segments of the AF, respectively, as discussed by Catani et al. (2005). A supplementary Tract-Based Spatial Statistics (TBSS) analysis in Tsang et al. (2009) also demonstrated correlations with approximate arithmetic beyond the left SLF/AF, including the bilateral SLF/AF, ILF, IFOF, CR, and corpus callosum (CC), thus suggesting that a broad network of white matter pathways is related to individual differences in children's arithmetic.

In a study by Van Beek et al. (2014), the critical segmentation of the AF into the three subcomponents was made, and an association was found between 12 year-old children's scores on addition and multiplication on a standardized timed arithmetic test (i.e. Tempo Test Arithmetic; de Vos 1992), and the FA of the anterior segment of the AF. However, the SLF as delineated by Tsang et al. (2009) and the anterior segment of the AF as delineated by Van Beek et al. (2014) coincide to a great extent. Historically speaking, the AF and the SLF have been thought to be the same tract, however, recently, attempts have been made to dissociate both tracts (e.g., Dick & Tremblay, 2012; Zhao et al., 2016). This ambiguity in defining both the SLF and the AF especially becomes apparent in comparing these two studies, as it is thus presumable that they are discussing the same tract; classic DTI analyses might not be fine-grained enough to properly disentangle both tracts (Zhao, Thiebaut de Schotten, Altarelli, Dubois, & Ramus, 2016). To resolve such issues, more novel and complex non-tensor models (e.g., spherical deconvolution) are necessary to make accurate statements about the tracts' relevance to arithmetic performance. Either way, the existing studies seem to establish the importance of fronto- and temporo-parietal white matter pathways in children's arithmetic.

These available studies on white matter pathways in children's arithmetic, however, have some major shortcomings. To begin, all of them have analyzed the diffusion data by means of classic DTI, which is subject to various methodological limitations. First, DTI is unable to resolve the orientation of multiple crossing fibers within a voxel, as it can only estimate the direction of one fiber per imaging voxel. Consequently, this leads to an oversimplification or inaccurate representation of the underlying anatomy in regions with multiple crossing white matter fibers. Using DTI, the major eigenvector in voxels with crossing fibers generally does not correspond to the actual orientation of any of the fibers (Assaf et al., 2004; Tournier et al., 2007). This is highly problematic, as the percentage of white matter voxels that contain multiple crossing fibers in the human brain is around 70-90% (Dell'Acqua et al., 2013; Farquharson et al., 2013). As many of these crossing fibers are situated in and around the parietal lobe, which is a critical region for arithmetic, and even contains multiple smaller regions with distinct functions within arithmetic (e.g., intraparietal sulcus, angular gyrus, supramarginal gyrus; Arsalidou et al., 2018; Peters and De Smedt, 2018), it is especially difficult to interpret the results from previous DTI studies on arithmetic, as tracts such as the SLF and AF that connect to these regions might have been difficult to disentangle with classic DTI.

Secondly, the interpretation of the FA index is not clear-cut, as it provides a quantitative measure per voxel that is determined by both microstructural (e.g., myelination of fibers, or size and density of cells) and macrostructural (e.g., number of crossing fibers) properties. A lot of anatomical information is thus reduced to just one index, implying that individual differences in FA could be due to a number of reasons, leading to difficulties in interpretation (Vanderauwera, Vandermosten, Dell'Acqua, Wouters, & Ghesquière, 2015). Because of these shortcomings, results of previous DTI studies should be interpreted carefully, as they might not accurately reflect the associations between white matter tracts and arithmetic.

These two methodological limitations can be resolved by using more complex non-tensor models, such as spherical deconvolution, which estimates a continuous 3D distribution of all possible fiber orientations within each voxel (Dell'Acqua et al., 2007, Tournier et al., 2004). Doing so, spherical deconvolution has the asset that it can characterize the orientation of more than one fiber per voxel, thus solving the crossing fibers problem. Furthermore, in comparison to other non-tensor models such as q-Ball imaging (Tuch, 2004), diffusion spectrum imaging (Wedeen, Hagmann, Tseng, Reese, & Weisskoff, 2005), and composite hindered and restricted model of diffusion (CHARMED; Assaf et al., 2004) imaging, which require the acquisition of higher or multiple b-values and consequently demand extended scanning sessions, spherical deconvolution has an acquisition time close to DTI, and is therefore much more suited to use with children.

The advantages of spherical deconvolution have been clearly shown in a comparative study by Farquharson et al. (2013), who pointed out that, in voxels containing two fiber populations, spherical deconvolution properly identifies both fiber populations, while DTI does not provide an orientation estimate that corresponds to either of the populations. More specifically, in their study, DTI consistently failed to identify well-known corticospinal connections extending to the majority of the sensorimotor cortex, while spherical deconvolution produced the expected fan-shaped configuration of the corticospinal fiber pathways that much more closely resemble the known anatomy. Furthermore, to extract information on microscopic properties of the white matter tracts, the hindrance modulated orientational anisotropy (HMOA) index can be derived for quantitative spherical deconvolution analyses. This index can be defined as the absolute amplitude of each lobe of the fiber orientation distribution (Dell'Acqua et al., 2013). In contrast to FA, which provides a quantitative measure per voxel indexing both micro- and macroscopic properties, HMOA is a tract-specific index, highly sensitive to changes in fiber diffusivity (e.g., myelination processes or axonal loss) and to differences in the microstructural organization of white matter (e.g., axonal diameter and fiber dispersion). The index thus provides information about microscopic properties along each fiber orientation, even in regions with fiber crossings. Accordingly, the HMOA index can detect small changes in the microstructural properties along single white matter tracts, which are not detectable with classic DTI. By applying this tract-specific index, we might thus be able to detect fiber diffusivity changes (e.g., developmental myelination processes) and can improve tractography to better map white matter complexity inside the brain (Dell'Acqua et al., 2013). Against this background, spherical deconvolution is particularly suited to overcome the limitations associated with the classic tensor model.

Another problem in many of the existing DTI studies in children's arithmetic, is that data were collected from children with wide age ranges (e.g., 7 to 10 or 10 to 15 years old). The problem here is that this period in time is characterized by large structural white matter development (e.g., Barnea-Goraly et al., 2005), and that, consequently, although statistically controlled for, the observed correlations between individual differences in arithmetic and white matter might still be swayed by maturation, instead of being purely related to mathematical achievement. As mathematical achievement also improves over child development, high experience-dependent plasticity can be expected in white matter (e.g., Casey, Tottenham, Limston, & Durston, 2006), which means that homogenous age groups (i.e., research samples with a small age range) should be studied in order to take such maturation effects into account.

To the best of our knowledge, the current study is the first to use spherical deconvolution to investigate which white matter tracts are related to differences in children's arithmetic fluency. We will also focus on children with a small chronological age range to minimize confounds of maturation. In addition, we not only implemented frequentist, but also Bayesian statistics for data analyses, as Bayesian statistics have the advantage of being able to quantify the evidence that the data provide for one hypothesis over another (Andraszewicz et al., 2015). In contrast to classic frequentist hypothesis testing, these Bayesian statistics are particularly informative when no association is observed, as they can quantify the evidence in favor of the null hypothesis of no association. These statistics are also not affected by the multiple comparison problem (Dienes, 2011). In light of the above-reviewed DTI literature, we expect to find relations between arithmetic fluency and the white matter integrity of the SLF/AF, and ILF.

5.2. Methods

5.2.1. Participants

The study started with 50 typically developing Flemish 4th graders, yet due to technical acquisition problems (n = 1) , excessive motion (n = 1), or problems during standardized testing (n = 1), data of 3 children were discarded. The remaining 47 participants (ages 9 to 10; M = 9.68, SD = 0.33; 26 boys, 21 girls; 8 left-handed) had no history of learning difficulties, or neurological or psychiatric disorders. All children were recruited via the elementary school they attended, in the vicinity of our university, and were given a financial compensation in return for participating. Written informed consent was obtained from a parent or legal guardian of each participating child. The study was approved by the Medical Ethical Committee of the University of Leuven (S59167).

All participants took part in two test sessions. During the first session, behavioral data were collected through standardized measures. This session always preceded the second session by two to three weeks (M = 19.54 days, SD = 6.34), which included the acquisition of the MRI data. This MRI acquisition session also partly contained fMRI data collection for another study, which is reported in Polspoel et al. (2017).

5.2.2. Standardized assessment

The standardized assessment session consisted of the evaluation of arithmetic, as well as intelligence, motor reaction time, and reading. First of all, the Tempo Test Arithmetic (TTA; de Vos, 1992) was used to measure the children's arithmetic competence. This is a standardized test of arithmetical fluency, similar to the Math Fluency subtest of the Woodcock-Johnson III tests of Achievement (Woodcock et al., 2003), and is very sensitive to individual differences in arithmetic fluency. The TTA is constructed of five columns of arithmetic items of increasing difficulty (i.e., one column per operation and a fifth column with mixed operations; each column starts with single digit items), for which each child gets one minute per column to write down as many correct answers as possible. Next, intellectual ability was measured by the WISC-III-NL Block Design and Vocabulary subtests, as measures of performance and verbal IQ respectively (Wechsler, 2005). To measure motor reaction time, two figures (always a circle, triangle, square, star, or heart) were presented on a computer screen. Each participant had to indicate which of both figures (i.e., left or right) was filled in white, by, as quickly as possible, pressing the corresponding key. Accuracy and reaction time were recorded for each trial (De Smedt & Boets, 2010). Finally, we measured the children's reading ability to investigate the specificity of our results (i.e., in comparison with a different symbolic skill). Reading ability was assessed using a combination of the One-Minute Test (OMT; Brus & Voeten, 1979) and the Klepel (van den Bos et al., 1994), which measure the reading of words and pseudowords, respectively; both tests consist of 116 words. For the OMT, the children get one minute to correctly read aloud as many words as possible; for the Klepel, the time limit is set to two minutes, and the children read aloud pseudowords. Other behavioral measures, such as strategy assessment, or sensitivity-to-interference, were also assessed in sub-samples of this study, but not reported here as they were linked to fMRI protocols of different studies (e.g., Polspoel et al., 2017).

5.2.3. MRI data acquisition and tractography

MRI scanning was done with a Philips Ingenia 3.0T CX MRI scanner with a SENSE 32-channel head-coil, located at the Department of Radiology of the University Hospital in Leuven, Belgium. Wash cloths were used to stabilize the children's heads and consequently minimize head motion. dMRI sagittal slices were obtained using the following parameters: 60 noncollinear directions b-value 2000 s/mm2, 30 noncollinear directions b-value 700 s/mm2 (which were eventually discarded for spherical deconvolution analyses), 6 nondiffusion-weighted images, 2.5 × 2.5 × 2.5 mm voxel size, 90º flip angle, repetition time (TR) 7000 ms, echo time (TE) 72 ms, and 240 × 125 × 240 mm field of view (approximately 12 minutes of scanning time). Anatomical T1 images were acquired with the following sequence: 0.98 × 0.98 × 1.2 mm voxel size, 256 × 256 acquisition matrix, 8º flip angle, TE 4.6 ms, 250 × 250 × 218 mm field of view (approximately 8 minutes of scanning time). As a part of data collection for different studies (e.g., Polspoel et al., 2017), the scanning session also included four fMRI runs of approximately 5 minutes, leading to a total scanning time of approximately 40 minutes.

All pre-processing was done using the Explore DTI software (Leemans, Jeurissen, Sijbers, & Jones, 2009), and existed of visual quality assurance, and rigorous motion, eddy current-induced distortion, and EPI distortion correction. After motion correction, data of participants which displayed excessive motion (n = 1), defined as a mean translation in any direction greater than the voxel size of 2.5 mm, were discarded. No normalization to a standard atlas took place. Whole-brain DTI tractography was performed with FA-threshold = 0.20, maximum turning angle = 30º, and step length between calculations = 1 mm. For the spherical deconvolution analyses, additional processing was done with the StarTrack software (Dell'Acqua et al., 2013), using the following parameters: iterations = 200, n = 0.04, and r = 8. Finally, the following parameters were used for the spherical deconvolution whole-brain tractography: absolute HMOA threshold = 0.06, relative HMOA threshold = 5%, maximum turning angle = 30º, and step length between calculations = 1 mm.

Tractography of the white matter tracts was performed with the TrackVis software (Wang, Benner, Sorensen, & Wedeen, 2007). All tracts were manually delineated for each subject using a region of interest (ROI) approach, based on anatomical landmarks in color-coded maps (Catani & Thiebaut de Schotten, 2008; Thiebaut de Schotten et al., 2011; Wakana et al., 2007). In this approach, each ROI represents an obligatory passage for the tract at hand. The colors in these maps refer to the direction the fibers run in; red are commissural fibers, green are associative fibers, and blue are projection fibers.

The white matter pathways that were investigated in the current study were based on the existing literature (Matejko & Ansari, 2015), as well as the ROIs used to perform the manual segmentation of each tract (Catani & Thiebaut de Schotten, 2008; Thiebaut de Schotten et al., 2011; Wakana et al., 2007). An overview of the connections of each tract and the ROIs used for delineation, can be found in Table 5.1. The same ROIs were used for both the spherical deconvolution and DTI delineations, in order to maximize the comparability of the HMOA and FA metrics, respectively. A visual overview of the tracts under study, delineated with spherical deconvolution, can be found in Figure 5.1. For comparison, the same tracts for the same participant, but delineated with classic DTI, can be found in Appendix (Figure 5.A1).

After manually delineating a white matter pathway, the TrackVis software offers statistical information of the tract at hand (e.g., HMOA value, amount of fibers, volume etc.), which can then be used for statistical testing.

5.2.4. Statistical analyses

All statistical analyses were performed with the JASP software package (JASP Team, 2017). We calculated Pearson correlations between the results of the TTA (i.e., of all columns separately and of the total score) and the HMOA values of all tracts under study, and their corresponding Bayes Factors. Since our sample had a small age range, and all participants were part of the same norm group, raw scores of the TTA (i.e., number of correctly solved items within one minute) were used for the correlation analyses. Only tracts with a minimum of 20 fibers were used for data analyses. The Bayesian approach was implemented, as it has the advantage to quantify the evidence that data provide for one hypothesis over another (Andraszewicz et al., 2015). Accordingly, Bayes factors of 1, 1-3, 3-10, 10-30, 30-100, or > 100 respectively point towards no, anecdotal, substantial, strong, very strong, or decisive evidence for the alternative hypothesis (Jeffreys, 1961). Additionally, Bayesian analyses allow us to verify the extent to which the data are in favor of the null hypothesis of no association (i.e., Bayes factors of 1-1/3, 1/3-1/10, 1/10-1/30, 1/30-1/100, < 1/100 respectively point towards anecdotal, substantial, strong, very strong, or decisive evidence for the null hypothesis; Jeffreys, 1961). These statistics are also not affected by the multiple comparison problem (Dienes, 2011). Results of frequentist approaches to statistical testing are also reported, implementing the Bonferroni method of controlling for multiple comparisons (p = Target Alpha Level / number of delineated tracts; $p = 0.05/25 = 0.002$). In order to control for other variables such as IQ, and motor reaction time, partial correlations were calculated with IQ and motor reaction time simultaneously added as control variables. Correlations were also calculated with our reading measure as to assure that any observed associations between the tracts and arithmetic fluency, are not observed with another symbolic skill, measured in a similar (i.e., time-limited) way, thus testing the specificity of the results. For the sake of comparison, all analyses were also conducted with the FA values when implementing classic DTI to analyze the neural data.

Table 5.1

Overview of connections and regions of interest (ROI) for each tract under study

Tract	**Connections**	**ROIs**
Inferior fronto-occipital fasciculus	Occipital cortex to frontal lobe, through deep temporo-basal areas and insula (Martino, Brogna, Robles, Vergani, & Duffau, 2010)	ROI 1: Coronal slice at anterior edge of the genu ROI 2: Occipital lobe on coronal slice at middle point between posterior edge of the cingulum and posterior edge of parieto-occipital fissure
Inferior longitudinal fasciculus	Occipital lobe to anterior part of temporal lobe, including fusiform gyri and parahippocampal regions (Catani et al., 2005)	ROI 1: Coronal slice at posterior edge of cingulum ROI 2: Entire temporal lobe at most posterior coronal slice where temporal lobe is separated from frontal lobe
Arcuate fasciculus	Perisylvian regions of frontal, parietal, and temporal lobes with each other – separated into a direct segment (ROIs 1 and 3), an anterior segment (ROIs 1 and 2, without 3) and a posterior segment (ROIs 3 and 4, without 1) (Catani et al., 2005)	ROI 1: Arch-shaped dorsal ROI on coronal slice at middle of posterior limb of internal capsule ROI 2: Association fibers on coronal slice at middle of splenium ROI 3: Lateral posterior ROI on axial slice at level of anterior commissure ROI 4: Lateral posterior ROI on axial slice, similar to ROI 3, yet 5 to 7 slices more superior
Superior longitudinal fasciculus	Large parieto-frontal connections, separated into a dorsal superior (SLF1; ROIs 1 and 2), middle (SLF2; ROIs 1 and 3), and ventral part (SLF3; ROIs 1 and 4) (Thiebaut de Schotten et al., 2011)	ROI 1: Entire parietal lobe on coronal slice at level of posterior commissure ROI 2: Superior frontal gyrus on coronal slice at level of anterior commissure ROI 3: Middle frontal gyrus on coronal slice at level of anterior commissure ROI 4: Precentral gyrus on coronal slice at level of anterior commissure
Uncinate fasciculus	Lateral orbitofrontal cortex to anterior temporal lobe (Von Der Heide et al., 2013)	ROI 1: Entire temporal lobe at most posterior coronal slice where temporal lobe is separated from frontal lobe ROI 2: All projections towards frontal lobe in the same slice as ROI 1
Corona radiata & corticospinal tract	Projection fibers, carrying neural traffic to and from cerebral cortex. (CST – specifically to and from primary motor cortex; Han et al., 2010)	ROI 1: Entire cerebral peduncle on axial level of decussation of superior cerebellar peduncle ROI 2 (CST): ROI around bundle of trajectories that reach primary motor cortex
Corpus callosum	Largest of commissural fibers, linking cerebral cortex of left and right hemisphere (Wakana et al., 2007)	*Forceps major* ROI 1 & 2: Coronal slice at most posterior edge of parieto-occipital fissure (bilaterally). *Forceps minor* ROI 1 & 2: Coronal slice in the middle of anterior edge of frontal cortex and genu (bilaterally).

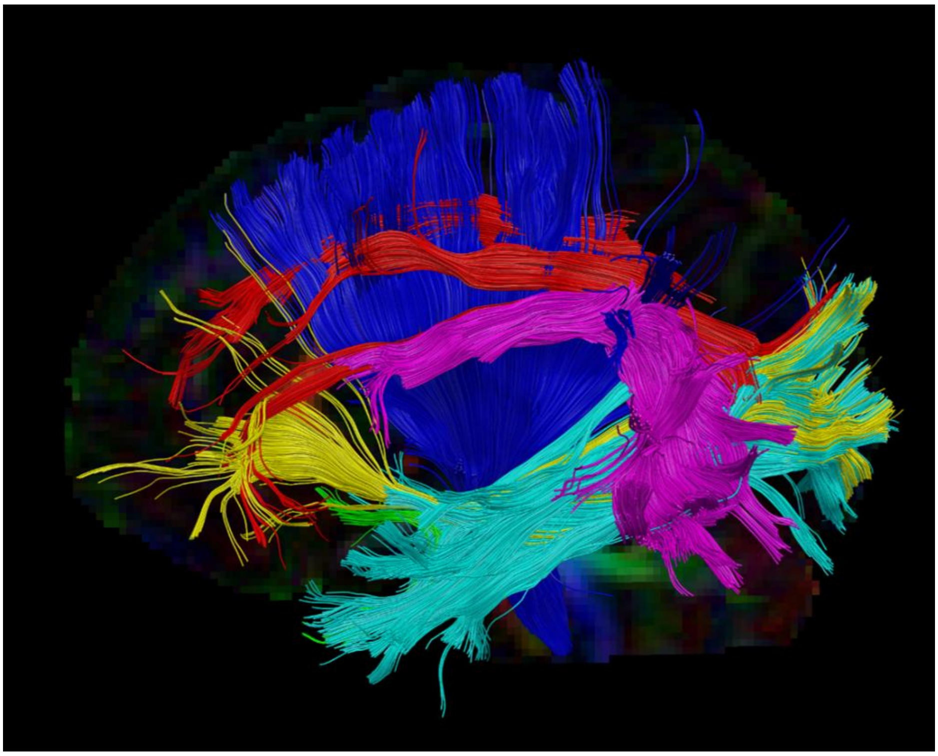

Figure 5.1. Overview of the white matter pathways under study, delineated with spherical deconvolution: in red the SLF I, SLF II, and SLF III; in fuchsia the AF; in yellow the IFOF; in cyan the ILF; in green the UF; in blue the CR.

Note: For purpose of clarity, the corpus callosum is not depicted in this image, yet it was also under study.

5.3. Results

5.3.1. Behavioral results

Figure 5.2 displays box plots with the descriptive statistics of arithmetic, intellectual ability, motor reaction time, and reading. The means of our sample were close to the expected population averages, and show proper variation. An important note is that even though the minimum score for some of the tasks was low, none of the participating children had been diagnosed with any type of learning disorder or intellectual disability.

5.3.2. Correlations with white matter integrity

Pearson correlations, using both Bayesian and frequentist approaches, were calculated between the HMOA values of the various tracts and participants' scores on the TTA. Table 5.2 summarizes the main results of the spherical deconvolution analyses and presents positive correlations for which we observed at least substantial evidence in favor of the hypothesis of an association between HMOA of a given tract and individual differences in TTA ($BF_{10} > 3$; see Jeffreys, 1961 for an interpretation of Bayes Factors), and for which a significant correlation using frequentist statistics ($p < .05$) was found.

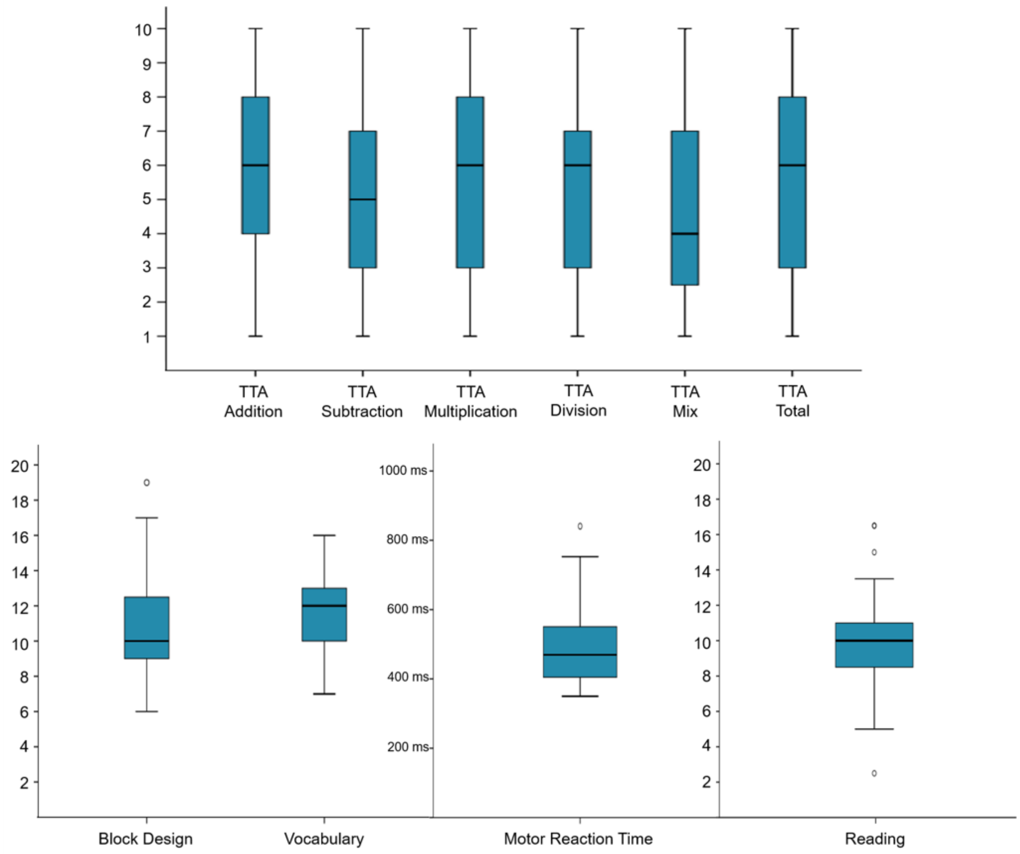

Figure 5.2. Box plots displaying performance on arithmetic, intellectual ability, motor reaction time, and reading.

Note: The scores for arithmetic, intellectual ability, and reading are standardized scores. The scores on the arithmetic test are standardized as M = 5, SD = 2, with a maximum of 10. The scores on the intelligence and reading tests are standardized as M = 10, SD = 3, with a maximum of 19. The scores for motor reaction time are raw scores displaying the average reaction time.

Correlations were found between the HMOA values of the right ILF and arithmetic performance across all operations. The Bayesian analyses indicated that the evidence for these associations was strong for addition and multiplication, very strong for division, and decisive for subtraction, the mixed column and the total score of the TTA. Using the frequentist approach, all correlations, except for addition, remained significant when using a Bonferroni method of controlling for multiple comparisons. The results also remained significant when simultaneously controlling for IQ, and motor reaction time. We also observed an association between the HMOA values of the right UF and subtraction, for which the evidence was substantial. Using the frequentist approach, significant correlations were also found for division, the mixed column and the total score of the TTA. These significant correlations, however, did not survive a Bonferroni method of controlling for multiple comparisons. The correlation for subtraction was also the only one that remained significant when controlling for IQ and motor reaction time.

Table 5.2

Correlations between the HMOA (spherical deconvolution) and FA (DTI) values of the right ILF and right UF, and participants' scores on the TTA for each operation and in total

		TTA addition		**TTA subtraction**		**TTA multiplication**		**TTA division**		**TTA mix**		**TTA total**	
		HMOA	**FA**	**HMOA**	**FA**	**HMOA**	**FA**	**HMOA**	**FA**	**HMOA**	**FA**	**HMOA**	**FA**
Right ILF	Pearson's *r*	.395	.249	.570	.451	.423	.370	.430	.252	.481	.404	.528	.391
	BF_{10}	14.25	1.37	1771.5	49.21	25.63	8.62	30.04	1.42	105.52	16.9	414.11	13.1
	p-value	.003	.046	< .001	<.001	.002	.005	0.001	.044	< .001	.002	< .001	.003
Right UF	Pearson's *r*	.096	-.008	.353	.274	.063	.076	.255	.233	.277	.249	.254	.204
	BF_{10}	0.327	0.174	6.419	1.902	0.261	0.285	1.476	1.144	1.971	1.370	1.468	0.830
	p-value	.260	.523	.007	.031	.337	.305	.042	.057	.030	.046	.042	.084

Note: The Bayes Factors (BF_{10}) report the amount of evidence found for a positive association between the tracts and the TTA. Such evidence is considered substantial, strong, very strong, or decisive for the alternate hypothesis if the Bayes factor is above 3, 10, 30, or 100, respectively, or for the null hypothesis if the Bayes factor is below 1/3, 1/10, 1/30, or 1/100, respectively.

Visual representations of the right ILF and right UF can be found in Figure 5.3. Scatterplots with fit lines of the associations between the HMOA values of these tracts and performance can be found in Figure 5.4. To test the specificity of these results to arithmetic fluency, we checked whether these associations could also be observed with a different symbolic measure (i.e., reading, calculated as the average score on both OMT and Klepel tests). This was not the case, as no significant correlations were observed between reading with either the right ILF (r = .213; p = .075; BF_{10} = 0.914) or right UF (r = .183; p = .109; BF_{10} = 0.673).

No evidence was found for associations between scores on the TTA and the HMOA values of any of the segments of either SLF or AF, even though this was hypothesized based on the available literature. The Bayes factors (BF_{10}) for almost all of these correlations were below 1, indicating that the null hypothesis of no association between arithmetic fluency and these tracts is more likely than the existence of an association. This evidence for the null hypothesis, however, was not substantial (i.e., BF_{10} < 1/3) across all segments and operations, but was often anecdotal (i.e., 1 < BF_{10} < 1/3; Jeffreys, 1961; see Appendix - Table 5.A1 for a more detailed overview).

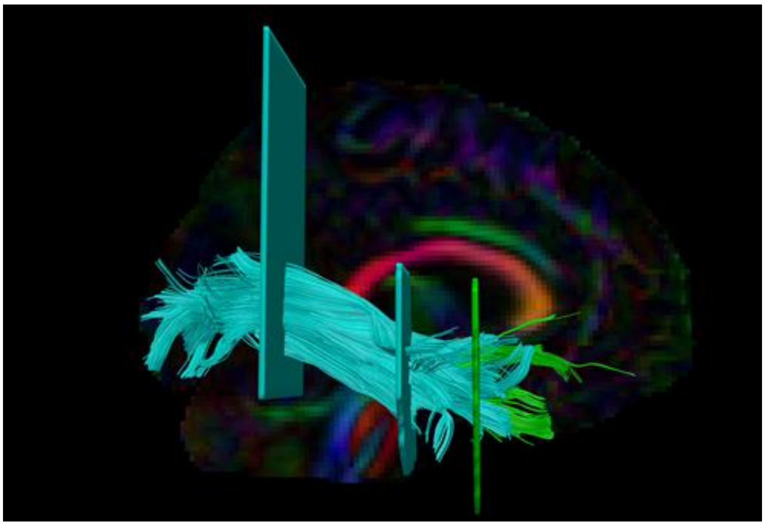

Figure 5.3. Visual representations of the white matter tracts found to be associated with individual differences in arithmetic fluency.

Note: Cyan: Right-hemispheric ILF delineated on a color-coded map; the ILF is delineated by an entire coronal slice at the posterior edge of the cingulum, and an ROI entailing the entire temporal lobe on the most posterior coronal slice in which the temporal lobe is not connected to the frontal lobe. Green: Right-hemispheric UF; the UF is delineated by an ROI on the entire temporal lobe at the most posterior coronal slice where the temporal lobe is separated from the frontal lobe, and a second ROI that includes all projections towards the frontal lobe in the same slice as the first ROI.

Finally, the associations between the tracts under study and individual differences in arithmetic fluency were also analyzed by using classic DTI metrics (i.e., the FA index). These analyses yielded similar results for the right ILF (see Table 5.2). The spherical deconvolution analyses, being more specific to white matter properties, also displayed stronger associations between arithmetic fluency and the right ILF than the DTI analyses. With DTI, only anecdotal evidence was found for associations between FA and addition and division for the right ILF. Using the frequentist approach, significant results were found across operations, yet only the results for subtraction survived controlling for multiple comparisons. For the right UF, which did display significant associations with arithmetic fluency when using spherical deconvolution, only anecdotal evidence was found when using classic DTI. Significant p-values were observed for the associations between FA and subtraction and the mixed column, yet these results did not survive when controlling for multiple comparisons. All significant correlations, however, stayed significant when simultaneously controlling for IQ and motor reaction time. As with the spherical deconvolution analyses, no evidence for an association with any of the other tracts was found. Furthermore, using classic DTI, parts of the SLF were only traceable in 5 out of 47 participants, due crossing fibers with either CR, AF, or corpus callosum. This made it impossible to examine the correlations between FA in the SLF and arithmetic (Appendix – Table 5.A1).

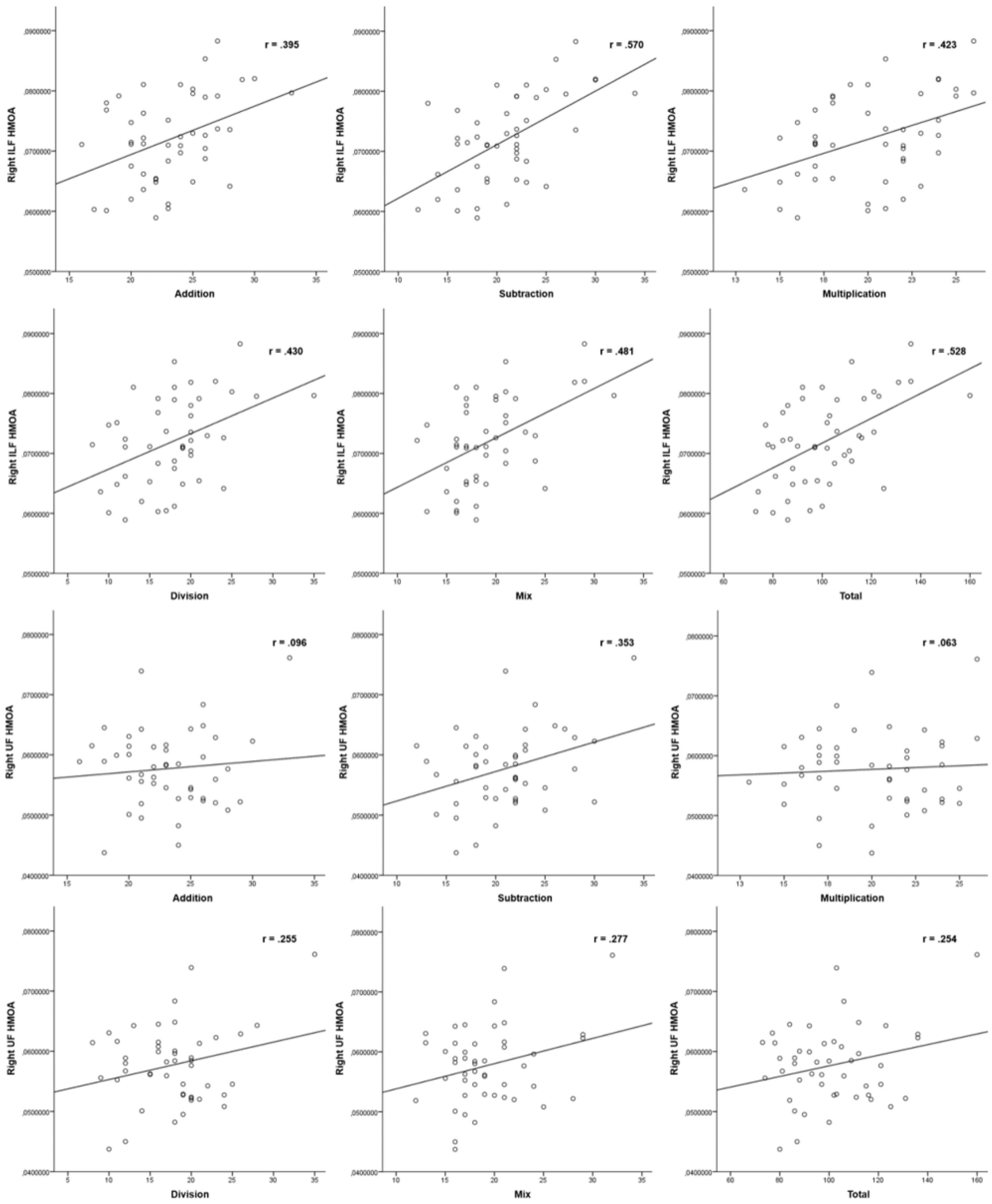

Figure 5.4. Scatterplots with fit lines of the associations between all the columns and the total score of the TTA, and the HMOA values of the right ILF and right UF.

5.4. Discussion

When it comes to arithmetic fluency, contemporary DTI research points to a variety of tracts (i.e., the AF, SLF, ILF, IFOF, UF, and others) as being related to children's arithmetic performance (Matejko & Ansari, 2015). However, the existing studies all applied classic DTI to study various white matter tracts, a method which is subject to methodological limitations (e.g., Assaf et al., 2004; Dell'Acqua et al., 2013;

Farquharson et al., 2013; Tournier et al., 2007) such as the fact that DTI can only estimate the direction of one fiber per imaging voxel, leading to an oversimplification or inaccurate representation of the underlying anatomy. Furthermore, these studies were all conducted in research samples with wide age ranges (e.g., 7 to 10 or 10 to 15 years old; Matejko & Ansari, 2015), which, even though statistically controlled for, might lead to maturational confounds and consequently to the over-interpretation of associations between differences in connectivity and differences in mathematical development. As a result, not that much is known about the actual relation between structural white matter tracts and children's arithmetic. These limitations, however, can be tackled by implementing spherical deconvolution, which can characterize the orientation of more than one fiber per voxel, and from which the HMOA index, which is tract-specific and provides information about the diffusion properties along each fiber orientation (Dell'Acqua et al., 2013), can be derived. The current study was the first to implement spherical deconvolution to investigate the associations of white matter tracts and individual differences in typically developing children's arithmetic fluency, and focused on children with a small age range (9- to 10-year-olds), as to minimize confounds of maturation.

Our results primarily point towards an association of the right ILF and individual differences in children's arithmetic fluency. The current findings echo previous results with classic DTI in which associations between FA in the ILF and individual differences in mathematics have been observed (Li et al., 2013; Navas-Sánchez et al., 2014; Rykhlevskaia et al., 2009; Van Eimeren et al., 2008). For example, correlations were found between FA in the left ILF and the arithmetic subtest of the WISC in 9-11 year-olds (Li et al., 2013) and the numerical operations subtest of the WIAT in 7-9 year-olds (Van Eimeren et al., 2008), albeit in the left hemisphere instead of the right. Furthermore, group differences were found in the FA of the bilateral ILF between controls and math-gifted children, with math-gifted children having higher FA values (Navas-Sánchez et al., 2014) and in the right ILF between controls and children with dyscalculia, with the control group having higher FA values (Rykhlevskaia et al., 2009). The current study in typically developing children thus found similar results, as correlations were found between the white matter integrity of the ILF and children's arithmetic fluency. Our findings, however, go beyond the existing literature, by focusing on a narrow age range, and by implementing a more reliable method of analyzing the diffusion data (i.e., spherical deconvolution, with the associated HMOA index).

What could this association between the ILF and arithmetic fluency potentially reflect? The ILF connects the occipital lobe to the anterior part of temporal lobe, including the fusiform gyri and parahippocampal regions. In arithmetic, this tract might be related to the efficiency with which children process Arabic numerals, as research has shown that inferior temporal regions are involved in the processing of visual representations of numerical symbols (Dehaene et al., 2003; Shum et al., 2013). More recently, increased activation in the inferior temporal gyrus has even been observed to be driven by broader mathematical processing, instead of a specific preference to Arabic numbers (Grotheer et al., 2018), which might also explain the ILF's relevance within arithmetic. The ILF also mediates the interaction between medial, inferior and anterior temporal cortices with Perisylvian areas, and is thus related to language (Catani & Mesulam, 2008). As such, it is possible that the ILF is important for exact verbal arithmetic skills (e.g., fact retrieval), as it could subserve as a first step in connecting the lingual, fusiform, and parahippocampal regions in the ventral visual stream, upwards to the dorsal visual stream and thus, possibly via other tracts such as the AF or SLF, eventually to the IPS and superior parietal lobe (Rykhlevskaia et al., 2009). Furthermore, recent fMRI research in children has indicated middle temporal regions as being increasingly activated during fact retrieval, which coincide with the anatomical location of the ILF (Polspoel et al., 2017). The function of the ILF could thus go beyond fluency in processing numerical symbols, but it may also be relevant for general arithmetic fluency. Further research is thus needed to define the exact role of the ILF within arithmetic fluency.

Correlations were also found between arithmetic – yet mainly in subtraction – and the HMOA values of the right UF. This result concurs with previous research, in which increased FA in the right UF was observed when using TBSS to compare math-gifted children to controls (ages 12 to 15; Navas-Sánchez et al., 2014). The exact function of the UF within arithmetic, however, is still unclear. The main role of the UF, which connects the lateral orbitofrontal cortex with the anterior temporal lobes, seems to lie within temporal lobe-based mnemonic associations (Von Der Heide, Skipper, Klobusicky, & Olson, 2013). As such, the UF could have an assisting role within memory, which could also explain the tract's role within arithmetic. Recently, a meta-analysis in children's arithmetic also highlighted the importance of the right insular cortex (i.e., a locus of the right UF) within calculation (Arsalidou et al., 2018), thus supporting the possibility of importance for this tract within children's arithmetic. A recent fMRI study on the neural differences between fact retrieval and procedural strategy use even points towards relevance for temporal regions and the orbitofrontal cortex during fact retrieval (Polspoel et al., 2017). It is thus plausible that the UF maintains this connection and is consequently important for more automated processes within arithmetic. The association between UF and individual differences in arithmetic, on the other hand, needs to be interpreted with caution, as the Bayesian analyses indicated that the evidence for this association was mainly anecdotal.

It is important to point out that none of the other previously found associations between white matter tracts (i.e., CC, CR, CST, AF, SLF, or IFOF; Matejko & Ansari, 2015) and mathematical abilities were found in the present study. The Bayesian statistics implemented in the current study even pointed to towards, albeit not always substantial, evidence for the null hypothesis of no association between arithmetic and the AF and SLF, even though these tracts were found to be related to children's arithmetic in previous studies (e.g., Tsang et al., 2009; Van Beek et al., 2014).

The absence of associations with these tracts in our results can be explained by various factors. First, in contrast to the current study in which we tried to minimize confounds of maturation, the available studies have been conducted in typically developing samples of children with wider age ranges than those of the current study (e.g., 7 to 9 years old; Van Eimeren et al., 2008; 10 to 15 years old; Tsang et al., 2009), leaving it unresolved to which extent found associations are due to maturation or to actual individual differences in performance. Second, the absence of a correlation with some of the previously observed tracts in the current study might also be due to the specificity of the tasks under study. The current study used results on the TTA (i.e., a timed arithmetic task based on fluency, with basic arithmetic items across all operations) for correlations with white matter indices. Some of the previous studies, however, did not focus on arithmetic fluency, but implemented a broader and/or untimed assessment of mathematical abilities. For example, Van Eimeren et al. (2008) assessed mathematics through the numerical operations and mathematical reasoning subtests of the WIAT-II, and in Navas-Sánchez et al. (2014), the math-gifted children were selected based on their enrolment in a program for mathematically talented children, and not on their arithmetic skills. This could also explain differences in results, as the tracts found in the current study might be specific to arithmetic fluency, but might not be found for individual differences in other mathematical skills, such as mathematical reasoning. Finally, all of the existing dMRI literature in the field of arithmetic has implemented classic DTI, which, as mentioned, has some methodological constraints (Assaf et al., 2004; Dell'Acqua et al., 2013; Farquharson et al., 2013; Tournier et al., 2007). Even though, as the results of the current study point out, DTI and spherical deconvolution analyses lead to similar results, the spherical deconvolution analyses are more accurate and provide stronger results. Consequently, it is possible that, in combination with a wider age range and with the use of different mathematical tasks, the use of classic DTI in previous studies might have led to the analyses not being powerful enough to consistently detect relationships with the right ILF, or to observing relationships with other tracts such as the AF or SLF.

The associations of the right ILF with arithmetic fluency were also observed across operations, even though in previous research, neural differences between operations have been observed, both functionally (e.g., De Smedt et al., 2011; Prado et al., 2011) and structurally (Van Beek et al., 2014). These neural differences, however, were most likely due to differences in the arithmetic strategies used to solve items of a certain operation (Polspoel et al., 2017; Prado et al., 2011). As the TTA largely consists out of single-digit items, the likelihood of a fact retrieval strategy in the children under study occurring across all operations was large, which might explain the fact that no clear operation differences were observed. However, our measure of arithmetic did not allow us to analyze the used strategies for each operation, as the problems were not selected to specifically elicit one strategy or the other, and we did not ask children to report on their strategy use, thus making it hard to form conclusions on this issue. We contend that future research implements more carefully designed tests of each operation (i.e., targeting particular strategies) or the collection of strategy data through verbal self-reports as an alternative avenue. Such research might also aid in clearly defining the role of the ILF within arithmetic.

In line with previous research (e.g., Farquharson et al., 2013), the current study shines light on dMRI research that goes beyond classic DTI, and implements more complex non-tensor models such as spherical deconvolution when studying arithmetic ability. Furthermore, it needs mentioning that the existing studies on white matter involvement in children's arithmetic often collected data from children with wide age ranges (e.g., 7 to 10 or 10 to 15 years old; Matejko & Ansari, 2015), which is a time period characterized by large white matter development (e.g., Barnea-Goraly et al., 2005). Consequently, merging data across wide age ranges, even though statistically controlled for, might lead to the over-interpretation of associations between differences in connectivity and differences in mathematical development, as results might still be affected by maturational confounds. To take this problem into account, the current study was conducted with a research sample of only 9 to 10 year-olds (i.e., fourth graders). Accordingly, it is crucial to emphasize the need for similar studies in children of different ages, such as first or second graders (i.e., children who are at the beginning of their arithmetic development) or children in secondary school. Studies with a longitudinal follow-up throughout development are also deemed necessary.

It is also important to emphasize that learning arithmetic does not occur in isolation, but that it is highly dependent on the general educational environment in which these skills evolve, as well as the emphasis on automatization processes within the mathematics curriculum (De Smedt, 2016). For example, a comparison of the fact retrieval frequencies in single-digit addition and subtraction in American (Geary et al., 2004) and Belgian (Torbeyns et al., 2004) third-graders revealed a relative retrieval frequency of 38% and 88%, respectively. As all participants of the current study came from Belgian elementary schools, high automatization skills were to be expected. In accordance, it is plausible that studies across cultures with differences in the emphasis on such automatization processes might point towards the involvement of different white matter tracts for the same set of arithmetic items as were found in our sample.

Finally, we would like to emphasize the necessity of similar research, not just across age groups or different cultures, but in atypical populations, such as math-gifted children, or children with developmental dyscalculia, as this could deliver additional insights into the neural development of these mathematical skills.

Alongside the abovementioned existing studies, the current study implemented a structural approach to connectivity (i.e., dMRI), leaving the possibility open of studying connections between neural regions involved in children's arithmetic in a functional manner. Functional connectivity analyses use fMRI to study consistent signal changes in anatomically distant regions, yet only a very limited number of such studies exist within arithmetic (Peters & De Smedt, 2018). In all, we feel that these suggestions yield the possibility of providing a fruitful contribution to the emerging field of educational neuroscience.

5.5. Appendix

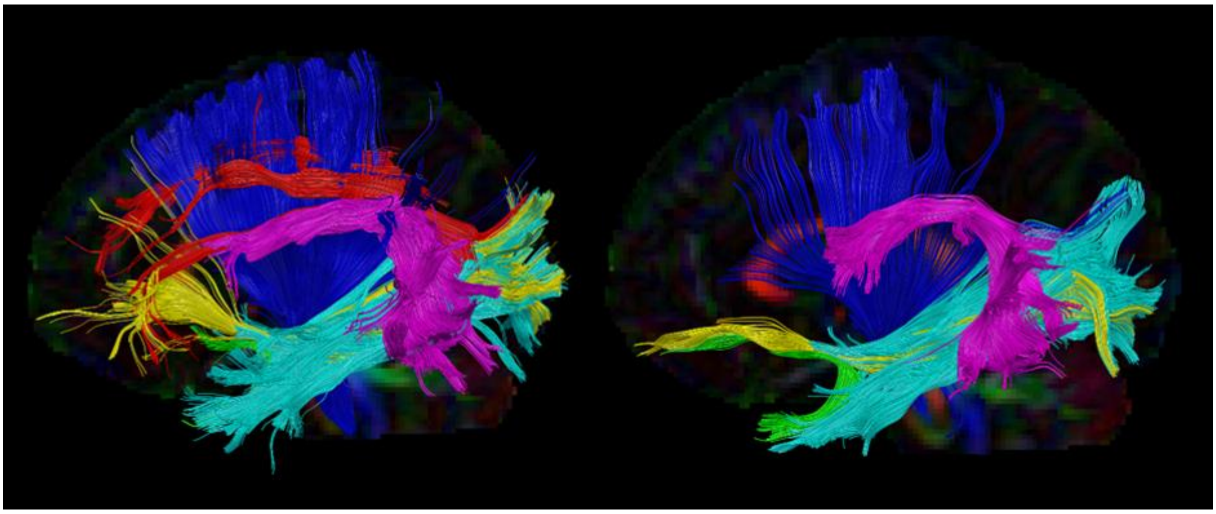

Figure 5.A1. Overview of the white matter pathways under study, delineated with spherical deconvolution (left) and classic DTI (right): in red the SLF I, SLF II, and SLF III; in fuchsia the AF; in yellow the IFOF; in cyan the ILF; in green the UF; in blue the CR.

Note: The tracts for both images were delineated on the same participant. Using classic DTI, the SLF is often not traceable, as was the case for this participant. For purpose of clarity, the corpus callosum is not depicted in this image, yet it was also under study.

Table 5.A1

Results of correlations between the HMOA values (spherical deconvolution) and FA values (DTI) of the AF, SLF, IFOF, CST, CR, and CC, and participants' scores on the TTA for each column and in total

		TTA addition		**TTA subtraction**		**TTA multiplication**		**TTA division**		**TTA mix**		**TTA total**	
		HMOA	**FA**	**HMOA**	**FA**	**HMOA**	**FA**	**HMOA**	**FA**	**HMOA**	**FA**	**HMOA**	**FA**
Left AFant	Pearson's *r*	-.048	-0.152	.138	-.059	.043	-.030	.110	.078	-.015	.013	.061	-.023
	BF_{10}	0.144	0.101	0.450	0.143	0.231	0.163	0.361	0.293	0.168	0.201	0.258	0.168
	p-value	.625	0.326	.178	.706	.386	.847	.231	.614	.541	.932	.342	.880
Right AFant	Pearson's *r*	.053	-0.001	.264	.148	.128	-.115	.121	.129	.150	.055	.168	.065
	BF_{10}	0.246	0.181	1.670	0.489	0.415	0.109	0.394	0.419	0.499	0.248	0.586	0.266
	p-value	.361	0.996	.036	.322	.196	.441	.209	.387	.157	.716	.129	.662
Left AFdir	Pearson's *r*	-.009	-.029	.087	.239	-.071	-.037	.032	.118	.073	.150	.033	.116
	BF_{10}	0.174	0.157	0.307	1.228	0.130	0.152	0.217	0.385	0.278	0.500	0.218	0.378
	p-value	.523	.845	.280	.105	.682	.807	.415	.429	.314	.313	.412	.438
Right AFdir	Pearson's *r*	.049	-.049	.205	.177	-.059	.032	.204	.257	.165	.090	.146	.136
	BF_{10}	0.240	0.161	0.837	0.596	0.137	0.233	0.833	1.251	0.569	0.323	0.482	0.438
	p-value	.372	.768	.084	.281	.652	.849	.084	.114	.134	.586	.163	.408
Left AFpost	Pearson's *r*	-.087	-.079	.035	.137	-.007	.093	.130	.167	.053	.173	.040	.123
	BF_{10}	0.122	0.126	0.221	0.445	0.175	0.320	0.423	0.579	0.245	0.609	0.227	0.399
	p-value	.720	.599	.407	.360	.520	.534	.191	.262	.363	.246	.395	.411
Right AFpost	Pearson's *r*	-.069	.041	.007	.240	-.025	.185	-.038	.260	-.025	.258	-.032	.234
	BF_{10}	0.132	0.228	0.189	1.231	0.160	0.684	0.151	1.573	0.161	1.545	0.155	1.153
	p-value	.676	.785	.480	.105	.567	.214	.601	.078	.565	.080	.586	.113
Left SLF1	Pearson's *r*	.080	/	.151	/	.236	/	.162	/	.262	/	.202	/
	BF_{10}	0.293	/	0.505	/	1.174	/	0.551	/	1.628	/	0.810	/
	p-value	.296	/	.155	/	.055	/	.139	/	.037	/	.087	/
Right SLF1	Pearson's *r*	-.006	/	.072	/	-.083	/	-.050	/	.051	/	-.001	/
	BF_{10}	0.176	/	0.277	/	0.124	/	0.143	/	0.242	/	0.181	/
	p-value	.516	/	.315	/	.710	/	.631	/	.368	/	.502	/
Left SLF2	Pearson's *r*	-.048	/	-.017	/	.036	/	.034	/	.044	/	.012	/
	BF_{10}	0.144	/	0.167	/	0.222	/	0.219	/	0.232	/	0.194	/
	p-value	.626	/	.545	/	.404	/	.411	/	.386	/	.467	/
Right SLF2	Pearson's *r*	-.196	/	-.133	/	-.169	/	.031	/	-.114	/	-.119	/
	BF_{10}	0.083	/	0.102	/	0.091	/	0.216	/	0.110	/	0.108	/
	p-value	.906	/	.814	/	.871	/	.417	/	.777	/	.787	/
Left SLF3	Pearson's *r*	-.044	/	.188	/	-.073	/	.040	/	-.065	/	.022	/
	BF_{10}	0.147	/	0.706	/	0.129	/	0.227	/	0.133	/	0.205	/
	p-value	.615	/	.103	/	.687	/	.394	/	.669	/	.441	/
Right SLF3	Pearson's *r*	-.081	/	.186	/	.017	/	-.027	/	.058	/	.039	/
	BF_{10}	0.125	/	0.689	/	0.199	/	0.159	/	0.253	/	0.226	/
	p-value	.707	/	.106	/	.456	/	.572	/	.350	/	.397	/
Left IFOF	Pearson's *r*	-.068	-.091	.213	.217	.012	-.079	.069	.082	.111	.135	.087	.077
	BF_{10}	0.132	0.120	0.910	0.947	0.194	.0126	0.271	0.296	0.363	0.440	0.307	0.286
	p-value	.649	.542	.151	.144	.935	.596	.647	.585	.459	.365	.560	.609

Right IFOF	Pearson's *r*	-.066	-.053	.130	.160	-.031	.037	-.110	.141	-.062	.082	-.031	.095
	BF_{10}	0.133	0.141	0.422	0.543	0.156	0.223	0.112	0.461	0.136	0.296	0.156	0.324
	p-value	.661	.722	.384	.283	.837	.806	.463	.345	.681	.585	.837	.525
Left CST	Pearson's *r*	.114	-.018	.150	.081	.097	.020	.030	.035	.225	.080	.137	.049
	BF_{10}	0.374	0.168	0.494	0.295	0.330	0.204	0.216	0.222	1.026	0.294	0.447	0.241
	p-value	.450	.905	.321	.593	.521	.895	.842	.817	.132	.597	.363	.748
Right CST	Pearson's *r*	.098	.014	.182	.205	.246	.174	.116	-.039	.180	.168	.184	.113
	BF_{10}	0.331	0.198	0.663	0.830	1.293	0.615	0.378	0.152	0.647	0.580	0.670	0.370
	p-value	.519	.927	.225	.171	.100	.246	.444	.798	.231	.265	.222	.455
Left CR	Pearson's *r*	-.123	.023	-.211	.258	-.386	.078	-.124	.047	-.286	.146	-.247	.131
	BF_{10}	0.108	0.208	0.081	1.503	0.052	0.289	0.107	0.238	0.066	0.481	0.073	0.423
	p-value	.417	.878	.160	.083	.008	.607	.411	.757	.054	.332	.098	.387
Right CR	Pearson's *r*	.091	-.014	.210	.149	.087	-.018	-.009	-.126	.139	.067	.116	.011
	BF_{10}	0.315	0.171	0.872	0.491	0.309	0.168	0.176	0.107	0.453	0.270	0.378	0.194
	p-value	.550	.926	.161	.324	.564	.908	.953	.403	.357	.658	.444	.944
CC Forc. Minor	Pearson's *r*	.137	-.026	.210	.096	.157	.045	.187	.183	.335	.120	.237	.106
	BF_{10}	0.446	0.159	0.883	0.326	0.528	0.235	0.697	0.671	4.718	0.389	1.193	0.352
	p-value	.359	.860	.157	.521	.293	.761	.209	.219	.021	.423	.109	.476
CC Forc. Major	Pearson's *r*	-.104	-.207	-.056	-.033	-.146	-.089	-.025	.105	.009	-.055	-.065	-.047
	BF_{10}	0.114	0.080	0.139	0.155	0.098	0.121	0.161	0.349	0.190	0.139	0.133	0.145
	p-value	.487	.162	.709	0.827	.327	.552	.869	.481	.954	.712	.662	.754

Note: Parts of the SLF were only traceable in 5 out of 47 participants with DTI, making the correlation analyses with the FA index impossible.

CHAPTER 6

The Added Value of Structural Brain Imaging Measures in Predicting Children's Academic Achievement: The Case of Arithmetic Fluency

Abstract

The current study aimed to investigate the added value of various structural brain imaging measures on top of behavioral measures in predicting individual differences in children's arithmetic fluency. Participants were 43 typically developing 9-10-year-olds. Individual differences in grey matter structure were investigated by looking at volume through voxel-based morphometry, and at cortical complexity through fractal dimensionality. Individual differences in white matter were examined through diffusion weighted magnetic resonance imaging and were analyzed via spherical deconvolution. As behavioral predictors, children's numerical magnitude processing, working memory, and rapid automatized naming were assessed. Motor reaction and intelligence were assessed as control variables. Neural and behavioral measures were added to a series of multiple regression models to predict arithmetic fluency. Symbolic number processing and RAN emerged as critical behavioral predictors. The white matter integrity of the right inferior longitudinal fasciculus and the cortical complexity of the left postcentral gyrus were found to be critical neuroanatomical predictors. Adding both behavioral and neural measures to one model revealed that the neuro-anatomical predictors of arithmetic fluency provided the best prediction of performance. These results highlight the value of brain imaging measures for the prediction of cognitive skills and strive towards a bridge between cognitive neuroscience and education.

6.1. Introduction

The ability to add, subtract, multiply or divide numbers, or arithmetic, is an essential skill for further mathematical and educational development (Kilpatrick et al., 2001). Within children's arithmetic, however, large individual differences exist, which has led recent research to try and gain insights into which cognitive factors might explain or predict these individual differences (e.g., Peng et al., 2016; Schneider et al., 2017; Vanbinst & De Smedt, 2016). Furthermore, over the last few years, many neuroimaging studies have also aimed at unraveling the neural basis of these individual differences in children's arithmetic, mainly on a functional, but also at a structural level (e.g., Arsalidou et al., 2018; Peters & De Smedt, 2018). From a perspective of educational neuroscience, which places itself at the junction of cognitive neuroscience, psychology, and educational research, these studies are exceedingly promising to understand the biological processes that play a role for educationally relevant skills, but additionally, in terms of the so-called process of neuro-prediction (De Smedt & Grabner, 2015), to predict educational outcomes, and generate predictions to be tested in educational research (De Smedt, 2018a; Hoeft et al., 2011; Howard-Jones et al., 2016; Supekar et al., 2013). For example, brain imaging data collected before the acquisition of certain behavioral skills, such as arithmetic or reading, could allow the identification of at-risk children before the start of formal education, leading to opportunities for early intervention (Diamond & Amso, 2008; Goswami, 2008; Ozernov-Palchik & Gaab, 2016). Moreover, these neuroimaging studies can also investigate if brain imaging measures are able to predict subsequent learning gains (Hoeft et al., 2011), or if they can predict responses to educational interventions (Supekar et al., 2013). In all, these brain imaging measures thus clearly have value in the prediction of the individual differences that exist in children's arithmetic. However, to this day, hardly any studies have studied the various relevant behavioral and neural correlates of arithmetic simultaneously in one sample of participants. Accordingly, it is highly interesting to consider the added value of different neural predictors for children's arithmetic fluency on top of well-known cognitive correlates, and whether the combination of behavioral and brain imaging measures leads to better prediction of children's arithmetic fluency. The current study aims to fill this gap by calculating multiple regression models with behavioral measures known to be strong predictors of arithmetic achievement (including numerical magnitude processing, working memory, and rapid automatized naming) and structural brain imaging data, including measures of both grey and white matter based on previous research (Arsalidou et al., 2018; Matejko & Ansari, 2015; Moeller et al., 2015; Peters & De Smedt, 2018) and collected through novel, more advanced imaging techniques (i.e., cortical complexity analyses and spherical deconvolution), on typically developing children's arithmetic fluency. To minimize maturational confounds, the current study focuses on children with a narrow age range (i.e., 9-10 year-olds; all 4th graders), at a point in development where considerable arithmetic knowledge has already been automatized.

6.1.1. Behavioral correlates

A prominent domain-specific cognitive predictor of mathematic achievement is numerical magnitude processing. This ability to understand and process numerical magnitude information is considered an important foundation for higher-level mathematical competence (Ansari, 2008; De Smedt, Noël, Gilmore, & Ansari, 2013; Schneider et al., 2017). Generally, children's magnitude representations have been studied through magnitude comparison tasks, in which children have to, as quickly as possible, indicate the largest of two presented numerosities (Ansari, 2008), in both a symbolic (digits) and non-symbolic (dots) format. Across development, individual differences in both tasks are linked to general arithmetic skills (e.g., Inglis, Attridge, Batchelor, & Gilmore, 2011; Price, Palmer, Battista, & Ansari, 2012), however, the association for symbolic numerical magnitude processing, which includes the notion that we can obtain abstract representations of numerical magnitude, is found to be a stronger and more robust predictor of arithmetic achievement (r = .30; Schneider et al., 2017). Furthermore, recent studies (e.g., Vanbinst et al. 2015a; Vanbinst, Ghesquière, & De Smedt, 2015b) have shown that symbolic numerical magnitude processing is especially important for children's ability to acquire and retrieve arithmetic facts, potentially over the entire primary school period (Vanbinst & De Smedt, 2016). Accordingly, symbolic numerical magnitude processing becomes important for further mathematical development as arithmetic facts could be stored in long-term memory in a meaningful way (i.e., according to their magnitude; Butterworth, Zorzi, Girelli, & Jonckheere, 2001; Robinson, Menchetti, & Torgesen, 2002).

Other than numerical magnitude processing, domain-general cognitive correlates of arithmetic achievement must also be considered for their predictive value. One of the most prominent domain-general predictors of arithmetic is working memory (for a meta-analysis, see Peng et al., 2016), which refers to the capacity of storing information for short periods of time when engaging in cognitively demanding activities (Baddeley, 1986). From a theoretical viewpoint, working memory is important for arithmetic development as arithmetic often involves the processing and storing of information simultaneously (e.g., remembering numbers during multi-digit calculations and word-problem solving; e.g., Raghubar, Barnes, & Hecht, 2010; Swanson & Jerman, 2006). The meta-analysis by Peng et al. (2016) confirmed the role of working memory as a domain-general predictor of mathematics performance and found that the relationship between working memory and arithmetic (i.e., whole-number calculations) is of medium strength (r = .35). Furthermore, the meta-analysis also showed that these associations were similar across domains of working memory (i.e., verbal, visuospatial, or numerical), yet they were different depending on the mathematical ability under study (Peng et al., 2016).

Another domain-general cognitive structure that can act as a predictor for arithmetic fluency lies in rapid automatized naming (RAN), or the fast retrieval of phonological information from long term memory (De Smedt, 2018b; De Smedt, Taylor, Archibald, & Ansari, 2010; Koponen, Salmi, Eklund, & Aro, 2013). For example, the well-known triple code model (Dehaene & Cohen, 1995) states that numbers can be represented in a verbal-phonological code, which is activated during the retrieval of arithmetic facts from semantic memory. In general, it has thus been suggested that poor (access to) phonological representations in long-term memory can interfere with retrieval, manipulation, and retention of phonological codes; if phonological representations for number words or facts in long-term memory are weak, they will be more difficult to retrieve quickly and accurately (Koponen et al., 2013; Simmons & Singleton, 2008). Rapid automatized naming thus reflects this ability to easily and rapidly access phonological information stored in long-term memory (Torgesen, Wagner, and Rashotte, 1994; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997). Accordingly, measures of RAN have been shown to predict arithmetic achievement (e.g., Chard et al., 2005; Hecht et al., 2001; Koponen, Mononen, Räsänen, & Ahonen, 2006; Landerl, Bevan, & Butterworth, 2004; Mazzocco & Grimm, 2013).

6.1.2. Structural grey matter imaging

Over the past few years, many studies have aimed to unravel the neural basis of children's arithmetic, but have mainly focused on grey matter functionality. The amount of studies looking at the structural grey matter correlates of children's arithmetic, however, is very scarce (Arsalidou et al., 2018; Peters & De Smedt, 2018). The few studies that did study these structural correlates (see Peters & De Smedt, 2018 for an overview) mainly implemented voxel-based morphometry, which typically uses T1-weighted volumetric MRI scans and performs statistical tests across voxels to identify volume differences between groups (Whitwell, 2009). Accordingly, some structural imaging studies have compared groups of children who differed in their level of arithmetic skill (Isaacs et al. 2001; Ranpura et al. 2013; Rotzer et al. 2008; Rykhlevskaia et al. 2009), and have mainly observed group differences (i.e., reduced grey matter volume for the children with deficits in calculation ability) in the bilateral intraparietal sulci, left inferior frontal gyrus, bilateral middle frontal gyri, and bilateral fusiform gyri.

Within voxel-based morphometry, it is also possible to perform regression analyses across voxels to assess neuroanatomical correlates of cognitive or behavioral skills, thus applying a more dimensional approach. Doing so, a small amount of studies also exists on the association between grey matter volume and arithmetic within typically developing children. For example, a study by Li et al. (2013) revealed that, in 9 to 11 year-old Chinese children, individual differences in arithmetic were positively correlated with grey matter volume in the left intraparietal sulcus. Using a longitudinal design, Evans et al. (2015) observed that grey matter volume of various parts of the arithmetic brain network (i.e., posterior parietal cortex, ventral occipito-temporal cortex, and prefrontal cortex) predict the growth in arithmetic across primary school. A study by Price et al. (2016) showed that grey matter volume in the left intraparietal sulcus at the end of the 1st grade is related to math competence at the end of the 2nd grade. Finally, Supekar et al. (2013) observed that the volume of the right hippocampus predicted the learning gains of one-on-one tutoring sessions in third-graders, with larger hippocampal volumes before the intervention predicting larger intervention gains.

A more recent study (Polspoel et al., *submitted*) also aimed to contribute knowledge on which grey matter regions are important for children's arithmetic at a structural level. Polspoel et al. (*submitted*) tested 47 typically developing 9-10 year-old children (a small age range was selected to minimize maturational confounds) and observed positive correlations between arithmetic fluency at this single point in development and the volume of the right fusiform gyrus. Furthermore, Polspoel et al. (*submitted*) went beyond the limitations of voxel-based morphometry, and also studied cortical complexity based on fractal dimensionality (i.e., the notion that fractal geometry can be used to quantify neural complexity) as discussed in Yotter et al. (2011). Unlike voxel-based morphometry, cortical complexity is sensitive to other differences in grey matter structure (e.g., the shape of structures) that are not indexed by volume or cortical thickness. Such fractal dimensionality analyses have been implemented in comparing patient groups (e.g., Alzheimer's disease; Ruiz de Miras et al., 2017 or Williams syndrome; Thompson et al., 2005) to controls, and have also been used to study age and gender related differences (Luders et al., 2004; Madan & Kensinger, 2016), and, most notably, differences in cognitive function (King et al., 2010; Im et al., 2006; Mustafa et al., 2012; Sandu et al., 2014). Implementing this novel technique, Polspoel et al. (*submitted*) observed positive correlations between children's arithmetic fluency and cortical complexity in the left postcentral gyrus, the right insular cortex (more specifically the right insular sulcus), and the left orbital sulcus.

6.1.3. Structural white matter imaging

As the brain regions of the well-known fronto-parietal arithmetic brain network (Menon, 2015) are not adjacent, but spatially distant from one another, it is also crucial to study the structural white matter connections between these regions. Furthermore, the connectivity between brain regions is integral to efficient cognitive processing (Johansen-Berg, 2010), making understanding the role of white matter pathways in arithmetic crucial for further clarification of the neural mechanisms of individual differences in arithmetic abilities (Matejko & Ansari, 2015). These white matter connections can be examined with diffusion-weighted Magnetic Resonance Imaging (dMRI), for which the most frequently applied model to relate the dMRI signal to the underlying neuroanatomy is Diffusion Tensor Imaging (DTI). This classic DTI model estimates the degree to which diffusion is not spherical but increased in a certain direction (fractional anisotropy or FA) per voxel (Tournier et al., 2007).

Structural connectivity research in children's arithmetic, however, is scarce and inconclusive. Many different white matter pathways have been found to be related to individual differences in arithmetic or other mathematical skills (see Matejko & Ansari, 2015, and Moeller et al., 2015 for reviews). These pathways include the arcuate fasciculus (Van Beek et al., 2014), superior longitudinal fasciculus (Li et al., 2013; Kucian et al., 2013; Navas-Sánchez et al., 2014; Rykhlevskaia et al., 2009; Tsang et al., 2009), inferior longitudinal fasciculus (Li et al., 2013; Navas-Sánchez et al., 2014; Rykhlevskaia et al., 2009), inferior fronto-occipital fasciculus (Li et al., 2013; Navas-Sánchez et al., 2014; Rykhlevskaia et al., 2009; Van Eimeren et al., 2008), uncinate fasciculus (Navas-Sánchez et al., 2014), corona radiata (Navas-Sánchez et al., 2014; Van Eimeren et al., 2008), corticospinal tract (Navas-Sánchez et al., 2014; Rykhlevskaia et al., 2009), and corpus callosum (Navas-Sánchez et al., 2014; Rykhlevskaia et al., 2009). All of these studies, however, applied classic DTI to study the various white matter tracts, which is subject to methodological limitations (e.g., Assaf et al. 2004; Dell'Acqua et al. 2013; Farquharson et al. 2013; Tournier et al. 2007). These limitations include the fact that DTI can only estimate the direction of one fiber per imaging voxel, leading to an oversimplification or inaccurate representation of the underlying anatomy. Furthermore, these studies were all conducted in research samples with wide age ranges (e.g., 7 to 10 or 10 to 15 years old; Matejko & Ansari 2015), which, even though statistically controlled for, might lead to maturational confounds and possible over-interpretation of associations between differences in connectivity and differences in mathematical development.

A study by Polspoel et al. (2018), on the other hand, correlated the white matter integrity of these previously observed arithmetic-related white matter pathways, to typically developing 9-10-year-old children's (i.e., the same 47 children as Polspoel et al., *submitted*) arithmetic fluency. In doing this, Polspoel et al. (2018) went beyond classic Diffusion Tensor Imaging (DTI), to tackle the existing methodological limitations by implementing the more novel and complex non-tensor model spherical deconvolution. Spherical deconvolution has the asset that it can characterize the orientation of more than one fiber per voxel, consequently providing the possibility of making more accurate statements about the observed associations with white matter tracts (Dell'Acqua et al., 2007; Tournier et al., 2004). Furthermore, to extract information on microscopic properties of the white matter tracts, the hindrance modulated orientational anisotropy (HMOA) index was derived for quantitative spherical deconvolution analyses. Using this more advanced technique to study the associations between white matter tracts and children's arithmetic fluency, Polspoel et al. (2018) observed significant positive correlations mainly between the right inferior longitudinal fasciculus (ILF) and children's arithmetic fluency across operations.

6.1.4. Current study

The current study aims to investigate the added predictive value of previously observed structural grey and white matter imaging measures, on top of well-known behavioral measures, for children's arithmetic fluency. In order to investigate this added value in predicting children's arithmetic fluency, the current study will combine and contrast behavioral measures (including numerical magnitude processing, working memory and RAN) and brain imaging measures based on previous reviews and meta-analyses (Arsalidou et al., 2018; Matejko & Ansari, 2015; Moeller et al., 2015; Peters & De Smedt, 2018) to calculate multiple regression models on typically developing children's arithmetic fluency. Doing so, the current study will use a sample with a narrow age range to minimize maturational confounds.

Based on the previous literature, we hypothesize that the collected behavioral measures will show unique associations with the children's arithmetic fluency. Furthermore, we expect that, similarly to previous research in reading (Hoeft et al., 2007), the regression models with both relevant behavioral and brain imaging measures will be more predictive of children's arithmetic fluency than regression models that only take either behavioral or brain imaging measures into account.

6.2. Methods

6.2.1. Participants

In total, data of 50 typically developing Flemish 4th graders were collected. These participants were the same as in Polspoel et al. (*submitted*), which reported the results on the associations between arithmetic and the voxel-based morphometry and cortical complexity data of this sample, and Polspoel et al. (2018), which reported the results on the associations between arithmetic and the dMRI data of this sample. Data of 7 children were discarded due to technical acquisition problems (n = 1), excessive motion (n = 1; see below for more details), problems during standardized testing (n = 1), or data quality (n = 4; see below for more details). Of the remaining 43 participants (ages 9 to 10; M = 9.68 years, SD = 0.34) 20 were girls, 23 were boys, and 8 children were left-handed. None of the participants had a history of learning difficulties, or neurological or psychiatric disorders. All participants were recruited via the elementary school they attended, all nearby the university, and were given a financial compensation for their participation. Written informed consent was obtained from a parent or legal guardian of each participating child. The study was approved by the Medical Ethical Committee of the University of Leuven (S59167).

All children were asked to take part in two test sessions. The first session contained the collection of behavioral data through various measures. In the second session, the MRI data were acquired. The second session always followed the first session by two to three weeks (M = 19.54 days, SD = 6.34), and also contained the collection of two different fMRI experiments that are not reported here (e.g., Polspoel et al. 2017).

6.2.2. Arithmetic assessment

First of all, children's arithmetic competence was measured through the Tempo Test Arithmetic (TTA; de Vos, 1992), as the main variable of interest. This standardized arithmetic test, similar to the Math Fluency subtest of the Woodcock-Johnson III tests of Achievement (Woodcock et al., 2003), is sensitive to individual differences in arithmetic fluency. The TTA contains five columns of increasingly difficult arithmetic items (i.e., one column per operation and a fifth column with mixed operations). Participants get one minute per column to write down as many correct answers as possible. The current sample included a small age range, leading to all participating children being part of the same norm group.

Consequently, the raw scores (i.e., the sum of the amount of correctly answered items per column) were used for statistical analyses.

6.2.3. Cognitive measures

6.2.3.1. Numerical magnitude processing

To assess the children's numerical magnitude processing (i.e., as a domain-specific predictor of arithmetic), a number comparison task was used (De Smedt & Gilmore, 2011). For this task, two numerosities are depicted on a computer screen (one on the left, one on the right) in either a symbolic (i.e., Arabic digits) or non-symbolic (i.e., dots) format. The goal is to indicate the larger of two response alternatives as quickly as possible. Both accuracy and reaction time were recorded, yet, as ceiling levels were reached for accuracy (symbolic M = 97.77%; non-symbolic M = 97.33%), only reaction time was used for analysis.

6.2.3.2. Working memory

As a domain-general predictor, working memory was assessed through the Corsi block-tapping test (Corsi, 1972), and a backwards digit span task (De Smedt et al., 2009). For the Corsi block-tapping test, children have to repeat a pattern of blocks tapped by the assessor, in increasing difficulty. The backwards digit span task involves the serial recall of spoken lists of digits between one and nine in reverse order. For both working memory tasks, the sequence increased with one block/number if two out of three sequences of the same size were correctly repeated. If not, the task was stopped. The amount of correctly repeated sequences was registered for analysis.

6.2.3.3. Rapid automatized naming

As another domain-general predictor, a RAN task was used to measure the children's fast retrieval of phonological information from long-term memory. The RAN task assesses how fast children can name colors, objects (all high-frequent, one-syllable Dutch words), numbers, and letters (van den Bos, 1998). Due to its numerical nature, however, the numbers subtest was not assessed to exclude possible confounds with the TTA. Each child was thus presented with three different sheets of paper, containing ten rows and five columns of images. The first sheet existed of rectangular shapes in five different colors (i.e., yellow, blue, green, red and black), the second one of five different objects (i.e., tree, duck, chair, bike and scissors), the third one of five different letters (i.e., a, d, o, p and s). For each page, participants have to name all colors, objects, or letters as fast as possible. Participant's timing was registered for every page separately. For data analysis, the average reaction time across pages was used.

6.2.4. Control measures

Finally, intelligence and motor reaction time were assessed as control variables. The WISC-III-NL Block Design and Vocabulary subtests were used to take intellectual ability into account, as measures of performance and verbal IQ, respectively (Wechsler, 2005). For intellectual ability, norm scores were used for the statistical analyses. Motor reaction time was measured to control for motoric speed during behavioral assessment, by having participants indicate which of two figures (always a circle, triangle, square, star, or heart; one on the left, one on the right) presented on a computer screen was filled in white by, as quickly as possible, pressing the corresponding key. Accuracy and reaction time were recorded for each trial, yet, as ceiling levels were reached for accuracy (M = 96.86%), only reaction time was used for analysis (De Smedt and Boets, 2010).

6.2.5. MRI data acquisition and pre-processing

All MRI scanning was done with a Philips Ingenia 3.0T CX MRI scanner with a SENSE 32-channel head-coil, located at the Department of Radiology of the University Hospital in Leuven, Belgium. Wash cloths were used to stabilize the children's heads and consequently minimize head motion. Next to the T1 and dMRI sequences, as a part of data collection for two different studies (e.g., Polspoel et al., 2017), the scanning session also included four fMRI runs of approximately 5 minutes each, leading to a total scanning time of approximately 40 minutes.

For the anatomical T1 images, the following parameters were implemented: 0.98 × 0.98 × 1.2 mm voxel size, 256 × 256 acquisition matrix, 8º flip angle, TE 4.6 ms, 250 × 250 × 218 mm field of view (approximately 8 minutes of scanning time). For the voxel-based morphometry and cortical complexity analyses, all preprocessing was done with the Computational Anatomy Toolbox (CAT12) within the Statistical Parametric Mapping software package for Matlab (SPM12, Wellcome Department of Cognitive Neurology, London), following the standard processing pipeline within the CAT12 software. Preprocessing included segmentation of the anatomical images; both grey matter and surface estimations were calculated. Next, for data quality and sample homogeneity testing, the Mahalanobis distance was used, which is a combination of weighted overall image quality (i.e., a measure of noise and spatial resolution before preprocessing) and mean correlation (i.e., a measure of the homogeneity of the data and thus the quality after preprocessing). Data with a Mahalanobis distance greater than two standard deviations of the sample average ($n = 4$) were discarded for further analysis. Finally, spatial smoothing was performed with 8 mm (voxel-based morphometry) and 20 mm (cortical complexity) FWHM Gaussian smoothing kernels. After preprocessing, mean values for volume and cortical complexity were estimated within each ROI (based on the existing literature; Arsalidou et al., 2018; Peters & De Smedt, 2018) and extracted for further analysis.

For the dMRI acquisition, sagittal slices were obtained using the following parameters: 60 noncollinear directions b-value 2000 s/mm2, 30 noncollinear directions b-value 700 s/mm2 (which were discarded for spherical deconvolution analyses), 6 nondiffusion-weighted images, 2.5 × 2.5 × 2.5 mm voxel size, 90º flip angle, repetition time (TR) 7000 ms, echo time (TE) 72 ms, and 240 × 125 × 240 mm field of view (approximately 12 minutes of scanning time). All pre-processing of the dMRI data was done using the Explore DTI software (Leemans et al., 2009), and existed of visual quality assurance, and rigorous motion, eddy current-induced distortion and EPI distortion correction. After motion correction, data displaying excessive motion (*n* = 1), defined as a mean translation in any direction greater than the voxel size of 2.5 mm, were discarded. No normalization to a standard atlas took place. Whole-brain DTI tractography was performed with the following parameters: FA-threshold = 0.20, maximum turning angle = 30º, and step length between calculations = 1 mm. For spherical deconvolution, additional processing steps were taken with the StarTrack software (Dell'Acqua et al., 2013): iterations = 200, *n* = 0.04, and *r* = 8. Finally, for the spherical deconvolution whole-brain tractography, the parameters were: absolute HMOA threshold = 0.06, relative HMOA threshold = 5%, maximum turning angle = 30º, and step length between calculations = 1 mm. Tractography of the white matter tracts was performed with the TrackVis software (Wang et al., 2007). All tracts were manually delineated for each subject using a region of interest (ROI) approach, based on anatomical landmarks in color-coded maps (Catani & Thiebaut de Schotten, 2008; Thiebaut de Schotten et al., 2011; Wakana et al., 2007). In such an approach, each ROI represents a mandatory (or prohibited) passage for the tract at hand. The colors in these maps refer to the direction the fibers run in; red fibers are commissural, green fibers are associative, and blue fibers are projection fibers. After manually delineating all white matter pathways, the TrackVis software offered statistical information of the tract at hand (i.e., HMOA value) which was then used for statistical testing.

6.2.6. Statistical analyses

First, both Bayesian and frequentist correlations were calculated for all behavioral and brain imaging measures with the results of the TTA. Bayesian statistics have the advantage of quantifying the evidence that data provide for one hypothesis over another (Andraszewicz et al., 2015). Accordingly, Bayes factors (BF_{10}) of 1, 1-3, 3-10, 10-30, 30-100, or > 100 respectively point towards no, anecdotal, substantial, strong, very strong, or decisive evidence for the hypothesis of an association between two variables (Jeffreys, 1961). For inclusion in the regression models, only the predictors for which at least substantial evidence ($BF_{10} > 3$) for an association between that measure and the total score of the TTA were considered. The control measures were always added to the models, regardless of any evidence for associations with the TTA. Both frequentist and Bayesian multiple regression models were calculated, first with the behavioral and brain imaging predictors separately, then with all predictors in the same model, for the total score of the TTA. For each model, all measures were simultaneously added to the model. Accordingly, p-values and inclusion Bayes factors (BF_{inc}; the change from prior to posterior inclusion odds and thus the extent to which the data support inclusion of the factor of interest, taking all models into account) were calculated, as well as which combination of variables constitutes the most predictive model. All models were checked for problems with multicollinearity for predictors that correlate highly with one another. All regression models were also calculated for the different columns of the TTA (i.e., for each operation and for the mixed column) separately; as similar results were observed between all operations and the total score, the results for these analyses are added to the appendix.

Finally, to assess the added value of one type of predictor (i.e., behavioral or brain imaging) on top of the other, hierarchical Bayesian regressions were calculated with the control variables and variables of one of two types added to the model first, then adding the variables of the second type of predictor to the model in a second step. Accordingly, the BF_{10} of the best predictive model in this second step indicates the increase in likeliness of predicting arithmetic fluency on top of the control variables and the predictors added in the first step, thus indicating the added predictive value of the variables in the second step of the hierarchical regression.

6.3. Results

6.3.1. Descriptive statistics

Table 6.1 displays the descriptive statistics of arithmetic, numerical magnitude processing, working memory, RAN, intellectual ability, and motor reaction time. The means of our sample were close to the expected population averages, and show proper variation. Even though the minimum scores for some of the tasks were low, none of the participating children had been diagnosed with any type of learning disorder or intellectual disability.

Table 6.1

Descriptive statistics of the arithmetic, numerical magnitude processing, working memory, rapid automatized naming, intellectual ability, and motor reaction time tasks

	Mean	**SD**	**Minimum**	**Maximum**
Arithmetic – Total (Raw)	102.1	19.01	73	160
Arithmetic – Total	5.54	2.82	1	10
Symbolic Number Comparison	804.6	180.7	567.7	1511
Non-Symbolic Number Comparison	1273	507.7	617.2	2682
Corsi	11.51	1.869	9	16
Digit Span Backwards	7.256	2.117	3	12
RAN	38.51	5.67	27.91	52.07
Block Design	10.86	3.21	6	19
Vocabulary	11.44	2.51	3	16
Motor Reaction Time	487.9	107.7	350.1	840.6

Note: Both raw and standardized scores are given for arithmetic. The scores on the arithmetic test are standardized as $M = 5$, $SD = 2$, with a maximum of 10. The scores for number comparison are raw scores displaying the average reaction time in ms. The Corsi and digit span scores are the raw scores displaying the amount of correctly repeated series. The rapid automatized naming scores are raw scores displaying the average reaction time in ms. The scores on Block Design and Vocabulary are standardized scores, standardized as $M = 10$, $SD = 3$, with a maximum of 19. The scores for motor reaction time are raw scores displaying the average reaction time in ms.

Table 6.2

Correlations between arithmetic assessment and all cognitive and control measures

		TTA Total	Symb Numb Comp	Non-Symb Numb Comp	Corsi	Backwards Digit Span	RAN Average	Block Design	Vocabulary	Motor RT
TTA Total	Pearson's *r*	/								
	BF_{10}	/								
	p-value	/								
Symb Numb Comp	Pearson's *r*	-.475	/							
	BF_{10}	28.360	/							
	p-value	.001	/							
Non-Symb Numb Comp	Pearson's *r*	-.288	.633	/						
	BF_{10}	1.037	4360.611	/						
	p-value	.061	<.001	/						
Corsi	Pearson's *r*	.393	-.287	-.215	/					
	BF_{10}	5.076	1.028	0.480	/					
	p-value	.009	.062	.166	/					
Backwards Digit Span	Pearson's *r*	.194	-.039	.080	.189	/				
	BF_{10}	0.403	0.196	0.215	0.387	/				
	p-value	.213	.806	.612	.225	/				
RAN Average	Pearson's *r*	-.445	.315	.145	-.218	-.162	/			
	BF_{10}	14.382	1.472	0.288	0.496	0.320	/			
	p-value	.003	.040	.353	.159	.299	/			
Block Design	Pearson's *r*	.441	-.365	-.241	.571	.195	-.223	/		
	BF_{10}	13.235	3.153	0.614	437.805	0.405	0.518	/		
	p-value	.003	.016	.119	<.001	.211	.150	/		
Vocabulary	Pearson's *r*	.066	.068	.138	.321	.301	-.050	.07	/	
	BF_{10}	0.207	0.208	0.277	1.603	1.220	0.200	0.209	/	
	p-value	.675	.664	.377	.036	.050	.749	.656	/	
Motor RT	Pearson's *r*	-.395	.647	.505	-.165	.039	.294	-.278	.055	/
	BF_{10}	5.247	7814.293	60.740	0.325	0.196	1.115	0.924	0.202	/
	p-value	.009	<.001	<.001	.292	.805	.056	.071	.725	/

6.3.2. Correlations

Table 6.2 shows the results of the correlation analysis between the total score of the TTA and all cognitive and control measures. As mentioned, only the predictors for which at least substantial evidence ($BF_{10} > 3$) for an association between the measure and the total score of the TTA were used for further analysis. Accordingly, only results of the symbolic number comparison, Corsi block-tapping, and RAN tests were used for further analysis. Results of the non-symbolic number comparison and backwards digit span tests were not analyzed further. The control measures (i.e., block design, vocabulary, and motor reaction time) were always added to the models, regardless of any evidence for associations with the TTA.

Results of the correlation analyses between the total score of the TTA and the brain imaging measures have been reported in Polspoel et al. (*submitted*) for the voxel-based morphometry and cortical complexity analyses, and in Polspoel et al. (2018) for the dMRI analyses. Based on the correlation results of the voxel-based morphometry and cortical complexity analyses (Polspoel et al., *submitted*), data of the volume of the right fusiform gyrus, and cortical complexity of the left postcentral gyrus, right insular sulcus, and left orbital sulcus were used for statistical analyses. Transverse slices with a visualization of these results can be found in Figure 6.1. Based on the correlation results of the dMRI (Polspoel et al., 2018), the HMOA values of the right ILF were used for further data analysis. A visual representation of this white matter tract can be found in Figure 6.2.

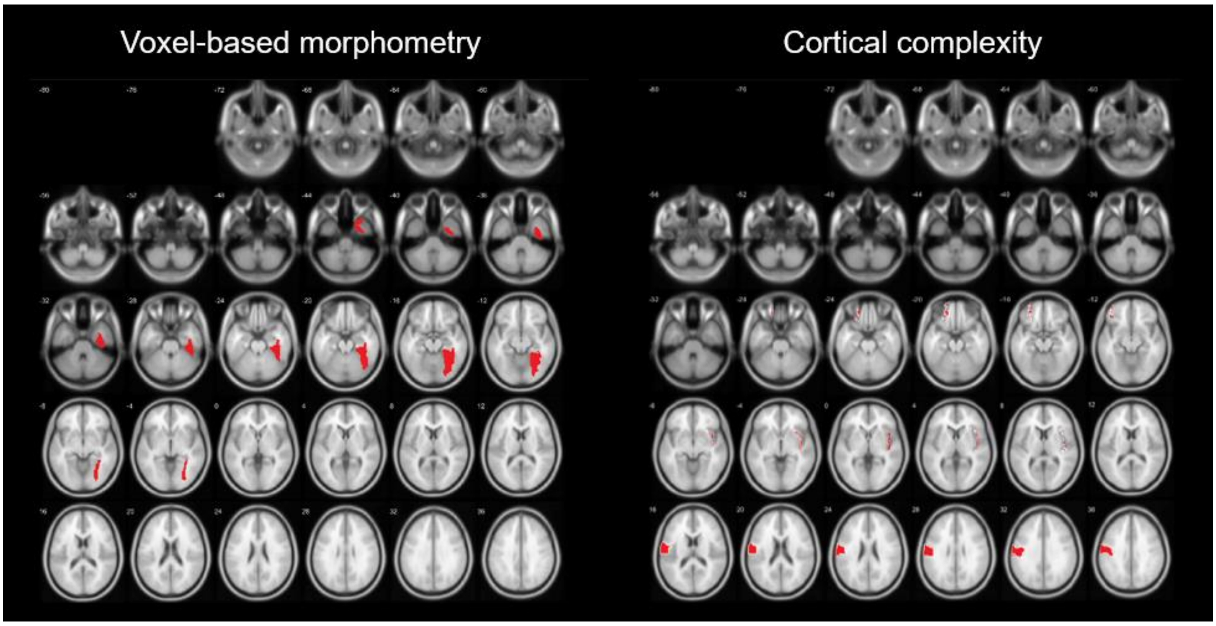

Figure 6.1. Visual representations of the structural ROIs for which the volume (left) or cortical complexity (right) is associated with children's arithmetic fluency.

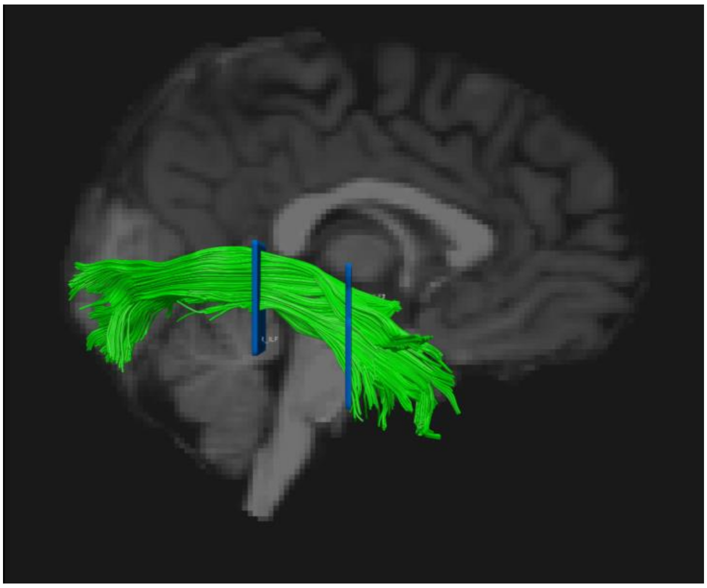

Figure 6.2. Visual representation of the right inferior longitudinal fasciculus. The ILF is delineated through an ROI on a coronal slice at the posterior edge of the cingulum, and an ROI entailing the entire temporal lobe on the most posterior coronal slice in which the temporal lobe is not connected to the frontal lobe.

6.3.3. Regression models

As some of the predictors in the multiple regression models correlated with one another, we checked for problems with multicollinearity by quantifying its severity. This was done by calculating the variance inflation factor (VIF), which is the ratio of variance in a model with multiple terms, divided by the variance of a model with one term alone (Allison et al., 1999). A rule of thumb is that multicollinearity is considered high and possibly problematic once the VIF is above 5 (Kutner, Nachtsheim, & Neter, 2004). In our data, however, the highest observed VIF was 2.2, indicating no problems with multicollinearity.

The regression model of all behavioral predictors on the total score of the TTA can be found in Table 6.3; based on the Bayesian regression analyses, the most predictive model is also stated. The adjusted R^2 for the entire model was $R^2 = .305$. With all behavioral predictors added to the model, no significant results or an inclusion Bayes factor above 3 are observed. The most predictive model, however, is a combination of symbolic magnitude processing, RAN and Block Design.

The regression model of the brain imaging predictors on arithmetic fluency can be found in Table 6.4. The adjusted R^2 for the entire model was $R^2 = .665$. For the total score of the TTA, the HMOA values of the right ILF and the cortical complexity of the left postcentral gyrus, right insular sulcus and left orbital sulcus all remain significant predictors (with a BF_{inc} above 3). The most predictive model is a combination of these four brain imaging measures.

Table 6.3

Multiple regression of the behavioral predictors on the total score of the TTA

	Correlations		*B*	*SE B*	β	*t*	*p*	BF_{inc}	R^2
	Zero-order	Partial							
Control Variables			**Most predictive model:** Block Design + Numb Comp + RAN (BF_{10} = 127.44)						
Block Design	.441	.192	1.141	0.971	0.193	1.176	.247	1.375	.305
Vocabulary	.066	.016	0.102	1.052	0.013	0.097	.924	0.381	
Motor RT	-.395	-.097	-0.018	0.030	-0.099	-0.582	.564	0.679	
Primary Variables									
Numb Comp	-.475	-.196	-0.023	0.019	-0.214	-1.200	.238	1.699	
Corsi	.393	.137	1.431	1.726	0.141	0.829	.413	0.868	
RAN	-.445	-.313	-0.919	0.464	-0.274	-1.981	.055	2.767	

Note: The provided R^2 values are adjusted R^2. All variables were simultaneously added to the model.

Table 6.4

Multiple regression of the brain imaging predictors on the total score of the TTA

	Correlations		*B*	*SE B*	β	*t*	*p*	BF_{inc}	R^2
	Zero-order	Partial							
Control Variables			**Most predictive model:** R ILF +LG Postcentral + RS Insula + LS Orbital (BF_{10} = 5.63e+6)						
Block Design	.441	.224	0.826	0.617	0.140	1.339	.190	0.809	.665
Vocabulary	.066	-.002	-0.010	0.701	-0.001	-0.014	.989	0.256	
Motor RT	-.395	-.188	-0.019	0.017	-0.110	-1.118	.271	0.649	
Primary Variables									
R ILF HMOA	.551	.597	1108.311	255.576	0.434	4.337	< .001	559.353	
R Fusiform Vol	.376	.036	1.073	5.135	0.023	0.209	.836	0.328	
LG Postcentral Comp	.539	.522	43.847	12.300	0.354	3.565	.001	79.683	
RS Insula Comp	.425	.376	39.063	16.515	0.238	2.365	.024	5.168	
LS Orbital Comp	.382	.373	17.676	7.532	0.222	2.347	.025	3.885	

Note: The provided R^2 values are adjusted R^2. All variables were simultaneously added to the model.

The regression model of all predictors on the total score of the TTA can be found in Table 6.5. The adjusted R^2 for the entire model was $R^2 = .656$. Significant results and an inclusion Bayes factor above 3 are only observed for the HMOA values of the right ILF and the cortical complexity of the left postcentral gyrus and right insular sulcus. The most predictive model for the total score contains these three brain imaging measures, but also the cortical complexity of the left orbital sulcus.

Table 6.5

Multiple regression of behavioral and brain imaging predictors on the total score of the TTA

	Correlations		*B*	*SE B*	β	*t*	*p*	BF_{inc}	R^2
	Zero-order	Partial							
Control Variables		**Most predictive model:** R ILF +LG Postcentral + RS Insula + LS Orbital (BF_{10} = 5.63e+6)							
Block Design	.441	.127	0.516	0.721	0.087	0.715	.480	0.532	.656
Vocabulary	.066	-.042	-0.181	0.778	-0.024	-0.233	.817	0.279	
Motor RT	-.395	-.104	-0.012	0.021	-0.070	-0.581	.566	0.476	
Primary Variables									
Numb Comp	-.475	-.056	-0.004	0.014	-0.042	-0.314	.756	0.596	
Corsi	.393	.074	0.550	1.337	0.054	0.411	.684	0.428	
RAN	-.445	-.222	-0.460	0.363	-0.137	-1.268	.214	1.161	
R ILF HMOA	.551	.594	1070.916	260.643	0.419	4.109	< .001	338.146	
R Fusiform Vol	.376	.039	1.252	5.719	0.026	0.219	.828	0.363	
LG Postcentral Comp	.539	.522	42.739	12.540	0.345	3.408	.002	73.339	
RS Insula Comp	.425	.351	35.947	17.223	0.219	2.087	.045	3.322	
LS Orbital Comp	.382	.249	12.330	8.608	0.155	1.432	.162	1.394	

Note: The provided R^2 values are adjusted R^2. All variables were simultaneously added to the model.

All analyses were also calculated for the different columns (i.e., all four operations and the mixed column) of the TTA. Yet, as similar results were observed between all operations and the total score, these analyses are not discussed further. The results of these analyses can be found in the appendix (Tables 6.A1 – 6.A5).

Finally, the hierarchical Bayesian regressions to assess the added value of one type of predictor on top of the other, indicated a strong increase in the likelihood of predicting arithmetic fluency when adding the brain imaging measures to the behavioral and control measures. The most predictive model in this analysis existed of the HMOA values of the right ILF, and cortical complexity of the left postcentral gyrus and right insular sulcus, and had a BF_{10} = 4445.56. Adding the behavioral to the brain imaging measures, however, did not improve the prediction of the model, as the most predictive model, only existing of RAN, had a BF_{10} = 0.687.

6.4. Discussion

The current study aimed to investigate the added value of various structural brain imaging measures over well-known behavioral predictors of arithmetic, in predicting typically developing children’s arithmetic fluency. First of all, we performed a correlation analysis between various behavioral measures and arithmetic fluency and, in accordance with the existing literature, observed significant associations for symbolic numerical magnitude processing (Ansari, 2008; De Smedt et al., 2013; Schneider et al., 2017; Vanbinst & De Smedt, 2016), working memory (Peng et al., 2016; Raghubar et al., 2010; Swanson & Jerman, 2006), and RAN (Chard et al., 2005; Hecht et al., 2001; Koponen et al., 2006; Landerl et al., 2004; Mazzocco & Grimm, 2013). These behavioral predictors were then added to a multiple regression model, together with structural grey and white matter imaging measures based on previous reviews and meta-analyses (Arsalidou et al., 2018; Matejko & Ansari, 2015; Moeller et al., 2015; Peters and De Smedt, 2018), which displayed statistical significance and at least substantial evidence for an association with the TTA. The results of the correlations of the TTA with the structural brain imaging measures are discussed in two different studies (i.e., Polspoel et al., 2018, and Polspoel et al., *submitted*), which took the methodological limitations of voxel-based morphometry and classic DTI into account.

The results of the correlations calculated between the behavioral measures and the results of the TTA were to be expected and corroborated findings of previous research. First of all, statistical significance and (at least) substantial evidence was observed for associations between the children's arithmetic fluency and numerical magnitude processing (De Smedt et al., 2013; Schneider et al., 2017). Noteworthy is that the associations found for numerical magnitude processing and the TTA were only observed for symbolic, and not for non-symbolic, number comparison. This agrees with the notion that symbolic numerical magnitude processing is a more robust predictor of individual differences in arithmetic achievement (De Smedt et al., 2013), for which evidence was found in the meta-analysis by Schneider et al. (2017), who reported significantly larger associations for symbolic numerical magnitude processing as a predictor of mathematics achievement. Secondly, significant correlations were also observed for working memory (i.e., Corsi block-tapping), corroborating previous studies (e.g., Peng et al., 2016; Raghubar et al., 2010; Swanson & Jerman, 2006), and highlighting the importance of this domain general skill as arithmetic often involves the processing and storing of information simultaneously. Finally, RAN was also observed to be significantly correlated to children's arithmetic fluency, highlighting the importance of being able to easily and rapidly access phonological information stored in long-term memory (Chard et al., 2005; Hecht et al., 2001; Koponen et al., 2006; Landerl et al., 2004; Mazzocco & Grimm, 2013; Torgesen et al., 1994; Torgesen et al., 1997).

Next, multiple regression models were calculated of the behavioral measures predicting the total score of the TTA. The results of these analyses show that, when putting all five behavioral predictors into the same model, none of the predictors remain significant over the other predictors added to the model. However, when calculating the most predictive model for the total score through Bayesian analyses, symbolic number comparison and RAN were added to the model, next to the control variable Block Design. These results thus stress the importance of symbolic numerical magnitude processing as a domain-specific predictor (De Smedt et al., 2013; Schneider et al., 2017), but also of RAN as a domain-general predictor of children's arithmetic fluency, again highlighting the importance of rapidly accessing phonological information within children's arithmetic (Chard et al., 2005; Hecht et al., 2001; Koponen et al., 2006; Landerl et al., 2004; Mazzocco & Grimm, 2013).

The used brain imaging measures for the current study include the volume of the right fusiform gyrus (analyzed through voxel-based morphometry; Whitwell, 2009), the cortical complexity of the left postcentral gyrus, right insular sulcus, and left orbital sulcus (calculated through the fractal dimensionality index; Yotter et al., 2011), and the white matter integrity of the right ILF (analyzed through spherical deconvolution; Dell'Acqua et al., 2007; Tournier et al., 2004). Multiple regression models were calculated for these five structural brain imaging measures on the total score of the TTA. The results of these analyses mainly point towards the cortical complexity of the left postcentral gyrus and the white matter integrity of the right ILF as important predictors for children's arithmetic, and to a lesser extent to the cortical complexity of the right insular sulcus and left orbital sulcus. The volume of the right fusiform gyrus, on the other hand, was not found to be a significant predictor when the other predictors were added to the model. The postcentral gyrus lies in continuity of the intraparietal sulcus and is adjacent to the superior parietal lobe, whose roles within the representation and manipulation of numerical quantity and arithmetic in general have been clearly established (Menon, 2015). The cortical complexity of the postcentral gyrus may then become important as the region acts as an extension of, and might affect its adjacent arithmetic-related regions. Previously, activation in the postcentral gyrus was mainly observed during grasping tasks (Simon et al., 2002), but the region has also been linked to the use of arithmetic strategies such as subvocalization and finger counting (Kesler et al., 2006). Furthermore, in both adults and children, activation in the right inferior parietal cortex (including the intraparietal sulcus and the postcentral gyrus) has also been related to nonsymbolic numerical and spatial processing (Kaufmann et al., 2008), indicating the region's importance for number-related tasks (Arsalidou et al., 2018).

Next, the ILF connects the occipital lobe to the anterior part of the temporal lobe, including the fusiform gyri and parahippocampal regions. Accordingly, the observed associations to children's arithmetic fluency might be related to the efficiency with which children process Arabic numerals as research has shown that inferior temporal regions are involved in the processing of visual representations of numerical symbols (Dehaene et al., 2003; Shum et al., 2013). Alternatively, arithmetic fluency might be related to broader mathematical processing as the function of the inferior temporal gyrus has been assumed to go beyond a specific preference to Arabic numbers (Grotheer et al., 2018). Finally, as the ILF also mediates the interaction between medial, inferior and anterior temporal cortices with Perisylvian areas, it might thus be related to language (Catani & Mesulam, 2008), making it vital for exact verbal arithmetic skills (e.g., fact retrieval), as it could subserve as a first step in connecting the lingual, fusiform, and parahippocampal regions in the ventral visual stream, upwards to the dorsal visual stream (Rykhlevskaia et al., 2009).

As mentioned, the results of the brain imaging regression model also points, to a lesser extent, to the importance of the right insular sulcus and left orbital sulcus. The insula has often been observed in arithmetic studies (Arsalidou and Taylor, 2011), but its exact function in arithmetic is still unclear. In general, the insula is known for directing attentional resources and decision-making (Arsalidou and Taylor, 2011; Menon, 2015), but has also been implicated to be important for emotional processing (Damasio et al., 2000) and speech-motor function (Fox et al., 2001). Similar to the insular cortex, the function of the orbitofrontal cortex within arithmetic could be associated with attention, decision-making, and executive function (Han et al., 2013). Further research is necessary to definitively identify the functions of both the insular and orbitofrontal cortices.

Finally, large multiple regression models were calculated with both behavioral and brain imaging predictors added to the model simultaneously. In these analyses, the brain imaging measures, especially the cortical complexity of the left postcentral gyrus and the HMOA values of the right ILF, seem to be stronger predictors of the children's arithmetic fluency than any of the behavioral measures. With the brain imaging measures added to the regression model, the behavioral measures never reach statistical significance, and are never added to the most predictive model. However, to definitively assess the added value of one type of predictor on top of the other (i.e., behavioral or brain imaging), hierarchical Bayesian regressions were calculated with the control variables and variables of one of two types added to the model first, then adding the variables of the second type of predictor to the model in a second step, and calculating the BF_{10} of the best predictive model to indicate the increase in likeliness of predicting arithmetic fluency on top of the control variables and the predictors added in the first step. These analyses indicated a strong increase in the likelihood of predicting arithmetic fluency when adding the brain imaging measures to the behavioral and control measures, but a reverse effect when adding the behavioral measures to the brain imaging measures.

Overall, our results thus show that the structural neuroimaging measures predict the children's arithmetic fluency better than the current standardized assessment tests. These results are not in line with similar research on children's reading skills (Hoeft et al., 2007), where a combination of behavioral and neuroimaging measures predicted reading outcome significantly better than either behavioral or neuroimaging models alone. Obviously, our results do not mean that neuroimaging measures have sufficient value and should be implemented for practical predictions or educational interventions on their own, but do stress their importance. Accordingly, combinations of behavioral and neural measures hold potential for understanding the biological processes that play a role for educationally relevant skills (Black, Myers, & Hoeft, 2015; De Smedt, 2018a). Furthermore, from a perspective of neuro-prediction, taking brain imaging measures into account could assist in predicting educational outcomes or in generating predictions to be tested in educational research, and improving the specificity and effectiveness of interventions (De Smedt, 2018a; Hoeft et al., 2011; Howard-Jones et al., 2016; Supekar et al., 2013). Similar to Hoeft et al. (2007), the current study strives towards a point where a bridge can be built between cognitive neuroscience and education.

Furthermore the current study was conducted with a research sample of only 9 to 10 year-olds (i.e., fourth graders), to minimize maturational confounds. Consequently, we feel it is essential to highlight the need for similar studies in children of different ages, such as in first or second grade (i.e., in early arithmetic development), or in secondary school (i.e., in more advanced levels of arithmetic). Accordingly, studies with a longitudinal follow-up throughout arithmetic development would also be highly informative to investigate how the balance between these predictors might shift over time, keeping educational-based neural plasticity in mind.

Finally, we would like to stress that arithmetic development is to a large extent dependent on the educational environment in which it develops. Accordingly, the emphasis on automatization processes within mathematics curricula might influence the results of (brain imaging) research on arithmetic fluency (De Smedt, 2016). Since all participants of the current study, however, came from Belgian elementary schools, high automatization skills were to be expected (Torbeyns et al., 2004). Hence, it is plausible that studies across cultures with differences in the emphasis on automatization might point towards different neural associations for the same set of arithmetic items as were found here, leading to different predictive models.

6.5. Appendix

Table 6.A1

Descriptive statistics of all separate columns of the arithmetic assessment

	Mean	SD	Minimum	Maximum
Addition (Raw)	23.26	3.730	16	33
Addition	5.84	2.886	1	10
Subtraction (Raw)	21.09	4.815	12	34
Subtraction	5.51	3.104	1	10
Multiplication (Raw)	20.09	3.379	13	26
Multiplication	5.86	2.731	1	10
Division (Raw)	18.28	5.128	9	35
Division	5.74	2.564	1	10
Mix (Raw)	19.40	4.387	12	32
Mix	4.98	2.816	1	10

Note: Both raw and standardized scores are given for arithmetic. The scores on the arithmetic test are standardized as $M = 5$, $SD = 2$, with a maximum of 10.

Table 6.A2

Correlations between all columns of the TTA and the cognitive and control measures used for the multiple regression models

		TTA Addition	**TTA Subtraction**	**TTA Mult**	**TTA Division**	**TTA Mix**	**Symb Numb Comp**	**Corsi**	**RAN Average**	**Block Design**	**Vocabulary**	**Motor RT**
TTA Addition	Pearson's *r*	/										
	BF_{10}	/										
	p-value	/										
TTA Subtraction	Pearson's *r*	.818	/									
	BF_{10}	4.782e +8	/									
	p-value	< .001	/									
TTA Mult	Pearson's *r*	0.727	.681	/								
	BF_{10}	476469.060	38505.342	/								
	p-value	< .001	< .001	/								
TTA Division	Pearson's *r*	.691	.616	.711	/							
	BF_{10}	62067.753	2175.364	189581.002	/							
	p-value	< .001	< .001	< .001	/							
TTA Mix	Pearson's *r*	.833	.824	.770	.691	/						
	BF_{10}	2.254e +9	9.002e +8	8.247e +6	63828.781	/						
	p-value	< .001	< .001	< .001	< .001	/						
Numb Comp	Pearson's *r*	-.520	-.474	-.435	-.242	-.477	/					
	BF_{10}	91.045	27.726	11.615	0.622	30.188	/					
	p-value	< .001	.001	.004	.117	.001	/					
Corsi	Pearson's *r*	.503	.492	.154	.223	.355	-.287	/				
	BF_{10}	58.109	43.274	0.305	0.517	2.666	1.028	/				
	p-value	< .001	< .001	.323	.150	.019	.062	/				
RAN Average	Pearson's *r*	-.472	-.347	-.413	-.316	-.461	.315	-.218	/			
	BF_{10}	26.329	2.332	7.379	1.491	20.295	1.472	0.496	/			
	p-value	.001	.023	.006	.039	.002	.040	.159	/			
Block Design	Pearson's *r*	.466	.494	.264	.312	.406	-.365	.571	-.223	/		
	BF_{10}	23.128	45.137	0.787	1.410	6.508	3.153	437.805	0.518	/		
	p-value	.002	< .001	.087	.042	.007	.016	< .001	.150	/		
Vocabulary	Pearson's *r*	.051	.239	-.033	-.067	.083	.068	.321	-.050	.070	/	
	BF_{10}	0.200	0.601	0.194	0.208	0.218	0.208	1.603	0.200	0.209	/	
	p-value	.744	.123	.833	.669	.596	.664	.036	.749	.656	/	
Motor RT	Pearson's *r*	-.374	-.310	-.478	-.322	-.308	.647	-.165	.294	-.278	.055	/
	BF_{10}	3.613	1.368	30.832	1.624	1.347	7814.293	0.325	1.115	0.924	0.202	/
	p-value	.014	.043	.001	.035	.044	< .001	.292	.056	.071	.725	/

Table 6.A3

Multiple regression of the behavioral predictors on all columns of the TTA

		Correlations								
		Zero-order	**Partial**	***B***	***SE B***	**β**	***t***	***p***	**BF_{inc}**	**R^2**
Addition	**Control Variables**									
		Most predictive model: Numb Comp + Corsi + RAN (BF_{10} = 2384.86)								
	Block Design	.466	.138	0.146	0.175	0.126	0.833	.410	0.705	.412
	Vocabulary	.051	-.063	-0.072	0.190	-0.048	-0.377	.709	0.357	
	Motor RT	-.374	-.022	-7.161e-4	0.005	-0.021	-0.132	.896	0.440	
	Primary Variables									
	Numb Comp	-.520	-.274	-0.006	0.003	-0.281	-1.711	.096	3.212	
	Corsi	.503	.305	0.599	0.311	0.300	1.925	.062	3.815	
	RAN	-.472	-.351	-0.188	0.084	-0.286	-2.247	.031	3.534	
Subtract	**Control Variables**									
		Most predictive model: Numb Comp + Corsi (BF_{10} = 221.82)								
	Block Design	.494	.240	0.353	0.238	0.235	1.484	.146	1.623	.349
	Vocabulary	.239	.214	0.339	0.258	0.177	1.316	.196	0.833	
	Motor RT	-.310	.026	0.001	0.007	0.026	0.156	.877	0.433	
	Primary Variables									
	Numb Comp	-.474	-.292	-0.008	0.005	-0.316	-1.832	.075	2.919	
	Corsi	.492	.180	0.465	0.423	0.180	1.099	.279	1.645	
	RAN	-.347	-.188	-0.131	0.114	-0.154	-1.148	.259	0.746	
Mult	**Control Variables**									
		Most predictive model: Motor RT + RAN (BF_{10} = 55.33)								
	Block Design	.264	.095	0.104	0.183	0.099	0.571	.571	0.475	.219
	Vocabulary	-.033	-.016	-0.019	0.198	-0.014	-0.095	.925	0.381	
	Motor RT	-.478	-.255	-0.009	0.006	-0.287	-1.585	.122	2.822	
	Primary Variables									
	Numb Comp	-.435	-.121	-0.003	0.004	-0.139	-0.734	.467	0.844	
	Corsi	.154	-.041	-0.081	0.325	-0.045	-0.248	.805	0.387	
	RAN	-.413	-.296	-0.163	0.087	-0.273	-1.861	.071	2.054	
Division	**Control Variables**									
		Most predictive model: Block Design + RAN (BF_{10} = 2.05)								
	Block Design	.312	.158	0.290	0.301	0.182	0.963	.342	0.980	.079
	Vocabulary	-.067	-.117	-0.230	0.327	-0.113	-0.705	.486	0.502	
	Motor RT	-.322	-.199	-0.011	0.009	-0.239	-1.215	.232	1.066	
	Primary Variables									
	Numb Comp	-.242	.066	0.002	0.006	0.081	0.394	.696	0.514	
	Corsi	.223	.078	0.253	0.536	0.092	0.472	.640	0.575	
	RAN	-.316	-.221	-0.196	0.144	-0.216	-1.357	.183	1.144	

Mix	**Control Variables**			**Most predictive model:** Block Design + Numb Comp + RAN (BF_{10} = 106.596)						
	Block Design	.406	.056	0.248	0.226	0.182	1.095	.281	1.037	.289
	Vocabulary	.083	.056	0.083	0.245	0.048	0.338	.737	0.402	
	Motor RT	-.308	-.291	0.002	0.007	0.059	0.339	.737	0.428	
	Primary Variables									
	Numb Comp	-.477	-.291	-0.008	0.004	-0.330	-1.828	.076	2.845	
	Corsi	.355	.080	0.194	0.403	0.083	0.481	.633	0.656	
	RAN	-.461	-.349	-0.242	0.108	-0.313	-2.237	.032	3.877	

Note: The provided R^2 values are adjusted R^2. All variables were simultaneously added to the model.

Table 6.A4

Multiple regression of the brain imaging predictors on all columns of the TTA

		Correlations		*B*	*SE B*	β	*t*	*p*	BF_{inc}	R^2
		Zero-order	**Partial**							
Addition	**Control Variables**	**Most predictive model:** Block Design + R ILF + LG Postcentral + LS Orbital (BF_{10} = 455400.53)								
	Block Design	.466	.320	0.262	0.133	0.225	1.968	.057	3.086	.597
	Vocabulary	.051	.033	0.029	0.151	0.020	0.195	.847	0.295	
	Motor RT	-.374	-.141	-0.003	0.004	-0.090	-0.830	.412	0.492	
	Primary Variables									
	R ILF HMOA	.407	.371	128.042	55.038	0.255	2.326	.026	4.522	
	R Fusiform V	.339	.025	0.161	1.106	0.017	0.146	.885	0.376	
	LG Postcentral C	.525	.476	8.360	2.649	0.344	3.156	.003	35.520	
	RS Insula C	.358	.205	4.345	3.556	0.135	1.222	.230	0.670	
	LS Orbital C	.505	.537	6.025	1.622	0.386	3.714	< .001	92.538	
Subtract	**Control Variables**	**Most predictive model:** Block Design + R ILF + LG Postcentral + RS Insula (BF_{10} = 2.35e+6)								
	Block Design	.494	.162	0.283	0.154	0.189	1.836	.075	1.919	.674
	Vocabulary	.239	.171	0.341	0.175	0.178	1.942	.060	1.116	
	Motor RT	-.310	-.025	-0.001	0.004	-0.028	-0.287	.776	0.317	
	Primary Variables									
	R ILF HMOA	.589	.399	289.111	63.925	0.447	4.523	< .001	691.905	
	R Fusiform V	.429	.099	1.449	1.284	0.120	1.128	.267	0.779	
	LG Postcentral C	.519	.307	10.704	3.076	0.342	3.479	.001	50.590	
	RS Insula C	.372	.133	6.222	4.131	0.150	1.506	.141	1.366	
	LS Orbital C	.285	.144	3.085	1.884	0.153	1.638	.111	0.886	
Mult	**Control Variables**	**Most predictive model:** Motor RT + R ILF + LG Postcentral (BF_{10} = 23147)								
	Block Design	.264	.016	0.012	0.127	0.011	0.092	.927	0.331	.548
	Vocabulary	-.033	-.177	-0.152	0.145	-0.113	-1.046	.303	0.435	
	Motor RT	-.478	-.424	-0.010	0.004	-0.312	-2.727	.010	10.503	
	Primary Variables									
	R ILF HMOA	.437	.519	186.839	52.776	0.411	3.540	.001	12.534	
	R Fusiform V	.097	-.310	-2.017	1.060	-0.239	-1.902	.066	0.822	
	LG Postcentral C	.520	.492	8.359	2.540	0.380	3.291	.002	49.795	
	RS Insula C	.360	.343	7.259	3.410	0.249	2.129	.041	1.259	
	LS Orbital C	.285	.187	1.727	1.555	0.122	1.110	.275	0.549	

Division	**Control Variables**			**Most predictive model:** R ILF + LG Postcentral + RS Insula (BF_{10} = 423.55)						
	Block Design	.312	.058	0.077	0.230	0.048	0.337	.738	0.469	.363
	Vocabulary	-.067	-.169	-0.261	0.261	-0.128	-0.999	.325	0.564	
	Motor RT	-.322	-.118	-0.004	0.006	-0.094	-0.691	.494	0.601	
	Primary Variables									
	R ILF HMOA	.488	.450	279.576	95.096	0.406	2.940	.006	18.086	
	R Fusiform V	.345	.072	0.802	1.911	0.062	0.420	.677	0.541	
	LG Postcentral C	.400	.326	9.203	4.577	0.276	2.011	.052	3.283	
	RS Insula C	.335	.262	9.734	6.145	0.220	1.584	.122	1.584	
	LS Orbital C	.264	.135	2.224	2.802	0.104	0.794	.433	0.617	
Mix	**Control Variables**			**Most predictive model:** R ILF + LG Postcentral + RS Insula + LS Orbital (BF_{10} = 53304.5)						
	Block Design	.406	.193	0.192	0.167	0.141	1.148	.259	0.809	.538
	Vocabulary	.083	.029	0.033	0.190	0.019	0.171	.865	0.314	
	Motor RT	-.308	-.030	-8.167e -4	0.005	-0.020	-0.173	.863	0.348	
	Primary Variables									
	R ILF HMOA	.487	.486	224.743	69.328	0.381	3.242	.003	42.842	
	R Fusiform V	.392	.083	0.678	1.393	0.062	0.487	.630	0.434	
	LG Postcentral C	.451	.348	7.220	3.337	0.253	2.164	.038	3.555	
	RS Insula C	.460	.403	11.503	4.480	0.304	2.568	.015	10.662	
	LS Orbital C	.386	.361	4.615	2.043	0.252	2.259	.030	3.321	

Note: The provided R^2 values are adjusted R^2. All variables were simultaneously added to the model.

Table 6.A5

Multiple regression of behavioral and brain imaging predictors on all columns of the TTA

		Correlations								
		Zero-order	**Partial**	***B***	***SE B***	**β**	***t***	***p***	**BF_{inc}**	**R^2**
Addition	**Control Variables**	**Most predictive model:** Numb Comp + Corsi + R ILF + LG Postcentral + LS Orbital (BF_{10} = 3.57e+6)								
	Block Design	.466	.093	0.076	0.145	0.065	0.522	.606	0.477	.639
	Vocabulary	.051	-.116	-0.101	0.156	-0.068	-0.649	.521	0.325	
	Motor RT	-.374	-.030	-7.082e -4	0.004	-0.020	-0.165	.870	0.418	
	Primary Variables									
	Numb Comp	-.520	-.173	-0.003	0.003	-0.135	-0.975	.337	0.955	
	Corsi	.503	.355	0.568	0.269	0.285	2.115	.043	5.732	
	RAN	-.472	-.158	-0.065	0.073	-0.099	-0.893	.379	0.599	
	R ILF HMOA	.407	.381	120.093	52.387	0.240	2.292	.029	3.163	
	R Fusiform V	.339	-.099	-0.636	1.149	-0.068	-0.553	.584	0.311	
	LG Postcentral C	.525	.515	8.425	2.520	0.347	3.343	.002	77.471	
	RS Insula C	.358	.147	2.867	3.462	0.089	0.828	.414	0.405	
	LS Orbital C	.505	.434	4.643	1.730	0.298	2.684	.012	17.495	
Subtract	**Control Variables**	**Most predictive model:** Numb Comp + Corsi + R ILF + LG Postcentral (BF_{10} = 1.32e+7)								
	Block Design	.494	.174	0.175	0.178	0.117	0.981	.334	0.628	.673
	Vocabulary	.239	.273	0.303	0.192	0.158	1.577	.125	0.864	
	Motor RT	-.310	.105	0.003	0.005	0.070	0.589	.560	0.308	
	Primary Variables									
	Numb Comp	-.474	-.248	-0.005	0.003	-0.187	-1.426	.164	1.529	
	Corsi	.492	.115	0.213	0.330	0.083	0.645	.524	1.364	
	RAN	-.347	-.092	-0.046	0.090	-0.054	-0.514	.611	0.380	
	R ILF HMOA	.589	.618	281.855	64.354	0.436	4.380	< .001	638.490	
	R Fusiform V	.429	.132	1.046	1.412	0.087	0.741	.464	0.496	
	LG Postcentral C	.519	.527	10.690	3.096	0.341	3.453	.002	87.142	
	RS Insula C	.372	.189	4.555	4.252	0.110	1.071	.292	0.769	
	LS Orbital C	.285	.148	1.771	2.125	0.088	0.833	.411	0.488	
Mult	**Control Variables**	**Most predictive model:** Motor RT + R ILF + LG Postcentral (BF_{10} = 23147)								
	Block Design	.264	-.022	-0.019	0.151	-0.018	-0.123	.903	0.359	.523
	Vocabulary	-.033	-.174	-0.160	0.163	-0.119	-0.981	.334	0.462	
	Motor RT	-.478	-.323	-0.009	0.004	-0.271	-1.898	.067	3.826	
	Primary Variables									
	Numb Comp	-.435	-.046	-7.606e -4	0.003	-0.041	-0.257	.799	0.533	
	Corsi	.154	.001	0.002	0.280	0.001	0.008	.993	0.387	
	RAN	-.413	-.186	-0.080	0.076	-0.134	-1.054	.300	0.889	
	R ILF HMOA	.437	.511	180.568	54.522	0.398	3.312	.002	11.644	
	R Fusiform V	.097	-.267	-1.845	1.196	-0.218	-1.543	.133	0.791	

	LG Postcentral C	.520	.486	8.127	2.623	0.370	3.098	.004	35.415	
	RS Insula C	.360	.321	6.809	3.603	0.233	1.890	.068	1.189	
	LS Orbital C	.285	.085	0.857	1.801	0.061	0.476	.637	0.467	
Division	**Control Variables**			**Most predictive model:** R ILF + LG Postcentral + RS Insula (BF_{10} = 423.553)						
	Block Design	.312	.094	0.140	0.266	0.088	0.527	.602	0.504	.358
	Vocabulary	-.067	-.168	-0.273	0.287	-0.133	-0.951	.349	0.624	
	Motor RT	-.322	-.217	-0.010	0.008	-0.205	-1.238	.225	0.665	
	Primary Variables									
	Numb Comp	-.242	.251	0.008	0.005	0.265	1.442	.159	0.496	
	Corsi	.223	-.046	-0.126	0.493	-0.046	-0.255	.800	0.428	
	RAN	-.316	-.144	-0.108	0.134	-0.119	-0.809	.425	0.688	
	R ILF HMOA	.488	.460	277.320	96.019	0.402	2.888	.007	14.705	
	R Fusiform V	.345	.133	1.573	2.107	0.123	0.747	.461	0.614	
	LG Postcentral C	.400	.323	8.779	4.620	0.263	1.901	.067	2.946	
	RS Insula C	.335	.310	11.517	6.345	0.260	1.815	.079	1.631	
	LS Orbital C	.264	.135	2.411	3.171	0.112	0.760	.453	0.604	
Mix	**Control Variables**			**Most predictive model:** RAN + R ILF + LG Postcentral + RS Insula (BF_{10} = 79258.76)						
	Block Design	.406	.134	0.144	0.191	0.105	0.755	.456	0.562	.547
	Vocabulary	.083	.043	0.049	0.206	0.028	0.239	.812	0.330	
	Motor RT	-.308	.108	0.003	0.006	0.084	0.605	.549	0.351	
	Primary Variables									
	Numb Comp	-.477	-.163	-0.003	0.004	-0.142	-0.919	.365	0.700	
	Corsi	.355	-.055	-0.108	0.354	-0.046	-0.306	.761	0.364	
	RAN	-.461	-.288	-0.161	0.096	-0.208	-1.677	.104	2.061	
	R ILF HMOA	.487	.482	211.079	68.957	0.358	3.061	.005	25.671	
	R Fusiform V	.392	.131	1.115	1.513	0.102	0.737	.467	0.522	
	LG Postcentral C	.451	.342	6.718	3.318	0.235	2.025	.052	3.115	
	RS Insula C	.460	.373	10.198	4.557	0.269	2.238	.033	6.431	
	LS Orbital C	.386	.204	2.647	2.277	0.144	1.162	.254	1.026	

Note: The provided R^2 values are adjusted R^2. All variables were simultaneously added to the model.

CHAPTER 7

General Discussion

An important aspect of arithmetic lies in the fact that arithmetic problems can be solved through various strategies, in which the development of automatization skills and fact retrieval from long-term memory are crucial for reaching the most efficient and fluent levels of arithmetic problem-solving. Although a vast body of behavioral literature exists on children's arithmetic fluency, its neural correlates are not completely understood. The overall aim of this doctoral project was thus to broaden our knowledge on both functional and structural neural correlates of arithmetic fluency. Doing so, we wanted to improve on the existing literature by complementing detected gaps and by using novel methods of analyzing the neural data, such as cortical complexity and spherical deconvolution analyses, which had never been used in research on children's arithmetic.

In this general discussion, a summary and discussion of the main findings of the abovementioned studies will be given, followed by a methodological reflection, and suggestions for future research.

7.1. Main findings and theoretical implications

7.1.1. Functional neural correlates

The first two studies of this doctoral dissertation aimed to gather knowledge on the functional neural correlates of children's arithmetic fluency, by identifying the brain regions that show increased activation for a fact retrieval strategy in comparison to a procedural strategy (*Chapter 2*), and by, within arithmetic fact retrieval, studying the neural basis of two established effects on arithmetic performance, namely the problem size and interference effect (*Chapter 3*).

In our first study (*Chapter 2*), we thus investigated the neural activation during children's arithmetic while taking individual differences in arithmetic strategy use into account. Previous research had already shown that brain activation in adults is modulated by the use of a fact retrieval or procedural strategy to solve arithmetic problems (Grabner et al., 2009; Tschentscher & Hauk, 2014), yet in children, only assumptions on strategy use were made, based on reaction time, problem size, operation, or presentation (De Smedt et al., 2011; Peters et al., 2016; Prado et al., 2014). As all problems of a particular size or operation are not necessarily solved through the same strategy, we applied a trial-by-trial approach to clearly investigate the neural differences between arithmetic fact retrieval and the use of procedural manipulations. Doing so, we looked at both multiplication and subtraction problems to see if previously observed operation differences (e.g., De Smedt et al., 2011; Prado et al., 2014) could also be found here, or if, similar to previous research on arithmetic strategies in adults (Tschentscher & Hauk, 2014), activation differences between operations disappear when taking strategy use into account.

During fact retrieval, we observed increased activation in the supramarginal and angular gyri, middle temporal gyri and frontal pole, which first of all concurs with the previous adult studies, where the angular gyri were shown to be related to fact retrieval strategy use. Similar to the data of Tschentscher and Hauk (2014), but in contrast to Grabner et al. (2009), this activation was found bilaterally. Extending the results of the fMRI studies in children that could only make assumptions on strategy use based on reaction time, problem size, operation, or presentation, we found stronger activation for retrieval in the middle temporal gyrus, observed during small multiplication items in Prado et al. (2014), and the bilateral angular and supramarginal gyri, observed during subtraction in a symbolic format in Peters et al. (2016). In contrast to our expectations and to previous developmental fMRI studies (i.e., De Smedt et al., 2011; Qin et al., 2014), our analyses did not reveal strong increases in hippocampal activation during fact retrieval (increased activation was only observed with a less stringent – i.e., FDR – correction for multiple comparisons). The longitudinal study by Qin et al. (2014), however, did point out that hippocampal engagement during arithmetic problem solving increases initially during childhood and subsequently decreases, reaching adult-like levels by adolescence. The weaker hippocampal activation found in our study could thus mean that the children of our research sample had already started the neural shift for arithmetic fact retrieval, from the hippocampus towards inferior parietal regions such as the angular gyrus (Grabner et al., 2009; Tschentscher & Hauk, 2014). Together, our results thus indicate the importance of both middle temporal and inferior parietal brain regions for arithmetic fact retrieval.

As, when using a procedural strategy during assessment, children were also asked to explain what their exact strategy was, this fMRI study also aimed to investigate neural activation during different possible procedural strategies. However, the children in our sample almost exclusively implemented a decomposition strategy for the procedural items. This was most likely due to the fact that our participants all came from Flemish primary schools, which put a high emphasis on decomposition strategies as being the most effective for solving problems when fact retrieval is not applicable. Accordingly, our study could only make claims on decomposition and not on any other procedural strategies, such as backwards counting for subtraction or repeated addition for multiplication, which should be investigated further in future research, as different neural activation might be observed across different procedural strategies. For the decomposition of operands strategy, we observed increased activation in a fronto-parietal network, including the bilateral inferior to superior parietal lobes, the inferior to superior frontal gyri, and bilateral areas in the occipital lobe and insular cortex. These results largely agree with the previous studies in adults, and again extend the studies in children that manipulated problem size, operation, or presentation format (De Smedt et al., 2011; Peters et al., 2016; Prado et al., 2014). Furthermore, this network seems to coincide with the multiple-demand network (Fedorenko et al., 2013), which, in adults, shows increases of activation for any kind of cognitive demand, independent of the content of the task. Accordingly, this fronto-parietal activation could also be explained by the inevitable association between strategy use and the task load of the procedural items, which is also evidenced by the increased reaction time for these items.

Most interestingly, our results did not reveal any activation differences between operations when taking strategy use into account, clearly indicating that arithmetic strategy rather than operation modulates brain activity. This in turn echoes the results of Tschentscher & Hauk (2014) in adults and emphasizes that assumptions on strategy use should not made based on operation alone. Accordingly, future imaging studies on arithmetic need to avoid indisputably linking certain operations to problem-solving strategies (e.g., multiplication to fact retrieval or subtraction to procedures), but instead consistently assess strategy use on a trial-by-trial basis.

In our second fMRI study (*Chapter 3*), we aimed to investigate the neural correlates of two known effects that can influence performance in arithmetic fact retrieval. In a similar design as previous adult studies (De Visscher et al., 2015; De Visscher et al., 2018), we studied the problem size effect, which states that more accurate and faster performance is often observed for small problems in comparison to large problems (De Brauwer et al., 2006), and the interference effect, which states that the quality of memory representations of multiplication problems depends on the problems that were previously learned, and that the more similar a problem is to previously learned problems, the more proactive interference will impact on encoding, leading to worse performance (De Visscher & Noël, 2014b). The problem size effect has been well established in children at both a behavioral (e.g., De Brauwer et al., 2006) and neural (e.g., De Smedt et al., 2011) level, but the interference effect is more recent and had never been studied in children at the neural level. Behaviorally, however, previous studies observed that the interference effect determines a substantial part of performance beyond the problem size effect (De Visscher & Noël, 2014a, 2014b; De Visscher et al., 2016). By investigating both the interference and problem size effect at the neural level in typically developing children, we intended to get a more detailed view of the brain network responsible for arithmetic fact retrieval, and to determine the specific functions of the relevant neural regions more precisely.

At the behavioral level, our results corroborated previous research findings (De Visscher and Noël, 2014b; De Visscher et al., 2016) and clearly showed that both problem size and interference level affect performance in multiplication, as both larger problems (i.e., problems with a product above 25, such as 6 × 7) and high interfering problems (i.e., problems with an interference level of 8 or higher, such as 4 × 6) were found to take significantly longer to solve. For accuracy, also in accordance with previous studies, no clear effects of problem size or interference level could be found, which, in the current sample, could be due to a ceiling effect on the multiplication task. The results from our imaging analyses, however, brought forth a different story. We observed clear neural differences between small and large problems; agreeing with previous studies (e.g., De Smedt et al., 2011; Molko et al., 2003; Stanescu–Cosson et al., 2000), increased activation for large items was observed in the bilateral fusiform gyri, the left superior parietal lobe, the left precentral gyrus, and the occipital cortex. However, no clear neural differences were observed for low and high interfering items, contrasting previous research in adults (De Visscher et al., 2015; De Visscher et al., 2018). In our univariate analyses, for example, no activation differences between low and high interfering problems were found when correcting for multiple comparisons, and our statistical pattern recognition analysis, which allows for the detection of differences between conditions with higher sensitivity than the conventional univariate analyses, also did not provide us with clear results.

The observed inconsistency across these results in children and previous results in adults, and across our behavioral and neural results, could be due to the notion that, as arithmetic development progresses, more neural distinctions are made between problems, based on the degree to which they are similar to previously learned problems. Accordingly, as the fourth graders of the current sample are still frequently in contact with these multiplication items in school, it is possible that clear neural distinctions are not yet made at this stage of development. Furthermore, it is possible that the used imaging modality (fMRI) was not perfectly suited to study the neural basis of the interference effect, as behavioral differences between low and high interfering items were only observed for reaction time, and as fMRI has a rather poor temporal resolution (Kim, Richter, & Uğurbil, 1997). Data collection methods with a greater temporal resolution such as electroencephalography (EEG) might have been more suitable, by focusing on the speed to which certain brain regions respond to either condition. However, because of the lower spatial resolution of EEG, its implementation would lead to other issues in localizing the involved brain regions (Burle et al., 2015). In all, our results indicate that the neural basis of the interference effect is not as strong in children as was previously observed in adults, but might develop over time (De Visscher et al., 2015; De Visscher et al., 2018), even though the research paradigm in our study was highly similar to that of the previous adult studies, and is not as strong as the problem size effect, which was clearly observed here.

7.1.2. Structural neural correlates

Two studies were also performed on the structural neural correlates of children's arithmetic fluency, looking at the volume and cortical complexity of theoretically relevant grey matter regions observed in previous arithmetic research (*Chapter 4*), and the structural integrity of white matter connections (*Chapter 5*).

In *Chapter 4*, we thus investigated structural properties of grey matter regions that were previously found to be related to arithmetic (Arsalidou et al., 2018; Peters and De Smedt, 2018). Previous studies had already used voxel-based morphometry and found an association of arithmetic and grey matter volume in the superior parietal lobe, the intraparietal sulcus, the inferior and middle frontal gyrus, the fusiform gyrus, and the hippocampus (see Peters and De Smedt, 2018 for an overview), but often used research samples with wide age ranges, and only took grey matter volume into account, disregarding other structural properties. Our study, however, used a research sample with a narrow age range (9- to 10-year-olds) and went beyond looking at volume alone, by also studying cortical complexity through fractal dimensionality. This cortical complexity looks at differences in the shape rather than the size of cortical structures (Yotter et al., 2011), and is highly sensitive for capturing relations between brain structure and cognitive function (King, et al., 2010; Im et al., 2006; Mustafa et al., 2012; Sandu et al., 2014). Very recently, within the field of mathematics, cortical complexity was used to study group differences in adults with dyslexia, dyscalculia, both disorders, and controls, yet revealed no evidence for differences in grey matter complexity associated with either dyslexia, dyscalculia, or their comorbid manifestation (Moreau, Wiebels, Wilson, & Waldie, 2019). However, as general associations had previously been observed between cognitive function and cortical complexity, we aimed to investigate if specific associations between arithmetic fluency and the cortical complexity of regions in children's arithmetic brain network were to be found, consequently aiming to provide a more comprehensive depiction of the structural neural correlates of children's arithmetic.

First of all, our results indicated associations between arithmetic fluency and the volume of the right fusiform gyrus, aligning with previous research where reduced volume of the fusiform gyrus was observed for children with developmental dyscalculia (Rykhlevskaia et al., 2009). The fusiform gyrus has previously been suggested to play a role in encoding complex visual stimuli (i.e., the recognition and discrimination of number-letter strings; Allison et al., 1999; Binder et al., 2006; Milner and Goodale, 2008), but has also been indicated to have a role in the early identification of problem difficulty, beyond mere digit recognition (Pinheiro-Chagas et al., 2018).

Second, our results pointed towards associations between arithmetic fluency and the cortical complexity of the left postcentral gyrus, right insular sulcus, and left orbital sulcus. The postcentral gyrus is adjacent to the superior and inferior parietal lobe, and thus also the intraparietal sulcus, and might act as an extension of its adjacent number-related regions. Here, it thus became apparent that the shape of the postcentral gyrus, which possibly affects, or is affected by, the superior and/or inferior parietal lobe, is a key correlate for typically developing children's arithmetic fluency. Furthermore, the postcentral gyrus was previously linked to the use of arithmetic strategies, such as subvocalization and finger counting (Kesler et al., 2006), and has been related to non-symbolic numerical and spatial processing (Arsalidou et al., 2018; Kaufmann et al., 2008). Accordingly, our study suggests an extension of the region's purpose for arithmetic fluency. The right insular cortex and left orbital sulcus, on the other hand, were previously associated with directing attentional resources and decision-making (Arsalidou and Taylor, 2011; Menon, 2015; Supekar and Menon, 2012). The function of these regions is thus most likely not arithmetic-specific, making our study highlight the importance of the directing of such attentional resources for arithmetic fluency. The right insula specifically has also been identified as a key region in emotional processing (Damasio et al., 2000), speech-motor function (Fox et al., 2001), and specific phobias, including mathematics anxiety (Lyons & Beilock, 2012).

Interestingly, our study did not find any associations, and even found evidence for no association, between the volume or cortical complexity of previously reported number and arithmetic brain regions, such as the superior and inferior parietal lobes. This discrepancy with previous studies could be due to various reasons, such as the wide age ranges used in previous studies (e.g., 8- to 14-year-olds), the different assessment of arithmetic (e.g., arithmetic fluency vs. broader mathematical reasoning), or the notion that the structural importance of the intraparietal sulcus might only become apparent when comparing children with extremely low arithmetic fluency or developmental dyscalculia to typically developing peers. Nonetheless, our study did direct the attention to the importance of other, less often discussed cortical regions for arithmetic fluency, as well as the importance of structural neural correlates other than volume.

As proper white matter connections between cortical regions are also crucial for cognitive development, our second structural MRI study (*Chapter 5*) aimed to investigate the associations of white matter tracts and individual differences in typically developing children's arithmetic fluency. These associations had already been studied in previous research, implicating a variety of tracts as relevant for arithmetic (Matejko and Ansari, 2015). However, these studies all applied classic diffusion tensor imaging (DTI), which is subject to methodological limitations, such as the problem with crossing fibers (e.g., Assaf et al., 2004; Dell'Acqua et al., 2013; Farquharson et al., 2013; Tournier et al., 2007). Furthermore, these studies were often conducted in research samples with very wide age ranges, such as 10- to 15-year-olds, which might lead to maturational confounds, leaving it unresolved to which extent previously found associations are due to maturation or to actual individual differences in performance. Our study aimed to take these limitations into account by using spherical deconvolution, a novel imaging technique that can characterize the orientation of more than one fiber per voxel and from which the hindrance modulated orientational anisotropy (HMOA) index can be derived, to study white matter pathways in a sample of 9- to 10-year-olds.

Results mainly pointed towards an association of arithmetic fluency and the right inferior longitudinal fasciculus (ILF), echoing results of previous studies where associations between fractional anisotropy in the ILF and individual differences in mathematics were observed (Li et al., 2013; Navas-Sánchez et al., 2014; Rykhlevskaia et al., 2009; Van Eimeren et al., 2008). As the ILF connects the occipital lobe to the anterior part of temporal lobe, including the fusiform gyri and parahippocampal regions, this association could be related to the efficiency with which children process Arabic numerals (Dehaene et al., 2003; Shum et al., 2013). Accordingly, these results can also be linked to our results in *Chapter 4*, where the importance of the fusiform gyrus was emphasized for its involvement in the processing of visual representations of numerical symbols. As mentioned, however, the inferior temporal gyrus has also been found to be driven by broader mathematical processing, instead of having a specific preference to Arabic numbers (Grotheer et al., 2018), which might also explain the ILF's relevance within arithmetic. Furthermore, as the ILF is important for the interaction between medial, inferior and anterior temporal cortices with Perisylvian areas, and as such is related to language (Catani & Mesulam, 2008), the tract could also be important for exact verbal arithmetic skills (e.g., fact retrieval), as it could act as a first step in connecting ventral stream areas upwards to the dorsal visual stream, via other tracts such as the arcuate fasciculus (AF) or superior longitudinal fasciculus (SLF), eventually to the intraparietal sulcus and superior parietal lobe (Rykhlevskaia et al., 2009). Accordingly, this result can also be linked to our results in *Chapter 2,* where middle temporal regions were found to be increasingly activated during self-reported fact retrieval. This is especially relevant when considering the structure of the Tempo Test Arithmetic (de Vos, 1992), which was used as a measure of arithmetic fluency, and is mainly constituted of single-digit items likely to be solved through fact retrieval.

To a lesser extent, correlations were also observed with the HMOA values of the right uncinate fasciculus, which connects the lateral orbitofrontal cortex with the anterior temporal lobes, and might be relevant for temporal lobe-based mnemonic associations (i.e., it might have an assisting role within memory and be relevant for automated processes in arithmetic; Von Der Heide et al., 2013). Taken together, the study thus again highlighted the importance of temporal brain structures for children's arithmetic fluency.

Notably, our study did not replicate other previously found associations between white matter tracts (e.g., AF and SLF) and arithmetic, and even pointed towards, albeit not necessarily substantial, evidence for the null hypothesis of no association between arithmetic and these tracts. This absence of associations, however, could be explained by various factors, such as the use of samples with wide age ranges in previous studies, possibly leading to maturational confounds, the use of group comparisons or the specificity of the tasks under study, such as broad mathematical reasoning tests (e.g., Van Eimeren et al., 2008) versus the more specific timed arithmetic test used here, or the implementation of classic DTI, which, in combination with the abovementioned wider age ranges and different mathematical tasks, might have led to the analyses not being powerful enough to consistently detect relationships with the right ILF, or to observe relationships with other tracts such as the AF or SLF. Accordingly, it would be interesting to do similar research across different ages and with a variety of mathematical measures, to clearly define which white matter pathways are related to which mathematical tasks across development.

7.1.3. Predictive value of brain imaging measures

Finally, in *Chapter 6*, we aimed to investigate the added value of the structural brain imaging measures observed in *Chapters 4 and 5* over well-known behavioral predictors of arithmetic, in predicting typically developing children's arithmetic fluency. This was done by first performing a correlation analysis between various behavioral measures and arithmetic fluency, which, in agreement with previous behavioral studies, displayed significant associations between arithmetic and symbolic numerical magnitude processing (Ansari, 2008; De Smedt et al., 2013; Schneider et al., 2017), working memory (Peng et al., 2016; Raghubar et al., 2010; Swanson & Jerman, 2006), and rapid automatized naming (Chard et al., 2005; Hecht et al., 2001; Koponen et al., 2006; Landerl et al., 2004; Mazzocco & Grimm, 2013).

Next, separate multiple regression models were calculated for these behavioral measures and for the observed structural grey and white matter imaging measures that showed at least substantial evidence for an association with arithmetic fluency. When putting all behavioral predictors into a regression model simultaneously, symbolic number comparison and rapid automatized naming emerged as crucial predictors of performance, stressing the importance of symbolic numerical magnitude processing as a domain-specific predictor (De Smedt et al., 2013; Schneider et al., 2017), but also of rapid automatized naming as a domain-general predictor, highlighting the importance of rapidly accessing phonological information in long-term memory for children's arithmetic (Chard et al., 2005; Hecht et al., 2001; Koponen et al., 2006; Landerl et al., 2004; Mazzocco & Grimm, 2013). When putting all previously observed brain imaging measures into a regression model simultaneously, the cortical complexity of the left postcentral gyrus, highlighting its possible importance as an extension of the superior and inferior parietal lobes and the intraparietal sulcus, and the white matter integrity of the right ILF, highlighting its possible importance for efficiently processing Arabic numerals and for exact verbal arithmetic skills, emerged as the most important predictors for children's arithmetic fluency.

All significantly correlated predictors, both behavioral and brain imaging, were then added to a multiple regression model together, which indicated that the brain imaging measures, especially the cortical complexity of the left postcentral gyrus and the HMOA values of the right ILF, are stronger predictors of arithmetic fluency than any of the behavioral measures. To definitively assess the added value of these brain imaging measures, we then performed a hierarchical Bayesian regression with control and behavioral measures added to the model first, then adding the brain imaging measures and calculating the BF_{10} of the best predictive model to indicate the increase in likeliness of predicting arithmetic fluency on top of the control and behavioral measures. Results of this analysis indicated a strong increase in the likelihood of predicting arithmetic fluency when adding the brain imaging measures to the behavioral and control measures, clearly emphasizing the importance of these neurophysiological measures. When adding the behavioral measures to the brain imaging measures, however, a reverse effect (i.e., worse prediction of arithmetic fluency) was observed.

These results are in accordance with previous similar research in the field of reading (Hoeft et al., 2011) as the observed structural neuroimaging measures provided a clear unique prediction (for children's arithmetic fluency), yet also differ from this research, as, in the study by Hoeft et al. (2011), the combination of behavioral and neuroimaging measures provided the best prediction, whereas here, the neuroimaging measures alone provided the best prediction. This is, however, similar to other research in reading (Vanderauwera, Wouters, Vandermosten, & Ghesquière, 2017), where cognitive, familial risk and neuroanatomical predictors of dyslexia were combined into a model to define the strongest unique predictor of dyslexia, which pointed towards a white matter measure (i.e., the FA values of the long segment of the arcuate fasciculus) as being the only significant predictor. Even though these results thus clearly show that the structural neuroimaging measures predict children's arithmetic fluency better than the cognitive predictors, we in no way mean that these neuroimaging measures by themselves have sufficient value and should be implemented for practical predictions or educational interventions on their own. Furthermore, it is perfectly possible that other cognitive predictors that were not assessed in this study do have additional predictive value on top the other behavioral and brain imaging measures. One such example could lie in the domain-general predictor inhibition, referring to the ability to suppress distracting information and inappropriate responses (Bull & Lee, 2014), which has been shown to be a significant predictor of mathematics after controlling for other executive functions, such as working memory (Espy et al., 2004). We, however, do feel that our results stress the importance of brain imaging measures, and suggest that a combination of behavioral and brain imaging measures holds potential for understanding the biological processes that play a role for educationally relevant skills (Black et al., 2015; De Smedt, 2018a). This, in turn, could assist in predicting educational outcomes, and in improving the specificity and effectiveness of educational interventions, as they could be followed-up at a neural level (De Smedt, 2018a; Hoeft et al., 2011; Howard-Jones et al., 2016; Supekar et al., 2013).

7.2. Methodological reflection

The current doctoral dissertation presented novel findings on the functional and structural neural correlates of children's arithmetic fluency, and used a variety of methods to do so. Accordingly, it is crucial to reflect upon both strengths and limitations of the implemented methodology, which will be discussed in the next few paragraphs.

First and foremost, in order to study the structural neural correlates of children's arithmetic fluency, we aimed to go beyond the existing literature, by implementing novel methods of analyzing the neural data. This becomes clear in the T1-imaging study in *Chapter 4*, where the structural correlates of children's arithmetic were not only analyzed through voxel-based morphometry, as was done in previous studies (Evans et al., 2015; Isaacs et al., 2001; Li et al., 2013; Price et al., 2016; Ranpura et al. 2013; Rotzer et al. 2008; Rykhlevskaia et al. 2009; Supekar et al., 2013), but also through cortical complexity, quantified by the fractal dimensionality index. This cortical complexity was previously used to study differences in cognitive function (King, et al., 2010; Im et al., 2006; Mustafa et al., 2012; Sandu et al., 2014), and most recently even to study group differences in adults with dyslexia, dyscalculia, both disorders, and controls (Moreau et al., 2019). Basically, cortical complexity is a method of doing surface-based morphometry, in which analyses are done on a segmentation of the cortical surface, which quantifies the spatial frequency of gyrification and fissuration of the brain surface (Luders et al., 2004). The fractal dimensionality index specifically was originally designed to quantify the structure of fractals (Kiselev et al., 2003; Yotter et al., 2011), but can be used to characterize cortical surface shape, and makes it possible for the complexity analysis to determine how periodically spaced and space-filling the fractal surface is. Doing so, the study in *Chapter 4* was thus able to provide novel insights into the structural neural correlates of children's arithmetic, by implementing a method that is sensitive to differences in grey matter structure that are not indexed by volume. By thus combining both voxel-based morphometry and cortical complexity analyses, we aimed to get a more holistic view on the structural grey matter correlates of children's arithmetic fluency. However, as cortical complexity is focused around the shape of cortical structures, a gap in this research lies in the fact that subcortical regions were not studied. The basal ganglia, for example, are subcortical brain regions that are part of procedural memory systems, which create a hierarchy of short-term representations that allow the manipulation of multiple discrete quantities. As such, the basal ganglia become more engaged depending on task complexity, with increased engagement as procedural memory requirements increase (Menon 2015). The basal ganglia are thus known to be important for the regulation of procedural memory (Chang, Crottaz-Herbette, Menon, 2007; Graybiel, 2005; Packard & Knowlton, 2002), making the study of structural characteristics of these regions important for providing additional insights into the neural basis of arithmetic fluency, but which was not possible for this type of analysis.

Such novel insights through the implementation of innovative imaging techniques are also present in the dMRI study in *Chapter 5*, in which spherical deconvolution was used to correlate the quality of white matter pathways to individual differences in children's arithmetic fluency, instead of classic DTI, which was implemented in the previous studies on the white matter correlates of children's mathematical abilities (Matejko and Ansari, 2015). It is true that our DTI and spherical deconvolution analyses led to similar results, bearing the question why spherical deconvolution should be preferred above classic DTI. However, being able to characterize the orientation of more than one fiber per voxel, spherical deconvolution provides a direct and more accurate (through the use of the more specific HMOA index; Dell'Acqua et al., 2007; Tournier et al., 2004) estimate of the underlying distribution of fiber orientations (Dell'Acqua & Tournier, 2017), emphasizing its usefulness as an essential tool for the study of the human brain. Nevertheless, the spherical deconvolution analysis used in this study has limitations as well, such as the fact that it assumes a zero-mean Gaussian distribution of underlying noise, even though this actual distribution is known to be non-Gaussian. This issue, however, could be resolved by other novel imaging techniques, such as diffusional kurtosis imaging (DKI), which extends classic DTI by estimating the kurtosis (i.e., a statistic for the skewed or non-Gaussian distribution) of the water diffusion probability distribution function. This could be important as water diffusion in biological tissues is non-Gaussian due to the effects of cellular microstructure, which is evident in the brain, where water diffusion is strongly restricted by myelinated axons (Steven, Zhuo, & Melhem, 2014; Wu & Cheung, 2010). Keeping all advantages and limitations in mind, the current doctoral dissertation thus strongly emphasizes and recommends the application of such novel data analysis methods in order to improve on the existing literature and broaden our knowledge on the neural correlates of various cognitive skills.

Second, throughout the studies discussed above, we consistently used a sample of 9- to 10-year-old children (i.e., all 4^{th} graders). This age range was selected as a substantial amount of arithmetic knowledge is already automatized at this point in development. Furthermore, we chose this small age range to minimize possible maturational confounds. Many of the existing studies on the neural substrates of children's arithmetic used research samples with wide age ranges (e.g., 7- to 10- or 10- to 15-years-old), even though this period in time is characterized by large neural development (e.g., Barnea-Goraly et al. 2005). Consequently, although statistically controlled for, the observed correlations between individual differences in arithmetic and brain structure and function might still be swayed by maturation, instead of being purely related to arithmetic achievement. Furthermore, mathematical achievement also improves throughout child development, meaning that high experience-dependent plasticity can also be expected in the brain (e.g., Casey et al. 2006; Huber, Donnelly, Rokem, & Yeatman, 2018). Taken together, we recommend that homogenous age groups (i.e., research samples with a small age range) should be studied in order to take maturation effects into account, as to not miss any important neurodevelopmental changes that occur during key stages of academic learning (Menon, 2015). The difficulty in this, however, lies in the notion that it is perfectly possible to collect data from a research sample with a small age range in which the desired results would be impossible to find, which might even be what happened in our interference study in *Chapter 3*. For example, it is possible that, as the fourth graders of our sample, in contrast to adults, were still frequently in contact with multiplication items, clear neural distinctions were not yet made, leading to the absence of the neural interference effect.

Third, it is important to stand still at how arithmetic was measured throughout this dissertation. In the fMRI studies, highly specific custom-made measures were used to study the neural basis of arithmetic problem-solving strategies (*Chapter 2*) and to study how well-known effects on arithmetic performance (more specifically on multiplication fact retrieval) affect activation in the arithmetic brain network (*Chapter 3*). Not only was strategy use assessed on a trial-by-trial basis in these tasks, they were also developed solely for these studies. In the structural studies (*Chapters 4 and 5*), however, arithmetic fluency was assessed in a more general fashion, through the Tempo Test Arithmetic (de Vos, 1992), meaning that these studies could not make specific statements on how the observed structural neural correlates of arithmetic fluency relate to such problem-solving strategies. The Tempo Test Arithmetic does mainly contain single-digit items, which are more likely to be solved through fact retrieval than through procedural manipulations, yet any statements on the relations between fact retrieval and these structural neural correlates would have been based on assumptions (mainly of problem size), which is exactly what we strongly advised against in *Chapter 2*. Accordingly, it was only possible for us to discuss the structural neural correlates of arithmetic fluency in general.

Fourth, this doctoral dissertation focused on children's arithmetic fluency, which is obviously time-sensitive, but implemented fMRI to study neural activation, which has a limited temporal resolution due to the intrinsic nature of the hemodynamic response function (i.e., the change in oxygenated blood flow that occurs after brain cells get activated to perform a task) that it is based on (Kim et al., 1997). This is especially relevant for the fMRI study on the interference effect in *Chapter 3*, for which the items of both interference conditions were highly similar (i.e., multiplication items of comparable problem size), were consistently solved through fact retrieval, and for which behavioral differences were only observed for reaction time. Consequently, instead of benefitting from the high spatial resolution of fMRI (Kim, Jin, & Fukuda, 2010), and looking at certain regions showing increased activation for one condition in comparison to the other, it might have been more suited to implement an imaging technique with a higher temporal resolution, such as event-related potentials (ERPs) in EEG (Burle et al., 2015). As these ERPs measure the actual direct electrophysiological response to a specific sensory, cognitive, or motor event, they can be valuable to better investigate the speed to which certain brain regions respond to certain stimuli, and could thus be used to study arithmetic fluency. For example, it is possible that an interference effect could have been found through ERPs when comparing the electrophysiological responses of a priori determined ROIs (e.g., based on previous research in adults; De Visscher et al., 2015; De Visscher et al., 2018) of low and high interfering multiplication problems. The downside of ERPs, however, then lies in the poorer spatial resolution the technique offers (Luck, 2005).

Fifth, a key aspect to keep in mind when analyzing MRI data regards the quality of the collected data, which can be strongly affected by motion. This is especially true for children, as children tend to move more than adults during data acquisition, possibly inducing undesirable noise in the neuroimaging data (Blumenthal, Zijdenbos, Molloy, & Giedd, 2002). Therefore, throughout the studies of this doctoral dissertation, we aimed to minimize motion by including a training session in a mock MRI scanner during behavioral assessment. Furthermore, during data analysis, we implemented strict criteria for motion correction and maintaining data quality. In the fMRI studies in *Chapters 2 and 3*, we implemented the rule that when participants displayed greater movement than the size of one voxel (i.e., 2.2 mm) on two consecutive images, only the items before the time point of excessive movement were included, and that, if a run did not contain at least one item for each condition the entire run was discarded (Vogel et al., 2016). For the structural grey matter analyses in *Chapter 4*, we aimed to maintain data quality by using the Mahalanobis distance, which combines weighted overall image quality, measuring noise and spatial resolution before preprocessing, and mean correlation, measuring the homogeneity of the data and thus the quality after preprocessing. Accordingly, we discarded data with a Mahalanobis distance greater than two standard deviations of the sample average. Finally, for the structural white matter analyses in *Chapter 5*, we discarded data for further analysis when, after motion correction, the data displayed excessive motion defined as a mean translation in any direction greater than the dMRI voxel size of 2.5 mm. Although the implementation of these criteria led to us losing a substantial amount of data, possibly leading to less statistical power for data analysis, the data quality criteria aimed to minimize noise and maximize data quality, actually giving rise to more accurate results. Of course, even more strict criteria could have been implemented, but we feel that, by using these criteria for motion and data quality, we achieved a proper balance between data quantity and quality.

Sixth, as mentioned, all research in the current doctoral dissertation was done with 9- to 10-year-old children. Accordingly, in order to ensure data quality and limit the amount of discarded data due to the abovementioned motion criteria, and to make the unusual experience of going through MRI scanning as pleasant as possible, we limited our scanning time to approximately 45-50 minutes per child (Ernst, Rumsey, & Munson, 2003). Unfortunately, due to this time constraint, it was impossible to incorporate extra runs into our imaging designs, which could have led to more statistical power to analyze the data. For example, in the fMRI study on the interference effect (*Chapter 3*), we were only able to include four fMRI runs per participant, instead of six, as was done previously in adults (De Visscher et al., 2015). This, of course, does not mean that the neural interference effect would have been found when six fMRI runs would have been used per participant, but it would have increased statistical power, giving us the opportunity to make more powerful statements about the collected neural data. Furthermore, we were also not able to add additional experiments to control for potentially confounding factors (e.g., on visual differences between conditions, which might have led to differences in occipital activation), or to implement runs with independent functional localizers based on previous research (e.g., to clearly localize the angular gyri for the interference study), in order to perform more powerful region of interest-based analyses.

Seventh, the functional neuroimaging data in this doctoral dissertation was consistently normalized to a standardized adult template. Even though this is in accordance with all previous neuroimaging studies on children's arithmetic that either normalized their data to MNI or Talairach space (Arsalidou et al., 2018), the use of such an adult template for children is not without problems, as children's brain structure does in fact differ from adults (e.g., Mills et al., 2016). A recent study (Phan et al., 2018a) even showed that using a pediatric atlas for 6-year-olds increases accuracy of segmentation for structural volume measurements. However, as most available pediatric templates focus on very young children (e.g., up to 4-year-olds; Phan, Smeets, Talcott, & Vandermosten, 2018b), at the time of data analysis, there were hardly any standardized pediatric templates available for 9- to 10-year-old children. This is why we mainly implemented whole-brain approaches for our fMRI studies, as it is difficult to anatomically localize specific regions in children based on region of interest software designed for adult brains.

Finally, a great strength of this doctoral project lies in the fact that, to perform statistical analyses, we used Bayesian statistics next to the more commonly used frequentist approach to statistics. In regard to the studies at hand, Bayesian statistics mainly have the advantage of being able to quantify the evidence that data provide for one hypothesis over another. Not only is this interesting to calculate the likelihood of the alternative hypothesis being true, Bayesian statistics, in contrast to classic frequentist hypothesis testing, are also able to quantify the evidence in favor of the null hypothesis when no association is observed (Andraszewicz et al., 2015). For example, this allowed us to, in *Chapter 5*, quantify evidence for the null hypothesis of no association between arithmetic and the AF and SLF, even though these white matter tracts were found to be related to children's arithmetic in previous research.

7.3. Ventures for future research

Although the results of the studies described above provide new insights in the functional and structural neural correlates of children's arithmetic, a lot remains not fully understood. In this section, we suggest possible ventures for future neuroimaging research in this field.

First, the structural studies in this doctoral dissertation used the Tempo Test Arithmetic (de Vos, 1992) to correlate children's arithmetic fluency to different components of neuroanatomy. Considering the structure of the Tempo Test Arithmetic (i.e., each column mainly contains single-digit items likely to be solved through fact retrieval – within the time limit, children often do not reach the larger items that are more often solved through procedural manipulations), only associations between brain structure and arithmetic fluency in general (however most likely based on fact retrieval) could be discussed. However, as we observed functional neural differences between the use of a fact retrieval or procedural strategy in *Chapter 2*, differences might also be observed in the structural neural correlates of both problem-solving mechanisms. Future research could thus use tasks specifically eliciting fact retrieval and procedural manipulations, preferably based on trial-by-trial self-reports (Siegler & Stern, 1998), in order to assess performance for both strategies and to more precisely study the structural neural correlates underlying arithmetic.

Second, all of the studies above were conducted with a research sample of 9- to 10-year-olds, to minimize maturational confounds. However, aside obvious neural development throughout childhood years (e.g., Barnea-Goraly et al., 2005), research has also implicated large development in arithmetic, with shifts in strategy use, especially in the frequency and efficiency of those strategies (Imbo & Vandierendonck, 2008; Siegler, 1996; Siegler et al., 1996; Vanbinst et al., 2015). Consequently, it is essential to highlight the need for similar functional and structural imaging studies in children of different ages, such as children in the early steps of their arithmetic development (i.e., first or second graders, who might show an increased involvement of hippocampal regions), or children who have reached more advanced levels of arithmetic (i.e., children in secondary school, whose arithmetic brain network might much closer resemble that of adults). Keeping this in mind, studies that longitudinally follow children throughout development could especially be informative to better understand how and when in time the neural networks for both retrieval and procedural strategies develop. Similarly, such functional longitudinal research might help to understand how and when the modulation by problem size and interference on the arithmetic fact retrieval network develops, as clear neural differences between low and high interfering items were observed in adults, but not in children. For example, it is possible that the neural effect of interference emerges at a later point in development, which can only be pinpointed by such longitudinal research. Furthermore, to clearly investigate if and how education-related development might affect both grey and white matter structures, or whether arithmetic-related structural neural differences are already present before the onset of formal arithmetic education, structural longitudinal research starting in preschool and following children throughout elementary school could also provide crucial insights. Longitudinal research is thus important, as through cross-sectional research, it is impossible to make strong statements on the causal link between brain structure and function, and arithmetic fluency (i.e., it is unclear whether individual differences in brain structure or function lead to differences in arithmetic performance, or vice versa). Based on a similar premise, another venture might lie in experimental research that manipulates the process of automatization and fact retrieval by teaching children to retrieve the answers to certain problems they would otherwise solve procedurally (e.g., single-double digit problems such as 3 × 14). Such research could then compare neural activation before and after the intervention, which would make it possible to see how automatization skills develop in the brain.

Third, as all research of this dissertation was done in participants with no history of learning difficulties, it would be highly interesting to do similar research in children with such difficulties, such as children with dyscalculia, who experience persistent deficits in acquiring basic mathematical competencies, or children with dyslexia, who experience severe impairments in reading (American Psychiatric Association, 2013). Such research would especially be interesting considering the fact that difficulties in arithmetic strategy use are considered the hallmark of dyscalculia, and considering that fact retrieval deficits have also been observed in children with dyslexia (Evans et al., 2015). Further research on the functional and neural correlates of arithmetic fluency in these atypical populations are thus deemed necessary, as they could provide additional insights into the neural correlates of arithmetic, and might be indicative for alternative remedial programs or mathematical instruction for children with learning disorders.

Fourth, throughout this doctoral dissertation, novel methods of analyzing structural neural data were used (i.e., cortical complexity and spherical deconvolution analyses). The benefits of these novel techniques were elaborately discussed in *Chapters 4 and 5*, yet, as we were among the first to implement these methods in the field of mathematical cognition and only looked at typically developing children, it stands to reason that future research should aim to apply similar methods in investigating adults or atypically developing children, where such techniques have not yet been used. Such studies could thus improve on the existing literature, by, among other things but in the example of our study in *Chapter 4*, regarding other structural characteristics than volume and their relation to arithmetic. Accordingly, this doctoral dissertation aims to emphasize how the implementation of novel data acquisition or analysis techniques can increase our understanding, not only of mathematics, but of the neural basis of various cognitive skills, be it reading, auditory processing, memory etc., and on clinical conditions characterized by deficits in those skills, such as dyslexia in the case of reading.

Finally, as mentioned above, the studies of this doctoral project were all performed with Flemish participants, all receiving formal schooling in the Flemish educational system, which strongly emphasizes the automatization of arithmetic facts, discourages, or even prohibits the use of counting strategies, and teaches the use of a decomposition of operands strategy as the most effective procedural strategy. Accordingly, it is crucial to highlight that arithmetic and the acquisition and use of arithmetic strategies is highly dependent on the math curriculum of the culture under study. For example, the emphasis on automatization may not be as strong in North America, where children are allowed to use counting strategies to solve arithmetic problems (e.g., Campbell & Xue, 2001). To this day, the neural correlates of these differences in arithmetic instruction remain unclear, but could be investigated through cross-cultural research designs. Future studies could thus explore how, for example, differences in the emphasis on fact retrieval or certain procedural strategies (i.e., different strategies, be it fact retrieval or procedural strategy use, or the use of different procedural manipulations, could be used for the same set of arithmetic problems) correlate with strategy-related brain activity. Similarly, from a perspective of experience-dependent neural plasticity, such studies could investigate if structural differences can be observed between children following different educational programs. For example, as the use of fact retrieval is emphasized to a much lesser extent in North America, results on arithmetic fluency tests such as the Tempo Test Arithmetic might actually correlate to structural characteristics of different neural regions, such as the intraparietal sulcus which is more involved in procedural manipulations. Similarly, such results might also correlate to the quality of other white matter pathways then the ILF, such as the arcuate fasciculus which connects the fronto-parietal brain regions of the procedural arithmetic brain network.

In all, we believe these suggestions will contribute to the existing body of literature, providing additional insights into the neural correlates of arithmetic. Only by longitudinally and cross-culturally studying arithmetic through specific tasks, and by implementing novel and more accurate neuroimaging techniques, can we gain a full overview of the neurobiological correlates that underlie this crucial skill.

REFERENCES

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex*, *9*, 415–430.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, MD: American Psychiatric Association.

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E. J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, *41*(2), 521-543.

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews. Neuroscience*, *9*(4), 278–291.

Ansari, D. (2010). Neurocognitive approaches to developmental disorders of numerical and mathematical cognition: The perils of neglecting the role of development. *Learning and Individual Differences*, *20*(2), 123–129.

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *Neuroreport*, *16*, 1769–1773.

Arsalidou, M., & Taylor, M. J. (2011). Is 2 + 2= 4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, *54*(3), 2328-2393.

Arsalidou, M., Pawliw-Levac, M., Sadeghi, M., & Pascual-Leone, J. (2018). Brain areas associated with numbers and calculations in children: Meta-analyses of fMRI studies. *Developmental Cognitive Neuroscience*, *30*, 239-250.

Arthurs, O. J., & Boniface, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? *Trends in Neurosciences*, *25*(1), 27–31.

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry: The methods. *NeuroImage*, *11*(6), 805-821.

Ashcraft, M. H., & Christy, K. S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in Grades 1-6. *Journal for Research in Mathematics Education*, *26*(5), 396–421.

Assaf, Y., Freidlin, R. Z., Rohde, G. K., & Basser, P. J. (2004). New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magnetic Resonance in Medicine*, *52*(5), 965–978.

Aydin, K., Ucar, K. K., Oguz, O. O., Okus, A., Agayev, Z., Unal, S., … Ozturk, C. (2007). Increased gray matter density in the parietal cortex of mathematicians: A voxel-based morphometry study. *American Journal of Neuroradiology*, *28*(10), 1859-1864.

Baddeley, A. D. (1986). *Working memory*. New York, NY: Oxford University.

Bailey, D. H., Littlefield, A., & Geary, D. C. (2012). The co-development of skill at and preference for use of retrieval-based processes for solving addition problems: Individual and sex differences from first to sixth grade. *Journal of Experimental Child Psychology*, *113*, 78–92.

Barnea-Goraly, N., Menon, V., Eckert, M., Tamm, L., Bammer, R., Karchemskiy, A., … Reiss, A. L. (2005). White matter development during childhood and adolescence: A cross-sectional diffusion tensor imaging study. *Cerebral Cortex*, *15*(12), 1848-1854.

Barrouillet, P., & Thevenot, C. (2013). On the problem-size effect in small additions: Can we really discard any counting-based account? *Cognition*, *128*(1), 35-44.

Barrouillet, P., Mignon, M., & Thevenot, C. (2008). Strategies in subtraction problem solving in children. *Journal of Experimental Child Psychology*, *99*(4), 233–251.

Basser, P. J., Mattiello, J., & LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, *66*(1), 259–267.

Ben-Shachar, M., Dougherty, R. F., & Wandell, B. A. (2007). White matter pathways in reading. *Current Opinion in Neurobiolgy*, *17*(2), 258–270.

Berteletti, I., Prado, J., & Booth, J. R. (2014). Children with mathematical learning disability fail in recruiting verbal and numerical brain regions when solving simple multiplication problems. *Cortex*, *57*, 143-155.

Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *NeuroImage*, *33*, 739–748.

Bishop, C. M. (2006). *Patter recognition and machine learning*. New York, NY: Springer.

Black, J. A., Myers, C. A., & Hoeft, F. (2015). The utility of neuroimaging studies for informing educational practice and policy in reading disorders. *New Directions in Child and Adolescent Development*, *147*, 49-56.

Blumenthal, J. D., Zijdenbos, A., Molloy, E., & Giedd, J. N. (2002). Motion artifact in magnetic resonance imaging: implications for automated analysis. *NeuroImage*, *16*, 89–92.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). *Region of interest analysis using an SPM toolbox*. Presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.

Brus, B. T., & Voeten, M. J. M. (1979). *Een Minuut Test (One minute Test)*. Lisse, The Netherlands: Swets & Zeitlinger.

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, *8*(1), 36-41.

Bulthé, J., De Smedt, B., & Op de Beeck, H. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage*, *87*, 311-322.

Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., & Vidal, F. (2015). Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *International Journal of Psychophysiology*, *97*(3), 210-220.

Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *The Quarterly Journal of Experimental Psychology*, *54*(4), 1005-1029.

Campbell, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network interference theory and simulation. *Mathematical Cognition*, *1*(2), 121–164.

Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, *130*(2), 299–315.

Cantlon, J. F., Libertus, M. E., Pinel, P., & Dehaene, S. (2009). The neural development of an abstract concept of number. *Journal of Cognitive Neuroscience*, *21*(11), 2217-2229.

Casey, B. J., Tottenham, N., Limston, C., & Durston, S. (2006). Imaging the developing brain: What we have learned about cognitive development. *Trends in Cognitive Sciences*, *9*(3), 104-110.

Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: History and current state. *Cortex*, *44*(8), 953-961.

Catani, M., & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, *44*(8), 1105–1132.

Catani, M., Jones, D. K., & ffytche, D.H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, *57*(1), 8-16.

Chang, C., Crottaz-Herbette, S., & Menon, V. (2007). Temporal dynamics of basal ganglia response and connectivity during verbal working memory. *NeuroImage*, *34*, 1253-1269.

Chard, D. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention*, *30*, 3–14.

Chilla, G. S., Tan, C. H., Xu, C., & Poh, C. L. (2015). Diffusion weighted magnetic resonance imaging and its recent trend-a survey. *Quantitative Imaging in Medicine and Surgery*, *5*(3), 407-422.

Cho, S., Ryali, S., Geary, D. C., & Menon, V. (2011). How does a child solve 7 + 8? Decoding brain activity patterns associated with counting and retrieval strategies. *Developmental Science*, *14*(5), 989-1001.

Cohen Kadosh, R., & Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences*, *32*, 313-373.

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, *84*(2), 132–147.

Corsi, P. M. (1972). *Memory and the medial temporal region of the brain* (Unpublished doctoral dissertation). McGill University, Department of Psychology, Montreal, Canada.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J., & Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, *3*, 1049–1056.

De Brauwer, J., Verguts, T., & Fias, W. (2006). The representation of multiplication facts: Developmental changes in the problem size, five, and tie effects. *Journal of Experimental Child Psychology*, *94*(1), 43–56.

De Smedt, B. (2016). Individual differences in arithmetic fact retrieval. In D. Berch, D. Geary, & K. Mann-Koepke (Eds.). *Mathematical cognition and learning* (Vol. 2; pp. 219-243). San Diego, CA: Elsevier Academic.

De Smedt, B. (2018a). Applications of cognitive neuroscience in educational research. In *Oxford Research Encyclopedia of Education*.

De Smedt, B. (2018b). Language and arithmetic: the potential role of processing. In A. Henik & W. Fias (Eds.), *Heterogeneity of function in numerical cognition* (pp. 51-74). San Diego, CA: Elsevier.

De Smedt, B., & Boets, B. (2010). Phonological processing and arithmetic fact retrieval: Evidence from developmental dyslexia. *Neuropsychologia*, *48*(14), 3973-3981.

De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology*, *108*(2), 278-292.

De Smedt, B., & Grabner, R. (2015). Applications of neuroscience to mathematics education. In A. Dowker & R. Cohen Kadosh (Eds.), The *Oxford handbook of numerical cognition* (pp. 613-636). Oxford: University Press.

De Smedt, B., Holloway, I. D., & Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. *NeuroImage*, *57,* 771-781.

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology*, *103*(2), 186-201.

De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*, 48-55.

De Smedt, B., Taylor, J., Archibald, L., & Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills? *Developmental Science*, *13*(3), 508-520.

De Visscher, A., & Noël, M.-P. (2013). A case study of arithmetic facts dyscalculia caused by a hypersensitivity-to-interference in memory. *Cortex*, *49*(1), 50–70.

De Visscher, A., & Noël, M.-P. (2014a). Arithmetic facts storage deficit: The hypersensitivity-to-interference in memory hypothesis. *Developmental Science*, *17*(3), 434-442.

De Visscher, A., & Noël, M.-P. (2014b). The detrimental effect of interference in multiplication facts storing: Typical development and individual differences. *Journal for Experimental Psychology: General*, *143*, 2380-2400.

De Visscher, A., Berens, S. C., Keidel, J. L., Noël, M.-P., & Bird, C.M. (2015). The interference effect in arithmetic fact solving: An fMRI study. *NeuroImage*, 116, 92-101.

De Visscher, A., Noël, M.-P., & De Smedt, B. (2016). The role of physical digit representation and numerical magnitude representation in children's multiplication fact retrieval. *Journal of Experimental Child Psychology*, *152*, 41-53.

De Visscher, A., Vogel, S. E., Reishofer, G., Hassler, E., Koschutnig, K., De Smedt, B., & Grabner, R. H. (2018). Interference and problem size effect in multiplication fact solving: Individual differences in brain activations and arithmetic performance. *NeuroImage*, *172*, 718-727.

de Vos, T. (1992). *Tempo-Test-Rekenen*. Nijmegen, The Netherlands: Berkhout.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*, 1-42.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*(1), 83-120.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*, 219–250.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3), 487–506.

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., & Benke, T. (2003). Learning complex arithmetic: An fMRI study. *Cognitive Brain Research*, *18*(1), 76-88.

Dell'Acqua, F., Rizzo, G., Scifo, P., Clarke, R. A., Scotti, G., & Fazio, F. (2007). A model-based deconvolution approach to solve fibercrossing in diffusion-weighted MR imaging. *IEEE Transactions on Biomedical Engineering*, *54*(3), 462–472.

Dell'Acqua, F., Simmons, A., Williams, S. C. R., & Catani, M. (2013). Can spherical deconvolution provide more information than fiber orientations? Hindrance modulated orientational anisotropy, a true-tract specific index to characterize white matter diffusion. *Human Brain Mapping*, *34*(10), 2464-2483.

Dell'Acqua, F., & Tournier, J.-D. (2018). Modelling white matter with spherical deconvolution: How and why? *NMR in Biomedicine*. doi:10.1002/nbm.3945

Desai, R., Liebenthal, E., Possing, E. T., Waldron, E., & Binder, J. R. (2005). Volumetric vs. surface-based alignment for localization of auditory cortex activation. *NeuroImage*, *26*, 1019-1029.

Diamond, A., & Amso, D. (2008). Contributions of neuroscience to our understanding of cognitive development. *Current Directions in Psychological Science*, *17*, 136-141.

Dick, A. S., & Tremblay, P. (2012). Beyond the arcuate fasciculus: Consensus and controversy in the connectional anatomy of language. *Brain*, *135*(12), 3529-3550.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*(3), 274-290.

Dowker, A. (2005). *Individual differences in arithmetic: Implications for psychology, neuroscience and education*. Hove, UK: Psychology Press.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475–483.

Epsy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2014). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology*, *26*(1), 465-486.

Ernst, M., Rumsey, J., & Munson, S. (2003). Update on functional neuroimaging in child psychiatry. In C. Fu, C. Senior, T. Russell, D. Weinberger, & R. Murray (Eds.), *Neuroimaging in Psychiatry* (pp. 51–80). Boca Raton: CRC.

Evans, T. M., Flowers, D. L., Napoliello, E. M., Olulade, O. A., & Eden, G. F. (2014). The functional anatomy of single-digit arithmetic in children with developmental dyslexia. *NeuroImage*, *101*, 644-652.

Evans, T. M., Kochalka, J., Ngoon, T. J., Wu, S. S., Qin, S., Battista, C., & Menon, V. (2015). Brain structural integrity and intrinsic functional connectivity forecast 6 year longitudinal growth in children's numerical abilities. *Journal of Neuroscience*, *35*, 11743–11750.

Farquharson, S., Tournier, J. D., Calamante, F., Fabinyi, G., Schneider-Kolsky, M., Jackson, G. D., & Connelly, A. (2013). White matter fiber tractography: Why we need to move beyond DTI. *Journal of Neurosurgery*, *118*(6), 1367-1377.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *PNAS*, *110*(41), 16616-16621.

Fox, P. T., Huang, A., Parsons, L. M., Xiong, J. H., Zamarippa, F., Rainey, L., & Lancaster, J. L. (2001). Location probability profiles for the mouth region of human primary motor-sensory cortex: Model and validation. *NeuroImage*, *13*, 196–209.`

Gaser, C., Nenadic, I., Buchsbaum, B., Hazlett, E., & Buchsbaum, M. S. (2001). Deformation-based morphometry and its relation to conventional volumetry of brain lateral ventricles in MRI. *NeuroImage*, *13*, 1140-1145.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool the number sets test. *Journal of Psychoeducational Assessment*, *27*(3), 265–279.

Geary, D. C., Bow-Thomas, C. C., & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology*, *54*(3), 372-391.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & Desoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, *88*(2), 121-151.

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11267–11271.

Glover, G. H. (2011). Overview of Functional Magnetic Resonance Imaging. *Neurosurgery Clinics of North America*, *22*(2), 133-139.

Goswami, U. (2008). Principles of learning, implications for teaching: A cognitive neuroscience perspective. *Journal of Philosophy of Education*, *42*, 381-399.

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., & Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia*, *47*, 604-608.

Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., & Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *NeuroImage*, *38*(2), 346-356.

Graybiel, A. M. (2005). The basal ganglia: Learning new tricks and loving it. *Current Opinion in Neurobiology*, *15*, 638-644.

Grotheer, M., Jeska, B., & Grill-Spector, K. (2018). A preference for mathematical processing outweighs the selectivity for Arabic numbers in the inferior temporal gyrus. *NeuroImage*, *175*, 188-200.

Han, B. S., Hong, J. H., Hong, C., Yeo, S. S., Lee, D. H., Cho, H. K., & Jang, S. H. (2010). Location of corticospinal tract at the corona radiata in human brain. *Brain Research*, *1326*, 75-80.

Han, Z., Davis, N., Fuchs, L., Anderson, A. W., Gore, J. C., & Dawant, B. M. (2013). Relation between brain architecture and mathematical ability in children: A DBM study. *Magnetic Resonance Imaging*, *31*, 1645–1656.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, *62*(2), 852-855.

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computational skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology*, *79*, 192–227.

Heirdsfield, A. M., & Cooper, T. J. (2002). Flexibility and inflexibility in accurate mental addition and subtraction: Two case studies. *Journal of Mathematical Behavior*, *21*, 57–74.

Hoeft, F., McCandliss, B. D., Black, J. M., Gantman, A., Zakerani, N., Hulme, C., . . . Gabrieli, J. D. E. (2011). Neural systems predicting long-term outcome in dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(1), 361-366.

Hoeft, F., Ueno, T., Reiss, A. L., Meyler, A., Whitfield-Gabrieli, S., Glover, G. H., … Gabrieli, J. D. E. (2007). Prediction of children's reading skills using behavioral, functional, and structural neuroimaging measures. *Behavioral Neuroscience*, *121*(3), 602-613.

Houdé, O., Rossi, S., Lubin, A., & Joliot, M. (2010). Mapping numerical processing, reading, and executive functions in the developing brain: an fMRI meta-analysis of 52 studies including 842 children. *Developmental Science*, *13*(6), 876–885.

Howard-Jones, P. A., Varma, S., Ansari, D., Butterworth, B., De Smedt, B., Goswami, U., . . . Thomas, M. S. C. (2016). The Principles and Practices of Educational Neuroscience: Comment on Bowers (2016). *Psychological Review*, *123*(5), 620-627.

Huber, E., Donnelly, P. M., Rokem, A., & Yeatman, J. D. (2018). Rapid and widespread white matter plasticity during an intensive reading intervention. *Nature Communications*. doi: 10.1038/P2theo0kz-018-04627-5

Huettel, S. A., Song, A. W., & McCarthy, G. (2014). *Functional Magnetic Resonance Imaging* (2nd ed.). Sunderland, MA: Sinauer Associates.

Im, K., Lee, J.-M., Yoon, U., Shin, Y.-W., Hong, S. B., Kim, I. Y., ... Kim, S. I. (2006). Fractal dimension in human cortical surface: Multiple regression analysis with cortical thickness, sulcal depth, and folding area. *Human Brain Mapping*, *27*(12), 994-1003.

Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology*, *96*, 284–309.

Imbo, I., & Vandierendonck, A. (2008). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: Differences between children and adults? *Psychological Research*, *72*(3), 331-346.

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011).Non-verbal number acuity correlates with symbolic math-emtics achievement: but only in children. *Psychonimic Bulletin & Review*, *18*(6), 1222-1229.

Isaacs, E. B., Edmonds, C. J., Lucas, A., & Gadian, D. G. (2001). Calculation difficulties in children of very low birthweight: A neural correlate. *Brain*, *124*, 1701-1707.

JASP Team (2017). *JASP* (Version 0.8.5) [Computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, United Kingdom: Oxford University.

Johansen-Berg, H. (2010). Behavioural relevance of variation in white matter microstructure. *Current Opinion in Neurology*, *23*(4), 351-358.

Jones, D. K., & Leemans, A. (2011). Diffusion tensor imaging. In M. Modo & J. W. M. Bulte (Eds.), *Magnetic resonance neuroimaging: Methods and protocols* (pp. 127-144). New York, NY: Humana.

Kaufmann, L., Vogel, S. E., Wood, G., Kremser, C., Schocke, M., Zimmerhackl, L.-B., & Koten, J. W. (2008). A developmental fMRI study of nonsymbolic numerical and spatial processing. *Cortex*, *44*(4), 376-385.

Kaufmann, L., Wood, G., Rubinsten, O., & Henik, A. (2011). Meta-analyses of developmental fMRI studies investigating typical and atypical trajectories of number processing and calculation. *Developmental Neuropsychology*, *36*(6), 763–787.

Keller, K., & Menon, V. (2009). Gender differences in the functional and structural neuroanatomy of mathematical cognition. *NeuroImage*, *47*, 342–352.

Kesler, S. R., Menon, V., & Reiss, A. L. (2006). Neurofunctional differences associated with arithmetic processing in Turner Syndrome. *Cerebral Cortex*, *16*(6), 849-856.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: The National Academies.

Kim, S. G., Richter, W., & Uğurbil, K. (1997). Limitations of temporal resolution in functional MRI. *Magnetic Resonance in Medicine*, *37*(4), 631-636.

Kim, S.-G., Jin, T., & Fukuda, M. (2010). Spatial resolution of fMRI techniques. In S. Ulmer & O. Jansen (Eds.), *fMRI: Basics and clinical applications* (pp. 15-21). Berlin, Germany: Springer.

King, R. D., Brown, B., Hwang, M., Jeon, T., & George, A. T. (2010). Fractal dimension analysis of the cortical ribbon in mild Alzheimer's disease. *NeuroImage*, *53*(2), 471-479.

Kiselev, V. G., Hahn, K. R., & Auer, D. P. (2003). Is the brain cortex a fractal? *NeuroImage*, *20*, 1765-1774.

Klein, E., Moeller, K., Glauche, V., Weiller, C., & Willmes, K. (2013). Processing pathways in mental arithmetic-evidence from probabilistic fiber tracking. *PLoS One*, *8*(1).

Koponen, T., Mononen, R., Räsänen, P., & Ahonen, T. (2006). Basic numeracy in children with specific language impairment: Heterogeneity and connections to language. *Journal of Speech, Language, and Hearing Research*, *49*, 58 –73.

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Educational Psychology*, *105*(1), 162-175.

Kucian, K., Ashkenazi, S. S., Hänggi, J., Rotzer, S., Jäncke, L., Martin, E., & von Aster,M. (2013). Developmental dyscalculia: A dysconnection syndrome? *Brain Structure and Function*, *219*(5), 1721-1733.

Kucian, K., von Aster, M., Loenneker, T., Dietrich, T., & Martin, E. (2008). Development of neural networks for exact and approximate calculation: A fMRI study. *Developmental Neuropsychology*, *33*(4), 447–473.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied multiple regression models* (4th ed.). New York, NY: McGraw-Hill.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition*, *93*, 99–125.

Lee, K. M. (2000). Cortical areas differentially involved in multiplication and subtraction: A functional Magnetic Resonance Imaging study and correlation with a case of selective acalculia. *Annals of Neurology*, *48*(4), 657–661.

Leemans, A., Jeurissen, B., Sijbers, J., & Jones, D. K. (2009). ExploreDTI: A graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In *Proceedings of the 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine* (p. 3537). Honolulu.

LeFevre, J. A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 216–230.

Li, Y., Hu, Y., Wang, Y., Weng, J., & Chen, F. (2013). Individual structural differences in left inferior parietal area are associated with schoolchildren's arithmetic scores. *Frontiers in Human Neuroscience*, *7*.

Lopes, R., & Betrouni, N. (2009). Fractal and multifractal analysis: A review. *Medical Image Analysis*, *13*, 634-649.

Luck, S. J. (2005). An introduction to the event-related potential technique. MIT.

Luders, E., Narr, K. L., Thompson, P. M., Rex, D. E., Jancke, L., Steinmetz, H., & Toga, A. W. (2004). Gender differences in cortical complexity. *Nature Neuroscience*, *7*(8), 799-800.

Lyons, I. M., & Beilock, S. L. (2012). When math hurts: Math anxiety predicts pain network activation in anticipation of doing math. *PLOS One*, *7*. doi:10.1371/journal.pone.0048076

Madan, C. R., & Kensinger, E. A. (2016). Cortical complexity as a measure of age-related brain atrophy. *NeuroImage*, *134*, 617-629.

Martino, J., Brogna, C., Robles, S. G., Vergani, F., & Duffau, H. (2010). Anatomic dissection of the inferior fronto-occipital fasciculus revisited in the lights of brain stimulation data. *Cortex*, *46*(5), 691–699.

Matejko, A. A., Price, G. R., Mazzocco, M. M. M., & Ansari, D. (2013). Individual differences in left parietal white matter predict math scores on the preliminary scholastic aptitude test. *NeuroImage*, *66*, 604–610.

Matejko, A., & Ansari, D. (2015). Drawing connections between white matter and numerical and mathematical cognition: A literature review. *Neuroscience & Biobehavioral Reviews*, *48*, 35-52.

Mazzocco, M. M. M., & Grimm, K. J. (2013). Growth in rapid automatized naming from grades K to 8 in children with math or reading disabilities. *Journal of Learning Disabilities*, *46*(6), 517-533.

McRobbie, D. W., Moore, E. A., Graves, M. J., & Prince, M. R. (2006). *MRI from picture to proton* (2nd ed.). Cambridge University.

Menon, V. (2015). Arithmetic in the child and adult brain. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 502-530). Oxford, United Kingdom: Oxford University.

Mills, K. L., Goddings, A.-L., Herting, M. M., Meuwese, R., Blakemore, S.-J., Crone, E. A., … Tamnes, C. K. (2016). Structural brain development between childhood and adulthood: Convergence across four longitudinal samples. *NeuroImage*, *141*, 273-271.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*, 774–785.

Moeller, K., Willmes, K., & Klein, E. (2015). A review on functional and structural brain connectivity in numerical cognition. *Frontiers in Human Neuroscience*, *9*.

Molko, N., Cachia, A., Riviere, D., Mangin, J.F., Bruandet, M., Le Bihan, D., … Dehaene, S. (2003). Functional and structural alterations of the intraparietal sulcus in a developmental dyscalculia of genetic origin. *Neuron*, *40*, 847–858.

Moreau, D., Wiebels, K., Wilson, A. J., & Waldie, K. E. (2019). Volumetric and surface characteristics of gray matter in adult dyslexia and dyscalculia. *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2019.02.002

Mustafa, N., Ahearn, T. S., Waiter, G. D., Murray, A. D., Whalley, L. J., & Staff, R. T. (2012). Brain structural complexity and life course cognitive change. *NeuroImage*, *61*, 694–701.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269.

Navas-Sánchez, F. J., Alemán-Gómez, Y., Sánchez-Gonzalez, J., Guzmán-De-Villoria, J. A., Franco, C., Robles, O., … Desco, M. (2014). White matter microstructure correlates of mathematical giftedness and intelligence quotient. *Human Brain Mapping, 35*(6), 2619–2631.

Ozernov-Palchik, O., & Gaab, N. (2016). Tackling the 'dyslexia paradox': Reading brain and behavior for early markers of developmental dyslexia. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(2), 156-176.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, *25*, 563-593.

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, *108*(4), 455-473.

Peters, L., & De Smedt, B. (2018). Arithmetic in the developing brain: A review of brain imaging studies. *Developmental Cognitive Neuroscience*, *30*, 265-279.

Peters, L., Polspoel, B., Op de Beeck, H., & De Smedt, B. (2016). Brain activity during arithmetic in symbolic and non-symbolic formats in 9-12 year old children. *Neuropsychologia*, *86*, 19-28.

Phan, T. V., Sima, D. M., Beelen, C., Vanderauwera, J., Smeets, D., & Vandermosten, M. (2018a). Evaluation of methods for volumetric analysis of pediatric brain data: The childmetrix pipeline versus adult-based approaches. *NeuroImage: Clinical*, *19*, 734-744.

Phan, T. V., Smeets, D., Talcott, J. B., & Vandermosten, M. (2018b). Processing of structural neuroimaging data in young children: Bridging thegap between current practice and state-of-the-art methods. *Developmental Cognitive Neuroscience*, *33*, 206-223.

Pinheiro-Chagas, P., Daitch, A., Parvizi, J., & Dehaene, S. (2018). Brain Mechanisms of arithmetic: A crucial role for ventral temporal cortex. *Journal of Cognitive Neuroscience*, *30*, 1757-1772.

Polspoel, B., Peters, L., Vandermosten, M., & De Smedt, B. (2017). Strategy over operation: Neural activation in subtraction and multiplication during fact retrieval and procedural strategy use in children. *Human Brain Mapping*, *38*(9), 4657-4670.

Polspoel, B., Vandermosten, M., & De Smedt, B. (2018). Relating individual differences in white matter pathways to children's arithmetic fluency: A spherical deconvolution study. *Brain Structure and Function, 224*(1), 337-350.

Polspoel, B., Vandermosten, M., & De Smedt, B. (*submitted*). The association of grey matter volume and cortical complexity with individual differences in children's arithmetic fluency.

Prado, J., Lu, J., Lio, L., Dong, Q., Zhou, X., & Booth, J. R. (2013). The neural bases of the multiplication problem-size effect across countries. *Frontiers in Human Neuroscience*, *7*.

Prado, J., Mutreja, R., & Booth, J. R. (2014). Developmental dissociation in the neural responses to simple multiplication and subtraction problems. *Developmental Science*, *17*(4), 537-552.

Prado, J., Mutreja, R., Zhang, H., Metha, R., Desroches, A. S., Minas, J. E., & Booth, J. R. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Human Brain Mapping*, *32*(11), 1932–1947.

Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica*, *140*, 50-57.

Price, G. R., Wilkey, E. D., Yeo, D. J., & Cutting, L. E. (2016). The relation between 1st grade grey matter volume and 2nd grade math competence. *NeuroImage*, *124*, 232–237.

Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D. C., & Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. *Nature Neuroscience*, *17*(9), 1263-1269.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual differences and cognitive approaches. *Learning and Individual Differences*, *20*, 110–122.

Raichle, M. E., Macleod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *PNAS, 98*(2), 676-682.

Ranpura, A., Isaacs, E., Edmonds, C., Rogers, M., Lanigan, J., Stinghal, A., … Butterworth, B. (2013). Developmental trajectories of grey and white matter in dyscalculia. *Trends in Neuroscience and Education*, *2*(2), 56-64.

Rickard, T. C., Romero, S. G., Basso, G., Wharton, C., Flitman, S., & Grafman, J. (2000). The calculating brain: An fMRI study. *Neuropsychologia*, *38*, 325–335.

Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex*, *15*(11), 1779–1790.

Robinson, C. S., Menchetti, B. M., & Torgesen, J. K. (2002). Toward a two-factor theory of one type of mathematics disabilities. *Learning Disabilities Research & Practice*, *17*(2), 81–89.

Rosenberg-Lee, M., Barth, M., & Menon, V. (2011). What difference does a year of schooling make? Maturation of brain response and connectivity between 2nd and 3rd grades during arithmetic problem solving. *NeuroImage*, *57*(3), 796-808.

Rosenberg-Lee, M., Tsang, J. M., & Menon, V. (2009). Symbolic, numeric, and magnitude representations in the parietal cortex. *Behavioral and Brain Sciences*, *32*, 350–351.

Rotzer, S., Kucian, K., Martin, E., Aster, M., von Klaver, P., & Loenneker, T. (2008). Optimized voxel-based morphometry in children with developmental dyscalculia. *NeuroImage*, *39*, 417-422.

Ruiz de Miras, J., Costumero, V., Belloch, V., Escudero, J., Avila, C., & Sepulcre, J. (2017). Complexity analysis of cortical surface detects changes in future Alzheimer's disease converters. *Human Brain Mapping*, *38*(12), 5905-5918.

Rykhlevskaia, E., Uddin, L. L. Q., Kondos, L., & Menon, V. (2009). Neuroanatomical correlates of developmental dyscalculia: Combined evidence from morphometry and tractography. *Frontiers in Human Neuroscience*, *3*.

Sandu, A.-L., Staff, R. T., McNeill, C. J., Mustafa, N., Ahearn, T., Whalley, L. J., & Murray, A. D. (2014). Structural brain complexity and cognitive decline in late life - A longitudinal study in the Aberdeen 1936 Birth Cohort. *NeuroImage*, *100*, 558-563.

Schmahmann, J. D., & Pandya, D. N. (2008). Disconnection syndromes of basal ganglia, thalamus, cerebrocerebellar systems. *Cortex*, *44*(8),1037-1066.

Schneider, M., Beeres, K., Coban, L. Merz, S., Schmidt, S., Stricker, J., & De Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, *20*(3).

Schrouff, J., Rosa, M. J., Rondina, J. M., Marquand, A. F., Chu, C., Ashburner, J., … Mourao-Miranda, J. (2013). *PRoNTo: Pattern Recognition for Neuroimaging Toolbox*. Neuroinformatics.

Shum, J., Hermes, D., Foster, B. L., Dastjerdi, M., Rangarajan, V., Winawer, J., … Parvizi, J. (2013). A brain area for visual numerals. *Journal of Neuroscience*, *33*(16), 6709–6715.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children’s addition. *Journal of Experimental Psychology General*, *116*(3), 250-264.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*, 258–275.

Siegler, R. S., & Shrager, J. (1984). Strategy choice in addition and subtraction: How do children know what to do? In C. Sophian (Ed.). *Origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, *127*(4), 377–397.

Siegler, R. S., Adolph, K. E., & Lemaire, P. (1996). Strategy choices across the life span. In L. R. Reder (Ed.), *Implicit memory and metacognition* (pp. 79–121). Mahwah, NJ: Erlbaum.

Siegler, R.S. (1996). *Emerging minds: The process of change in children’s thinking*. New York, NY: Oxford University.

Simmons, F. R., & Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? A review of research into arithmetic and dyslexia. *Dyslexia*, *14*(2), 77-94.

Simon, O., Mangin, J.-F., Cohen, L., Le Bihan, D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, *33*, 475-487.

Soares, J. M., Marques, P., Alves, V., & Sousa, N. (2013). A hitchhiker's guide to diffusion tensor imaging. *Frontiers in Neuroscience*, *7*, 31.

Stanescu–Cosson, R., Pinel, P., Moortele, P. F. V. D., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia. A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain*, *123*, 2240–2255.

Steven, A. J., Zhuo, J., & Melhem, E. R. (2014). Diffusion Kurtosis Imaging: An emerging technique for evaluating the microstructural environment of the brain. *American Journal of Roentgenology*, *202*(1), 26-33.

Supekar, K., Swigart, A. G., Tenison, C., Jolles, D. D., Rosenberg-Lee, M., Fuchs, L., & Menon, V. (2013). Neural predictors of individual differences in response to math tutoring in primary-grade school children. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(20), 8230-8235.

Supekar, K., Uddin, L. Q., Prater, K., Amin, H., Greicius, M. D., & Menon, V. (2010). Development of functional and structural connectivity within the default mode network in young children. *NeuroImage*, *52*(1), 290-301.

Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research*, *76*, 249–274.

Thevenot, C., Barrouillet, P., Castel, C., & Uittenhove, K. (2016). Ten-year-old children strategies in mental addition: A counting model account. *Cognition*, *146*, 48-57.

Thiebaut de Schotten, M., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D. G. M., & Catani, M. (2011). A lateralized brain network for visuospatial attention. *Nature Neuroscience*, *14*, 1245-1246.

Thompson, P. M., Lee, A. D., Dutton, R. A., Geaga, J. A., Hayashi, K. M., Eckert, M. A., … Reiss, A. L. (2005). Abnormal cortical complexity and thickness profiles mapped in Williams syndrome. *The Journal of Neuroscience*, *25*(16), 4146-4158.

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2004). Strategic aspects of simple addition and subtraction: The influence of mathematical ability. *Learning and Instruction*, *14*(2), 177-195.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, *27*(5), 276-286.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second- to fifth-grade children. *Scientific Studies of Reading*, *1*, 161–185.

Tournier, J. D., Calamante, F., & Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, *35*(4), 1459–1472.

Tournier, J. D., Calamante, F., Gadian, D. G., & Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, *23*(3), 1176–1185.

Tsang, J. M., Dougherty, R. F., Deutsch, G. K., Wandell, B. A., & Ben-Shachar, M. (2009). Frontoparietal white matter diffusion properties predict mental arithmetic skills in children. *PNAS*, *106*(52), 22546-22551.

Tschentscher, N., & Hauk, O. (2014). How are things adding up? Neural differences between arithmetic operations are due to general problem solving strategies. *NeuroImage*, *92*, 369-380.

Tuch, D. S. (2004). Q-ball imaging. *Magnetic Resonance in Medicine*, *52*(6), 1358–1372.

Uittenhove, K., Thevenot, C., & Barrouillet, P. (2016). Fast automated counting procedures in addition problem solving: When are they used and why are they mistaken for retrieval? *Cognition*, *146*, 289-303.

Van Beek, L., Ghesquière, P., Lagae, L., & De Smedt, B. (2014). Left fronto-parietal white matter correlates with individual differences in children's ability to solve additions and multiplications: A tractography study. *NeuroImage*, *90*, 117-127.

van den Bos, K. P. (1998) IQ, phonological awareness and continuous-naming speed related to Dutch poor decoding children's performance on two word identification tests. *Dyslexia*, *4*, 73–89.

van den Bos, K. P., Spelberg, H. C. L., Scheepstra, A. S. M., & De Vries, J. R. (1994). *De Klepel: Pseudowoordentest*. Nijmegen, The Netherlands: Berkhout.

Van Eimeren, L., Grabner, R. H., Koschutnig, K., Reishofer, G., Ebner, F., & Ansari, D. (2010). Structure-function relationships underlying calculation: a com-bined diffusion tensor imaging and fMRI study. *NeuroImage*, *52*(1), 358–363.

Van Eimeren, L., Niogi, S. N., McCandliss, B. D., Holloway, I. D., & Ansari, D. (2008). White matter microstructures underlying mathematical abilities in children. *Cognitive Neuroscience and Neuropsychology*, *19*(11), 1117-1121.

Vanbinst, K., & De Smedt, B. (2016). Individual differences in children's mathematics achievement: The roles of symbolic numerical magnitude processing and domain-general cognitive functions. *Progress in Brain Research*, *227*, 105-130.

Vanbinst, K., Ceulemans, E., Ghesquière, P., & De Smedt, B. (2015a). Profiles of children's arithmetic fact development: A model-based clustering approach. *Journal of Experimental Child Psychology*, *133*, 29-46.

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2012). Numerical magnitude representations and individual differences in children's arithmetic strategy use. *Mind, Brain and Education*, *6*(3), 129–136.

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2015b). Does numerical processing uniquely predict first graders' future development of single-digit arithmetic? *Learning and Individual Differences*, *37*, 153-160.

Vanderauwera, J., Vandermosten, M., Dell'Acqua, F., Wouters, J., & Ghesquière, P. (2015). Disentangling the relation between left temporoparietal white matter and reading: A spherical deconvolution tractography study. *Human Brain Mapping*, *36*(8), 3273-3287.

Vanderauwera, J., Wouters, J., Vandermosten, M., & Ghesquière, P. (2017). Early dynamics of white matter deficits in children developing dyslexia. *Developmental Cognitive Neuroscience*, *27*, 69-77.

Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nature Reviews Neuroscience*, *10*(11), 792-802.

Verschaffel, L., Torbeyns, J., De Smedt, B., Luwel, K., & Van Dooren, W. (2007). Strategy flexibility in children with low achievement in mathematics. *Educational and Child Psychology*, *24*(2), 16-27.

Vogel, S. E., Matejko, A. A., & Ansari, D. (2016). Imaging the developing human brain using functional and structural Magnetic Resonance Imaging: Methodological and practical guidelines. *Practical Research with Children*, 46–69.

Von Der Heide, R. J., Skipper, L. M., Klobusicky, E., & Olson, I. R. (2013). Dissecting the uncinate fasciculus: disorders, controversies and a hypothesis. *Brain*, *136*(6),1692-1707.

Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., … Mori, S. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *NeuroImage*, *36*(3), 630–644.

Wang, R., Benner, T., Sorensen, A. G., & Wedeen, V. J. (2007). Diffusion toolkit: A software package for diffusion imaging data processing and tractography. In *15th Annual Meeting, International Society for Magnetic Resonance in Medicine*. Berlin.

Wechsler, D. (2005). *Wechsler Intelligence Scale for Children – WISC-III-NL*. Amsterdam, The Netherlands: Pearson.

Wedeen, V. J., Hagmann, P., Tseng, W. Y. I., Reese, T. G., & Weisskoff, R. M. (2005). Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine*, *54*(6), 1377–1386.

Whitwell, J. L. (2009). Voxel-based morphometry: An automated technique for assessing structural changes in the brain. *Journal of Neuroscience*, *29*, 9661-9664.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2003). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Wu, E. X., & Cheung, M. M. (2010). MR diffusion kurtosis imaging for neural tissue characterization. *NMR in Biomedicine*, *23*, 836-848.

Wu, S., Chang, T. T., Majid, A., Caspers, S., Eickhoff, S. B., & Menon, V. (2009). Functional heterogeneity of inferior parietal cortex during mathematical cognition assessed with cytoarchitectonic probability maps. *Cerebral Cortex*, *19*, 2930–2945.

Yotter, R. A., Ziegler, G., Nenadic, I., Thompson, P. M., & Gaser, C. (2011). Local cortical surface complexity maps from spherical harmonic reconstructions. *NeuroImage*, *56*, 961-973.

Zago, L., Pesenti, M., Mellet, E., Crivello, F., Mazoyer, B., & Tzourio-Mazoyer, N. (2001). Neural correlates of simple and complex mental calculation. *NeuroImage*, *13*, 314–327.

Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds: The problem-size effect. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 331-345). New York, NY: Psychology.

Zhao, J., Thiebaut de Schotten, M., Altarelli, I., Dubois, J., & Ramus, F. (2016). Altered hemispheric lateralization of white matter pathways in developmental dyslexia: Evidence from spherical deconvolution tractography. *Cortex*, *76*, 51-62.

Zhou, X., Chen, C., Zang, Y., Dong, Q., Chen, C., Qiao, S., & Gong, Q. (2007). Dissociated brain organization for single-digit addition and multiplication. *NeuroImage*, *35*(2), 871-880.