The structure of the mathematical brain



Tudor Popescu

St Edmund Hall, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Trinity Term, 2014

Abstract

Humans have an innate ability to deal with numerosity and other aspects of magnitude. This ability is generally honed through education in and experience with mathematics, which necessarily changes the brain structurally and functionally. These changes can be further manipulated through non-invasive electrical brain stimulation. Studying these processes in the case of maths not only constitutes research of great practical impact – given the importance of numerical skills in today's society – but also makes use of maths as a suitable domain in which to study plasticity. In this thesis, I aimed to explore how expertise with numbers shapes brain and behaviour, and also the degree to which processing numbers is similar to other domains in terms of the necessity of healthy brain regions believed to underlie normal processing within and across these domains.

In Study 1, behavioural and structural brain differences were found cross-sectionally between mathematicians and non-mathematicians. A double dissociation between those groups was found between grey matter density in the frontal lobe and behavioural performance: their correlation was positive for mathematicians but negative for controls. These effects may have been caused by years of experience, by congenital predispositions, or, plausibly, by both of these factors, whose disambiguation is non-trivial. Study 2 used transcranial random noise stimulation (tRNS) to assist arithmetic learning. A novel montage was used to enhance brain function during the stage when it is believed to be most involved. Real as compared to sham tRNS enhanced reaction times (RTs) and learning rate on a calculation-based task, but not on a retrieval-based task. The effects were only observed in conditions of high task difficulty. Study 3 examined structural MRI measures before and after arithmetic training to determine how either frontal or parietal tRNS applied with the task changes the structure of the brain longitudinally as compared to sham. Previous results (including those of Study 2) of behavioural facilitation in terms of enhanced RTs to calculation problems were replicated, and further interpreted. Both frontal and parietal tRNS modulated the changes that occurred, pre-to-post training, in terms of cortical volume and gyrification of frontal, parietal and temporal areas. Study 4 investigated the shared neural and cognitive resources used for processing numerical magnitude and musical

pitch, by probing how stimulus-response compatibility (SRC) effects for each of the two dimensions compare in a group of mainly temporoparietal lesion patients with numerical impairments versus controls. A double dissociation was found in that numerically impaired patients did not show the number-based SRC effect but did show the pitch-based one, while control subjects demonstrated the opposite trend.

Overall, the results of these studies leave us with three main messages. First, expertise in numbers and mathematics, whether acquired through years of experience (Study 1) or through a few days of tRNS-assisted training (Study 3), appears to be associated with complex changes in the morphology of several brain structures. Some – but not all – of these structures are maths-relevant, and, in the case of tRNS-assisted training, they are distal to the site of the stimulating electrodes. Second, tRNS can improve performance in arithmetic (Studies 2 and 3), although the mechanisms by which this occurs are not yet fully understood, neither neurally nor behaviourally. Third, I found (Study 4) that brain lesions leading to impairment in the number domain do not necessarily affect processing in other domains – such as pitch – that are otherwise linked to number via a putative common code in the parietal lobes.

Long abstract

The brain is in a perpetual state of change, not only during development but also as a result of each and every one of our actions, perceptions, emotions and cognitions. As we learn and acquire expertise, this experience-dependent plasticity interacts with (genetically-predetermined) congenital predispositions in complex ways. This means that drawing causal links between specific areas of learning and certain changes in the structure of the brain is not a trivial task. Nevertheless, the field of cognitive neuroscience has been making progress in establishing likely correspondences between cognitive functions and structural parameters of the brain relating to both the cortex and the connectivity between cortical regions. This progress has been possible due to the recent development of complementary techniques, for instance, in terms of the resolution provided (e.g. the spatial resolution of magnetic resonance imaging, MRI, compared to the temporal resolution of electroencephalography, EEG), and the type of inference allowed (the correlational approach of MRI compared with the causal approach of transcranial magnetic stimulation, TMS). In particular, developments in the acquisition sequences and analysis pipelines of MRI have allowed researchers to find evidence of plasticity not just across years of experience (Aydin et al., 2007) but also across weeks (Scholz, Klein, Behrens, & Johansen-Berg, 2009), days (Takeuchi et al., 2011) and even hours (Sagi et al., 2012).

In the field of numerical cognition, the function of the number-processing and calculating brain has been well described for more than a decade (e.g., R. Cohen Kadosh et al., 2005; L. Cohen, Dehaene, Chochon, Lehéricy, & Naccache, 2000; Delazer et al., 2003, 2005; Göbel, Johansen-Berg, Behrens, & Rushworth, 2004). Studies suggest that the process of numerical learning in adults involves similar neural changes to those observed in development, only at a different time scale. Most notably, these changes consist of a spatial frontal-to-parietal shift in recruited brain areas (for a review, see Ansari, 2008), accompanied, behaviourally, by increasingly efficient arithmetic strategies (Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Jordan, Hanich, & Kaplan, 2003). Throughout and after development, proficiency in arithmetic and other high-level numerical

processing is believed to build on a much lower-level faculty, referred to as the "number sense" (Cantlon, Platt, & Brannon, 2009). At the core of the number sense lies the so-called approximate number system (ANS), the ability to encode and process discrete or analogue quantity without the need for culturally-defined symbols such as digits. The ANS is considered by some to be a foundation for the human invention of mathematics and thus a domain-specific predictor for proficiency in it (Dehaene, Piazza, Pinel, & Cohen, 2003; Piazza, Pinel, Le Bihan, & Dehaene, 2007), whereas others suggest that individual differences in maths are instead best explained by domain-general factors (Geary, 1993; Göbel & Snowling, 2009). The improvement of numerical cognition function using non-invasive transcranial electrical stimulation of the brain (tES), in particular transcranial direct current stimulation (tDCS) and transcranial random noise stimulation (tRNS), has recently been well described (e.g., Cappelletti et al., 2013; R. Cohen Kadosh, Soskic, Iuculano, Kanai, & Walsh, 2010; Snowball et al., 2013). Less well understood is how exactly the structure of the brain makes some people better with numbers than others, and how experience with numbers affects everyone's brain. The few studies that have begun to address these questions have focused on neural correlates of existing child (Tsang, Dougherty, Deutsch, Wandell, & Ben-Shachar, 2009) and adult expertise (Aydin et al., 2007) and mental arithmetic training (Takeuchi et al., 2011). These results are, however, divergent as to the direction of change in measures characterising the brain's grey and white matter that is associated with better maths ability. Some of the outstanding questions relate to the extent to which behavioural scores (in maths and in general) predict structural differences presumably caused by expertise; and to the structural changes that accompany the (by-now) reliable behavioural facilitation effects of tES-assisted training.

Number is not processed in the brain in isolation, and there is no single "number region" (Arsalidou & Taylor, 2011). Quite the contrary, number seems to share many neural resources with other dimensions, and there is converging evidence that the parietal lobes encode relative magnitude not just for number (e.g. Dehaene et al., 2003) but also for time (e.g. Walsh, 2003), space (e.g. Knops, Thirion, Hubbard, Michel, & Dehaene, 2009), luminance (R. Cohen Kadosh et al., 2005) and musical pitch (e.g. Foster & Zatorre, 2010a) among others. It is therefore difficult to single out a particular parietal (or other) structure that supports judgements along a particular dimension. On the other hand, behavioural evidence for this entanglement is provided by reliable pair-wise interactions that exist between these dimensions, e.g. between number and space (Spatial-Numerical Associations of Response Codes, SNARC; Dehaene, Bossini, & Giraux, 1993), pitch and space (Spatial-Musical Associations of Response Codes, SMARC; Rusconi, Kwan, Giordano, Umiltà, &

Butterworth, 2006), and number and pitch (Seeing Number As Pitch, SNAP; Marghetis, Walker, Bergen, & Núñez, 2011).

This thesis aims to address a few of the outstanding questions revolving around the concept of number in the brain. In Study 1, I looked at the specific ways in which years of maths experience shape the brain; I also sought to find which structural parameters of which regions correlate with numerical task performance of the mathematician and of the non-mathematician. In Study 2, I built on an existing arithmetic learning paradigm to assess how enhancement of numerical processing can be achieved by electrically stimulating different parts of the brain in succession within the same training period. In Study 3, I returned my focus to brain structure, this time by measuring how different stimulation conditions change structure when accompanied by cognitive training in arithmetic. Finally, in Study 4, I sought to unravel the mechanisms that provide a common metric encoding number and other relative judgements, specifically pitch, by examining their interaction in the healthy brain compared with the brain lesioned in critical number areas. I expand on the premises, methods and results of each of these studies below.

Study 1 (Chapter 2) aimed to fill in an existing gap in the neuroscience of mathematical cognition. We have increasing knowledge about how learning a given cognitive or motor skill (becoming an expert in it) changes the structural parameters of certain parts of the brain, as well as an increasing understanding of how improvement in maths changes the focus of functional activation between regions of the brain. However, not much is known about how becoming good at maths changes the structure of the brain, nor about which brain regions in particular are structurally predictive of behavioural performance in maths. An important precursor to the current work comes from Aydin et al. (2007), in which voxel-based morphometry (VBM) was used to find group differences in grey matter density (GMD) across the entire brain between mathematicians and non-mathematicians. While several regions were identified in that study, a major drawback was that no behavioural data was collected, and thus there was no information about which (if any) of the observed GMD differences correlated with behavioural performance.

In Study 1, I employed a wide range of MRI methodology, in order to characterise the brain in terms of as many structural parameters as possible. These included two approaches that are currently used in neuroimaging to assess grey matter parameters: voxel-based and surface-based morphometry (VBM and SBM). Across the different methods, I assessed grey matter density and thickness, gyrification of the cortical surface, white matter integrity expressed as fractional anisotropy (FA) and mean diffusivity (MD), and

direction and strength of fibre tracts recreated using probabilistic tractography. In Study 1, mathematicians and non-mathematician (humanities) academics, matched for age, gender and number of years of education, underwent structural and diffusion MRI scans and were tested on a broad range of cognitive dimensions, including working memory, attention, IQ, numerical and social skills.

At the behavioural level, mathematical expertise was associated with better performance in a diverse range of behavioural tasks, even for dimensions that are not purely mathematical, such as fluid intelligence, logic, and verbal reasoning. At the anatomical level, for mathematicians relative to controls I found increased GMD in the right superior parietal lobule (SPL) but decreased GMD in the right intraparietal sulcus (IPS) and left inferior frontal gyrus (IFG), and decreased surface area in the left fusiform gyrus (FG). The decreased GMD in right IPS is a surprising result, as this has previously been identified in individuals with dyscalculia (Cappelletti & Price, 2014; Molko et al., 2003; Rotzer et al., 2008; Rykhlevskaia, Uddin, Kondos, & Menon, 2009). For controls, behavioural performance correlated negatively with GMD and local gyrification index in the left superior frontal gyrus (SFG). There was a double dissociation with the mathematicians group, whose behavioural scores correlated positively with GMD in the left SFG. In terms of white matter, there were no significant group differences in fractional anisotropy, however tractography revealed a stronger right SFG to left SFG tract in mathematicians than in controls. Such homotopic tracts between homologous regions have previously been shown to display a high degree of variability between regions and between individuals, and so the left-to-right SFG tract, given the behavioural correlation exhibited by this cortex, might be particularly suited to experience-dependence plasticity.

The results of Study 1 suggest a complex relationship between mathematical expertise and cortical metrics in parietal and prefrontal regions previously implicated in numerical cognition. The divergent results across the identified changes (increases in some regions, decreases in others) comes as no surprise given similar divergences between different structural studies in the same area of expertise, e.g. greater cortical thickness for musicians (Bermudez, Lerch, Evans, & Zatorre, 2009) but also for tone-deaf people with amusia (Hyde et al., 2007); or even within a single study, e.g. both increased and decreased GMD found in musicians vs non-musicians (James et al., 2013). The cross-sectional nature of this study prevents a causal inference, i.e. a claim that the observed changes are a result of experience-dependent plasticity. Nevertheless, it is plausible that a bidirectional feedback loop exists between experience and structural brain changes, and that the nature of this loop changes across the lifespan.

Study 2 (Chapter 3) was the first of two studies that used tRNS (transcranial random noise stimulation) as a non-invasive brain stimulation method to enhance arithmetic learning over a 5-day training period. A previous tRNS study, with which the author of this thesis was involved, acted as the model for the training paradigm (Snowball et al., 2013). As in that previous study, the behavioural paradigm was based on a distinction between rote learning (fact retrieval) and procedural learning. The training involved learning to solve arithmetic problems by applying an algorithm (Calculation) or by memory retrieval (Drill). tRNS was applied to the bilateral prefrontal cortex for the first 3 days of training and to the posterior parietal cortex for the last 2 days; these regions were previously shown to be implicated in learning in this order. Both trained and untrained problems were used in a post-training testing session. Task difficulty was manipulated between groups by having subjects train on either a small or a large set of problems.

tRNS decreased overall reaction times (RTs) and increased the rate of learning for Calculation problems during the training phase; it also decreased Calculation RTs for both trained and untrained (new) problems during the testing phase. These behavioural enhancements reached significance only for the more difficult condition. There were no effects of tRNS on control tasks of attention and working memory. These results confirmed the capacity for transcranial electrical stimulation (and tRNS in particular) to produce specific facilitative effects on numerical cognition. They also suggested that task difficulty is an important factor that modulates the behavioural enhancement effect, as shown in one previous study that found the effectiveness of tES increased with task difficulty (Pope & Miall, 2012). Study 2 suggested that facilitation might only occur if task difficulty reaches a certain level, and it is possible that the size of the facilitation might subsequently depend linearly on this difficulty. This conclusion was used when adapting this same behavioural paradigm to the following experiment, Study 3.

In Study 3 (Chapter 4), the most extensive of the four empirical studies performed for this thesis, I merged the two overarching questions that I set out to address: namely, how does tES modulate mathematical expertise, and how does improved expertise/skill acquisition affect brain structure? I used the training paradigm developed for Study 2 but, based on that study's results, no longer manipulated task difficulty; instead, I chose the larger problem set for Calculation and the smaller one for Drill, as those conditions were associated with significant effects and with good retention during the testing phase, respectively. During training, tRNS was applied to one group as prefrontal stimulation for all 5 days, to another group as posterior parietal stimulation, and to a third group as sham. Therefore, in a first instance,

this study should potentially replicate the results of Snowball et al. (2013) and those of Study 2. In Study 3, all subjects were MRI-scanned pre- and post-training.

This longitudinal design (3 groups, 2 time points) allowed me to compare each of the real stimulation groups (frontal, parietal) to the sham control and thus infer how each of these affects the brain when used as tRNS-assisted training. As opposed to the "chicken and the egg" inferential paradox of the cross-sectionally designed Study 1, here I was able to unambiguously ascribe any observed group changes from pre- to post-training to the tRNS intervention. The same MRI acquisition and analyses were performed as in Study 1.

Behaviourally, tRNS decreased RTs on the Calculation task for both frontal and parietal stimulation as compared to sham. Exploring this effect with further analyses did not reveal an accompanying change in learning rate, as in Study 2, but did reveal group differences in the average within-session (online) and between-session (offline) learning. These effects did not, however, refer to the Calculation RTs for which a decrease was found, suggesting that in this case, the online/offline distinction was not helpful and that some other phenomena might underlie the observed decrease in RTs. Interestingly though, online effects were negatively correlated with offline effects in two instances (Calculation accuracy and Drill RT), both within and across stimulation groups. This observation might be helpful outside of the tES-assisted training realm, in generally classifying training subjects as "offline" and "online" learners.

Anatomically, at a whole-brain level there were no group differences in the pre-to-post-training change in GMD and FA, and no correlations between those measures and baseline and learning-related behavioural scores. Null results were also obtained for these measures when regions of interest (ROIs) were defined based on the stimulating electrodes and on significant clusters obtained in Study 1. Pre-to-post change in FA along homotopic tracts based on these ROIs was also not different between groups. Several group effects were, however, found in terms of the pre-to-post change in surface-based measures (volume and thickness). These effects went in both directions (i.e. both increases and decreases in the frontal and parietal groups as compared to sham) and their clusters were not proximal to the stimulation site. This suggests that it is not straightforward to predict either the regions of the brain receiving the maximal tRNS effect, or the direction of change in structural parameters of these regions as a result of tRNS.

Study 4 (Chapter 5) was a smaller-scale behavioural investigation into the lower-level representation of number as a dimension that is believed to share a common – parietally-supported – metric with several other

dimensions. One dominant theory in the field (ATOM) has long posited that time and space are similarly cognitively represented (Walsh, 2003), but newer results also suggest that relative judgements of musical pitch recruit similar parietal areas as do relative judgements of numerical magnitude (Foster & Zatorre, 2010a). At the level of behavioural observation, the well-known SNARC effect (Spatial-Numerical Association of Response Codes), in which small numbers are responded to faster with the left hand and larger numbers with the right hand (Dehaene et al., 1993), later found its pitch equivalent in the SMARC effect (Spatial–Musical Association of Response Codes; Rusconi et al., 2006). Both the SNARC and the SMARC are implicitly assumed to rely on an intact representation of the relevant dimension in the brain, and so in this study I tested whether, on one hand, the number-space interaction (SNARC) also exists in lesion patients having numerical impairments; and on the other, whether these impairments generalise to the presumably related dimension of pitch, whose own interaction with space was probed with the SMARC. At the group level, a double dissociation was found in that controls did not show SMARC but showed SNARC, whereas the opposite was true for the patient group. Furthermore, the single cases whose SNARC differed qualitatively from the normative value of the controls group also differed in terms of accuracy of number-to-space mapping (number line), suggesting a more general numerical impairment. While the absence of a SMARC in controls can be explained by the difficulty in replicating this effect (in comparison to the SNARC), the dissociation in patients suggests that, at least in the lesioned brain, number and pitch dissociate sufficiently for impairment in the former not to necessarily imply impairment in the latter, at least in terms of the ability for each to interact with space. Further studies should elucidate whether this holds true in the other direction (patients with pitch impairments). This study addressed the specific question of pairwise interactions between dimensions believed to have a common metric in the parietal lobe, and found a double dissociation between the two. These results might inspire further research that will ascertain whether so-called "specific" impairments such as dyscalculia (number) and amusia (pitch) arise from the dysfunction of shared or domain-specific processes. Such knowledge may eventually inform the selection of rehabilitation strategies.

Acknowledgements

Looking back at what have been my "Oxford years" is an emotional, reflective and humbling mental exercise. There are, of course, many people without whom the challenging endeavour of leaving Oxford with a doctorate would have been even more challenging, or simply less enjoyable. I dedicate this slightly less technical part of my thesis to them.

"There for me", not just as support during hard times but also constituting the most important part of what made Oxford so pleasant for me, were friends I made here and who will undoubtedly stay with me for life. I have in mind my dearest friends Ciprian Ploştinar, Dragoş Cîrstea, Eliza Gheorghe, Julian Nowag, Maria Ioannou, Carlos Velasco, Tewfiq Al-Sharaiyra and Devin Terhune. You have made my time in Oxford truly memorable, and I am very fortunate to have you as my best friends.

I have learnt a lot from my first supervisor, Dr Roi Cohen Kadosh, not only about science per se (and how to do it creatively and efficiently) but also about independent work and about the sacrifices that are sometimes necessary in the academic career. I thank Roi – and the Wellcome Trust – for entrusting me with the Research Assistant position that has funded most of my DPhil. I also thank the Ratiu Family Foundation for entrusting me with their scholarships, which were part of what made my DPhil a feasible financial enterprise.

I hold tremendous respect and gratitude for my second supervisor, Professor Glyn Humphreys, whose helping and inspiring influence have been outstanding. I cannot imagine how I would have succeeded to the end of this long and winding DPhil journey had it not been for his presence. I have been lucky enough to also receive valuable personal and scientific guidance – indeed, mentorship – from several other people in the Department of Experimental Psychology, to whom I am extremely grateful. Of these people, my warmest gratitude no doubt goes to my dear friends Devin Terhune and Mihaela Duţã, who have helped and supported me countless many times. I very much thank Dr Rogier Mars for continued help with my increasingly complicated diffusion MRI analyses. I am thankful to Dr Ann Dowker, who has, throughout all stages of my DPhil, often imparted valuable insight and critical feedback on my work and ideas, with regards to various aspects of mathematical cognition. My MRes supervisor at Durham, now also at Oxford, Professor Hannah Smithson, was my first – and, in many ways, still my most influential – academic mentor, to whom I owe an important part of my scientific formation. Part of the reason for this was that Hannah was instrumental in my decision to switch to – and stay in – the field of Experimental Psychology, a decision I never looked back on.

Several others colleagues from EP have also been very helpful, supportive and/or simply (but importantly) fun presences during my time here. Some of these colleagues are my good friends Carlos Velasco, Maria Witek, Urs Schüffelgen, Jacques Launay, Beatrix Krause, Dante Mantini and Nele Demeyere. Also up the hill at FMRIB, the astonishing density of IQ points per voxel has been very beneficial to my growth as a cognitive neuroscientist and to the quality and understanding of my MRI methodology. Of particular note is the help of Dr Adam Thomas, whose patience with my ubiquitous lists of never-ending questions cannot possibly go by unremarked. From the same building, but mainly providing excellent FSL support for me and others electronically, are members of FMRIB's Analysis Group, in particular Mark Jenkinson, Anderson Winkler, Stephen Smith and Gwenaelle Douaud. I thank Marie Schaer of the University of Geneva for invaluable help and guidance with the complex and often confusing world that is FreeSurfer. Also invaluable there was the help – through the FreeSurfer mailing list – of that package's authors, in particular Louis Vinke, Douglas Greve, Bruce Fischl and Martin Reuter.

In parallel with my DPhil, my growing interest in music psychology has benefited tremendously from the enlightening figure of Professor Eric Clarke of the Music Faculty. I owe Professor Clarke a great deal of gratitude for his advice, support and friendship, and also for his valuable feedback and encouragement during the time when my number-pitch study was taking shape. Still in the realm of music, I thank Reverend Gareth Hughes, Chaplain of Hertford College, for helping me pursue my desire to teach myself the piano by granting me access to the Blüthner Grand. The hour of music-making at that piano, in the magical surrounding that is the Hertford Chapel, is what ended many of my long days of work, and will be among my fondest memories from Oxford.

I thank the City and the University of Oxford for offering me a unique environment for work and life. The abundance, within this environment, of intellectual stimulation of the highest calibre, unimaginable self-development opportunities, and a rich cultural and musical life, have been for me the most fertile medium in which I could have wished to complete the final part of my formal education. This medium has, no doubt, become my second home – a home I will always be thinking back fondly to. As intimidating as it is to think of oneself as becoming part of "Oxford's history", I feel that, with this thesis, this is the chance that I was given, for which I am infinitely grateful.

I have saved the best for last. The single brightest beacon of inspiration, love, encouragement and unreserved support has been and continues to be my beloved mother, Aniela Popescu, to whom my gratitude cannot possibly be expressed in words, and to whom this thesis is dedicated. And at the Florentine end of our long-standing relationship, I thank my beloved girlfriend, Mădălina Moraru, for her unwavering love and support. Soon her own thesis will also be submitted, and so our separate PhD missions in two of Europe's most beautiful – but so far apart – cities will have come to an end.

Table of contents

ABSTRACT II

LONG ABSTRACT IV

ACKNOWLEDGEMENTS XII

TABLE OF CONTENTS XIV

1 GENERAL INTRODUCTION 1

1.1. Nature and nurture: learning and plasticity in numerical cognition 1
1.2. Learning & development 3
1.3. Behavioural and brain correlates of maths ability 5
1.3.1 The cognitive foundations of maths achievement 5
1.3.2 Neuroanatomical markers of maths achievement 8
1.4. Number, magnitude and maths: The scope and methods of numerical cognition 9
1.5. Methodology 11
1.5.1 Using brain stimulation to boost cognition 11
1.5.2 Using structural MRI to quantify (differences in) cognition 16
1.6. This thesis 23

2 STUDY 1: A CROSS-SECTIONAL SNAPSHOT INTO THE STRUCTURE OF THE MATHEMATICAL BRAIN 24

2.1. Abstract 24
2.2. Introduction 25
2.2.1 The neuroanatomy of mathematical expertise 25
2.2.2 Supporting functions for mathematics 26
2.2.3 The current study 28
2.3. Methods 28
2.3.1 Subjects 28
2.3.2 Behavioural tasks 28
2.3.3 MRI acquisition 31
2.3.4 Analyses 31
2.4. Results 36
2.4.1 Behavioural results 36
2.4.2 Brain results 40
2.5. Discussion 48
2.5.1 Behavioural results 48
2.5.2 Brain results 50
2.5.3 Limitations & conclusions 55
2.6. Acknowledgements 56

3 STUDY 2: ENHANCING ARITHMETIC LEARNING USING TRNS 58

3.1. Abstract 58
3.2. Introduction 59
3.2.1 The neuroanatomy of mental arithmetic 59
3.2.2 Brain stimulation and behavioural performance 60
3.3. Methods 62
3.3.1 Subjects 62
3.3.2 Training phase 62
3.3.3 Testing phase 64
3.3.4 Control tasks 65
3.3.5 Brain stimulation 66
3.3.6 Analyses 67
3.4. Results 68
3.4.1 Training phase 68
3.4.2 Testing phase 75
3.4.3 Control tasks 77
3.4.4 Stimulation perception 78

3.5. Discussion 78
3.5.1 Training phase 78
3.5.2 Testing phase 80
3.5.3 tRNS selectivity 81
3.5.4 Conclusion 82
4 Study 3: The structural correlates of tRNS-assisted arithmetic learning 84
4.1. Abstract 84
4.2. Introduction 85
4.2.1 Strategies of mental arithmetic in the brain 85
4.2.2 Evidence of plasticity from longitudinal correlates of training 87
4.2.3 Using brain stimulation to affect maths cognition via consolidation (offline) effects 90
4.2.4 The current study 91
4.3. Methods 92
4.3.1 Subjects 92
4.3.2 MRI acquisition 93
4.3.3 Behavioural tasks 93
4.3.4 Brain stimulation 96
4.3.5 Behavioural analyses 97
4.3.6 MRI analyses 98
4.4. Results 104
4.4.1 Behavioural results 104
4.4.2 Brain results 114
4.5. Discussion 118
4.5.1 Behavioural results 118
4.5.2 Brain results 122
4.5.3 Limitations 125
4.5.4 Conclusions 128
4.6. Acknowledgements 128
5 Study 4: Number-pitch-space interactions in the lesioned brain 130
5.1. Abstract 130
5.2. Introduction 131
5.2.1 Maths and music 131
5.2.2 Space as a medium for interaction 133
5.2.3 The present study 141
5.3. Methods 142
5.3.1 Subjects 142
5.3.2 Behavioural tasks 144
5.3.3 SNARC/SMARC analyses 148
5.4. Results 150
5.5. Discussion 153
5.5.1 Limitations & conclusions 156
5.6. Acknowledgements 156
6 General Discussion 158
6.1. Summary of findings 159
6.2. Overall conclusions 162
6.2.1 Maths performance and brain structure 162
6.2.2 tES to enhance maths performance 162
6.2.3 Number-pitch-space interactions 164
6.3. Limitations 164
6.4. Future directions 165
7 References 168
8 Appendices 188
8.1. Study 4: Patient information (BCoS charts, lesion imaging) 188
8.1.1 P01 189
8.1.2 P02 190
8.1.3 P03 191
8.1.4 P04 192
8.1.5 P07 193
8.2. Study 4: Individual patient plots 194
8.2.1 SMARC (dAcc) 195
8.2.2 SMARC (dRT) 196
8.2.3 SNARC (dAcc) 197
8.2.4 SNARC (dRT) 198
9 List of abbreviations 199

1 General Introduction

1.1. Nature and nurture: learning and plasticity in numerical cognition

It is, by now, well established that not only nature (genetic predisposition) but also nurture (experience) have important contributions in shaping and moulding the brain (see e.g. review by Jolles & Crone, 2012). Given the limitations of examining microscopical brain structure in humans, evidence for experience-dependent neuroplasticity has to mainly come from investigations of macroscopic changes in the brain. The available methodology for this includes magnetic resonance imaging (MRI), positron emission tomography (PET), electro- (EEG) or magneto-encephalography (MEG) and near-infrared spectroscopy (NIRS).

The case for claiming that any observed brain changes have come about via plasticity-like mechanisms is, however, not straightforward to make, and is dependent upon an appropriately-controlled experimental design. The simplest approach from a practical point of view, but also the weakest, is to compare brain structure between single scans acquired for experts and for controls (e.g., Hänggi, Koeneke, Bezzola, & Jäncke, 2010; Hu et al., 2011); ideally, the groups would be matched for all measures except for the presence of the expertise in question. Any effect of a group difference will, however, be limited by the doubt that one has to cast over the cause that has led to the observed effect: this cause might well be the expertise factor – which, as mentioned above, is ideally the most salient aspect that differentiates the two groups – but might plausibly also be unrelated to experience and instead be only or predominantly due to the nature factor (e.g., genetic predisposition). This effectively reduces the inference process to a "chicken or the egg"-type speculation as to whether initial propensities in brains structure led to expertise, or, vice-versa, whether acquired expertise has modified brain structure. While still not unequivocally disentangling cause from effect, additional evidence implicating a given brain region in expertise of a given type can be brought if a significant

correlation is observed between a certain measure of structural difference and some measure of expertise. The latter can be either time-based, e.g. number of years spent in a given profession (Aydin et al., 2007; Maguire et al., 2000) or average number of hours of weekly practice (Bengtsson et al., 2005), or performance-based, e.g. learning rate during training (Sampaio-Baptista et al., 2014). *Study 1 of this thesis (Chapter 2) falls in the first category of studies (comparison between matched groups only, without time- or performance-based measures of expertise), due to practical limitations existing at the time, and unfortunately does not bring brain-expertise correlations due to the limited range of ages present in the expertise group.*

Another approach that might suggest the occurrence of neuroplasticity is administering training to a group of subjects and comparing brain structure post- versus pre-training (Bueti, Lasaponara, Cercignani, & Macaluso, 2012). The weakness of such a design, however, is that, while it is plausible any observed changes are causally related to the training received, they could also simply be due to any other environmental or individual phenomena assumed to also have occurred between the two scanning time points, such as for instance (assuming a training of sufficient length) maturation. The only way to eliminate this doubt as well is to use a design that is controlled in terms of both group (treatment vs. no-treatment, and potentially also treatment that is not related to the behaviour of interest) and time (pre- and post-training scans). Such studies may (Sagi et al., 2012; Sampaio-Baptista et al., 2014; Takeuchi et al., 2011) or may not (Draganski et al., 2004; Scholz et al., 2009) also bring the additional proof of a brain-expertise correlation, as explained above. *Study 3 (Chapter 4) fits in this category, as the different brain stimulation groups were scanned before and after training, and their structural changes were correlated with a measure of their performance.* Table 1 compares all design parameters discussed above, across a selection of well-cited studies that report MRI-derived structural changes as a result of training or existing expertise (whether or not maths-relevant).

Table 1: Comparison of studies observing brain structural changes as a result of in-study training or existing expertise.

Study	Domain of expertise/training	Control group	Multiple scans	Correlation with time- or performance-based measure of expertise
(Aydin et al., 2007)	mathematics education	✓	✗	✓ (time)
(Bengtsson et al., 2005)	piano playing	✓	✗	✓ (time)
(Maguire et al., 2000)	taxi driving	✓	✗	✓ (time)
(Hu et al., 2011)	abacus training	✓	✗	✗
(Hänggi et al., 2010)	ballet dancing	✓	✗	✗
(Bueti et al., 2012)	duration estimation	✗	✓	✓ (performance)
(Sampaio-Baptista et al., 2014)	juggling	✓	✓	✓ (performance)
(Draganski et al., 2004)	juggling	✓	✓	✗
(Scholz et al., 2009)	juggling	✓	✓	✓ (performance)
(Sagi et al., 2012)	spatial learning	✓	✓	✓ (performance)
(Takeuchi et al., 2011)	mental arithmetic	✓	✓	✓ (performance)

The structural changes observed in the studies summarised in Table 1 have arisen as a result of training or expertise in a wide range of domains, and across a time course ranging from hours (Sagi et al., 2012), days (Takeuchi et al., 2011), and weeks (Scholz et al., 2009) to years (Hu et al., 2011) and even decades (Aydin et al., 2007; Bengtsson et al., 2005). Other studies have, instead (or additionally), documented functional (Bueti et al., 2012; Olesen, Westerberg, & Klingberg, 2004; Van Opstal, Fias, Peigneux, & Verguts, 2009) or connectivity (Schulz et al., 2013) changes in the brain arising presumably as a result of training or expertise.

1.2. Learning & development

The developing brain differs from the adult one not just in terms of structural architecture but also in terms of functional organisation (for a review see e.g. B. J. Casey, Tottenham, Liston, & Durston, 2005). This implies, for instance, and as the studies mentioned below suggest, placing different weights upon the recruitment of different regions within a network, or of different networks altogether. With development, there is decreasing reliance on (mostly frontal) brain regions that subserve general cognitive functions. There is, instead, increasing reliance on more domain-specific regions, that become increasingly fine-tuned. This

general trend emerges from studies across several domains of cognition involving tasks such as suppression of overlearned stimulus-response mappings in favour of new ones (B. J. Casey, Thomas, Davidson, Kunz, & Franzen, 2002), executive functions associated with response inhibition (Tamm, Menon, & Reiss, 2002), word generation (Brown et al., 2005) and implicit sequence learning (Thomas et al., 2004). In the case of numerical cognition, typical development seems to involve a shift in brain areas recruited for number processing, from prefrontal (Kaufmann et al., 2006; Kucian, von Aster, Loenneker, Dietrich, & Martin, 2008) areas to parietal areas in older children, such as the bilateral IPS (for a review, see Ansari, 2008). This shift in recruitment of brain areas is accompanied at the behaviour level by a shift in strategies, from inefficient procedures such as counting to faster memory retrieval (Geary et al., 2004; Jordan et al., 2003). There are, however, substantial individual differences in the degree to which children apply this change of strategy over the course of this critical period (Geary et al., 2004), while children with dyscalculia might never make the transition completely (S. Cho, Ryali, Geary, & Menon, 2011).

A similar shift might accompany adult arithmetic learning, on a much smaller time scale than that seen in development. To study the time course of functional recruitment in this situation, Ischebeck et al. (2007) asked subjects to perform in the scanner multiplication problems that were presented either frequently ("repeated") or infrequently ("novel"). As subjects gained more relative experience for repeated versus novel problems, the pattern of brain activity shifted from frontal to parietal areas and, within the parietal lobes, from the bilateral IPS to the left angular gyrus. This trend became significant after a small (<10) number of repetitions, and it resembled similar trends observed in similar studies that used more intensive training (Delazer et al., 2003, 2005; Grabner, Ischebeck, et al., 2009; Ischebeck et al., 2007; Ischebeck, Zamarian, Schocke, & Delazer, 2009; Pauli et al., 1994). It is interesting to note that it is not just increasing proficiency in arithmetic alone that implies the involvement of the angular gyrus. For instance, Grabner et al. (2009) trained subjects with both an arithmetic and a visuo-spatial paradigm, and bilateral AG activation was subsequently observed for the trained versus untrained contrast in both domains.

It seems, thus, that training studies with adults show a trajectory of functional recruitment – when trained material is contrasted with untrained material – similar to that described above in the case of increasing development. In developmental studies, the effect of the learning process is of course confounded with that of actual physiological maturation, thus making a direct comparison with adult learning studies

difficult. Nevertheless, it could be said that the effects of development on the brain are, in some sense, similar to those seen with increasing expertise in adulthood via learning.

1.3. Behavioural and brain correlates of maths ability

1.3.1 The cognitive foundations of maths achievement

Rather than being considered one of the unique features of human cognition (such as language or music), there is strong evidence (reviewed in R. Cohen Kadosh, Lammertyn, & Izard, 2008) from developmental, comparative, cross-cultural and neurophysiological studies suggesting that numbers are represented as mental magnitudes by a neural circuitry basic enough to also be identified in other species, as well as across human development. The important observation that a basic "number sense"[1], not relying on any symbols denoting magnitudes, exists even in human infants and non-human species (Agrillo, Piffer, Bisazza, & Butterworth, 2012; Cantlon et al., 2009) has made some researchers speculate about this approximate number sense/system (referred to as the ANS) being a foundation upon which human mathematical understanding can be built (e.g. Piazza et al., 2007). Knowing what predicts proficiency in symbolic number processing and maths in general, is useful both practically, e.g. from an educational or interventional perspective, and theoretically (e.g., cognitive and developmental models of numerical cognition). Developmental psychologists have appreciated that zooming in on the limited time window just before numbers are learnt would allow them to better understand how these two different modes of numerical understanding, i.e. symbolic and non-symbolic, interact with and complement each other. Some studies that have done this have proposed that children form an understanding of symbolic number by making a direct mapping between the symbols (Arabic digits) and the non-symbolic analogue representation conferred to them by the ANS (Mundy & Gilmore, 2009; Noël & Rousselle, 2011). The advent of this exact number system (ENS) would then, in turn, lead to an increase in precision (i.e., more fine-tuned representations) in the ANS. The two systems (ENS, ANS) are subsequently presumed to remain distinct throughout adulthood (Lyons, Ansari, & Beilock, 2012). The fact that magnitude comparison tasks show similar behavioural effects for both symbolic and non-symbolic magnitude (Fias, Lammertyn, Reynvoet,

[1] Note that the term "number sense" has a slightly different meaning in cognition than it does, for instance, in education. For a detailed theoretical discussion about this concept, see (Greeno, 1991). Throughout this thesis, I use it mainly as a short-hand for the concept of Approximate Number Sense (ANS).

Dupont, & Orban, 2003) confirms this conjecture of a link and "dialogue" between the two modes of thinking about number (but see R. Cohen Kadosh & Walsh, 2009). However, while magnitude comparison tasks using symbolic number prove to be good predictors of children's maths performance (De Smedt, Verschaffel, & Ghesquière, 2009; Holloway & Ansari, 2009b; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013), data on similar tasks using non-symbolic magnitudes have been comparatively inconclusive so far (De Smedt, Noël, Gilmore, & Ansari, 2013).

Domain-*specific* predictors, such as individual differences in the acuity of the ANS, form one school of thought as to what explains variance in maths achievement (Dehaene et al., 2003). This acuity is normally expressed as an index known as the internal Weber fraction, *w* (see Study 1, section 2.2.2, page 26, for more details on this index). However, inconsistent methods used for measuring this index of acuity led to such a link not being well replicated in the literature (Leibovich & Henik, 2013; e.g., Price, Palmer, Battista, & Ansari, 2012; Sasanguie et al., 2013; Szucs, Nobes, Devine, Gabriel, & Gebuis, 2013). This inconsistency has also been speculated to be due to ANS performance being mediated by inhibition: in a dot array paradigm, the resolution of congruency effects (small/large numerosity denoted by a small/large dot areas) relies on interference suppression, a factor not normally controlled for (Szucs et al., 2013). The other school of thought proposes that a better explanation is provided by individual differences in domain-*general* factors instead, such as abilities relating to language that affect mathematical reasoning mediated by verbal codes (Göbel & Snowling, 2009; Pimperton & Nation, 2010); and also to visuo-spatial processing, working memory or long-term memory (Geary, 1993; Raghubar, Barnes, & Hecht, 2010). A lot of the evidence for this line of reasoning comes from the mathematical disability literature, which reports such disabilities being comorbid with, for instance working memory deficits (Passolunghi & Siegel, 2004) or with language-related impairments such as aphasia (Delazer, Girelli, Semenza, & Denes, 1999).

In addition to the ANS, another domain-specific (i.e., number-sense related) factor proposed to possibly predict maths achievement has to do with the distance effect (Moyer & Landauer, 1967), which is the observation that the speed of a number comparison task is proportional to the numerical distance between them. However, as with measures of the ANS, the magnitude of the distance effect proves to also not be a reliable predictor: links are mainly found in young children (Holloway & Ansari, 2009b), while more numerous studies report its absence in older children (e.g. De Smedt & Gilmore, 2011; Sasanguie et al., 2013; Sasanguie, Van den Bussche, & Reynvoet, 2012). Additionally, even this fragile link might be explained not by

the effect's size per se, believed to quantify the precision of number representation on the number line, but simply by the confounding effect of response selection and general reaction time variability (Sasanguie et al., 2013).

A domain-specific predictor of maths achievement that proves more reliable than ANS acuity and size of the distance effect is performance on the mental number line. In this classical number cognition paradigm, subjects are asked to indicate the position of a given numerosity – presented either symbolically (Arabic digit) or non-symbolically (array of dots) – on an unmarked number line, with only the values at the edges of the line displayed; a score is then computed based on the difference between the indicated position (subjective mapping) and the actual position of the number on the line (objective mapping). Several studies have shown that a higher score corresponds to better maths achievements (Booth & Siegler, 2006, 2008; Sasanguie, De Smedt, Defever, & Reynvoet, 2012; Sasanguie, Van den Bussche, et al., 2012; Siegler & Booth, 2004). When plotting subjective mapping against objective mapping, the nature of the best fitting function is also predictive of maths performance. For young children this function is logarithmic, i.e. small numbers are more spaced out and larger numbers are 'compressed' together. As children develop and acquire more expertise with numbers, this best-fitting function becomes linear, indicating that numbers are spaced relatively evenly along the number line (Booth & Siegler, 2006; Sasanguie, De Smedt, et al., 2012). Confirming the relationship between number line accuracy and formal maths education is data collected from remote tribes with no formal maths education. For symbolic numbers the best fitting curve is consistently logarithmic (Dehaene, Izard, Spelke, & Pica, 2008), suggesting that the log-to-linear shift only occurs as a result of training our innate intuitions.

One important concept in the context of fundamental numerical abilities is that of subitising. Subitising refers to a specialised process in visual perception conferring the ability to accurately take in the numerosity of sets of up to 3 or 4 objects at a glance, and without counting (Butterworth, 2010; Mandler & Shebo, 1982). The fact that this basic property of the small numerosity system is impaired in dyscalculic children (Koontz, 1996), and, on the other hand, predicts calculation ability in healthy children (Penner-Wilger et al., 2007), suggests its foundational role in arithmetic. However, since the notion of "arithmetic ability", far from being unitary, actually consists of a large set of independent components and supporting skills (Dowker, 2005), it is plausible that each of these different components has different predictors (Dowker, 2008). Causal models of

how different aspects of cognition might contribute to behaviourally-testable performance in arithmetic are complex (Butterworth, Varma, & Laurillard, 2011).

Another topic of contention in numerical cognition concerns the neural representation (or coding) of number and magnitude. A proposal that has long been widely accepted is the one by Dehaene's group (Dehaene, Dehaene-Lambertz, & Cohen, 1998), according to which this representation is abstract in nature, i.e. it is independent of whether the magnitude is denoted symbolically (with Arabic numerals or number words) or non-symbolically (dot arrays), and in an analogue (continuous) or discrete (step-wise) manner. This proposal has been challenged by Cohen Kadosh and Walsh (2009), who reviewed evidence against an abstract coding of magnitude and for distinct numerical representations, mediated by specific neuronal populations in the parietal lobes. This evidence includes data from paradigms of automatic number processing such as the numerical Stroop task (Henik & Tzelgov, 1982).

In the current thesis, I used several of the measures reviewed above, as well as other paradigms:

- Number acuity task, to compute the Weber fraction (studies 1 and 3);
- Number line task (studies 1, 3 and 4);
- Numerical Stroop (size congruity) task, a reaction time task that assesses the degree of automaticity in symbolic number comparison, using digits of varying physical size (study 1);
- In addition, for studies 2, 3 and 4 I have used the Numerical Operations and Mathematical Reasoning sub-components of the Wechsler Individual Achievement Test (WIAT; Wechsler, 2005), in order to measure subjects' higher-level mathematical performance. The former subcomponent requires subjects to solve written calculation problems and simple equations involving the four basic operations of addition, subtraction, multiplication and division, and also some more advanced ones such as powers and square roots. The latter subcomponent tests subjects orally on problems with whole and decimal numbers, interpreting graphs and identifying mathematical patterns.

1.3.2 Neuroanatomical markers of maths achievement

At the level of the brain, superior performance in arithmetic has been correlated with changes in both grey matter (Takeuchi et al., 2011) and the fractional anisotropy (FA) integrity measure of white matter (Tsang et al., 2009). There is no consensus, however, of the direction that these effects take with increased

proficiency; for instance, the above-mentioned studies reported a *decrease* in grey matter density and an *increase* in FA, respectively. I discuss this in more detail in Study 1's Discussion (section 2.5, page 48).

One of the ways to provide a causal relationship between numerical abilities and brain architecture is by studying subjects with numerical impairment (e.g. L. Cohen et al., 2000; Price, Holloway, Räsänen, Vesterinen, & Ansari, 2007). Although for a long time given much less attention than their linguistic counterparts (e.g. aphasia), numerical difficulties, whether of a congenital (developmental dyscalculia), acquired (acalculia) or social/emotional nature (maths anxiety), are becoming increasingly acknowledged as a problem affecting both the individual and society. This parallels a recognition of the importance of numerical abilities in people's education, careers and health (Beddington et al., 2008). Even in the absence of one of these conditions affecting numerical ability, the range of individual differences in maths performance and, more generally, in numerical abilities, is large enough to motivate psychologists to search for what might be some of the underlying factors determining such variability. While some of them (like maths anxiety) can in turn be attributed to individual and educational factors, the source of the most salient of these disorders, developmental dyscalculia, is yet unknown, although speculations suggest a core deficit in manipulation of quantity (Price et al., 2007). In terms of the neuroanatomical basis of dyscalculia, there are tentative suggestions about altered grey matter density in certain number-relevant regions including the IPS and the prefrontal cortex (Cappelletti & Price, 2014; Molko et al., 2003; Rotzer et al., 2008; Rykhlevskaia et al., 2009).

1.4. Number, magnitude and maths: The scope and methods of numerical cognition

In its attempt to understand how the mind encodes and processes quantity and magnitude, numerical cognition looks at different stages of the cognitive process. Arguably the most essential of these processes relates to the basic understanding of magnitude, either as numerosity (non-symbolic) or as the concept of (symbolic) number. At its most basic level, numerosity is studied neurophysiologically, by probing the response of individual neurons to various quantities (Nieder, 2012; Nieder & Miller, 2004). Although a homologue neural coding scheme for approximate numerosity is also assumed to exist in the human brain (Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004), the direct probing of quantity representation in humans is normally performed with behavioural paradigms such as comparison of symbolic number (Chochon, Cohen, Moortele, & Dehaene, 1999; R. Cohen Kadosh et al., 2005; Fias et al., 2003; Göbel et al., 2004; Lavidor,

Brinksman, & Göbel, 2004; Pinel, Dehaene, Rivière, & LeBihan, 2001), comparison of non-symbolic numerosity such as dot arrays (Cappelletti et al., 2013; Piazza et al., 2004; Szucs et al., 2013) and discrete-analogue response tasks (DART; Castelli, Glaser, & Butterworth, 2006). Some of these paradigms have also been taken to the functional scanner, for instance in the form of the fMRI adaptation paradigm or fMRA (Ansari, Dhital, & Siong, 2006; R. Cohen Kadosh et al., 2011; R. Cohen Kadosh, Cohen Kadosh, Kaas, Henik, & Goebel, 2007; Piazza et al., 2004, 2007).

Moving to the next hierarchical level of number processing, processes underlying mental arithmetic are assessed using tasks involving any combination of the four main operations: addition, subtraction, multiplication and (more rarely) division. A consistent finding of studies using such tasks is that multiplication, which relies heavily on memorised facts, relies on left-hemisphere regions important for language, such as the left IFG and left AG (Grabner, Ansari, et al., 2009; Grabner, Ischebeck, et al., 2009; Menon, Rivera, White, Glover, & Reiss, 2000). On the contrary, addition and subtraction make *some* use of stored facts (particularly small-range addition) but mostly involve actual quantity manipulation, rely instead only on (different) parietal sites such as IPS bilaterally (Arsalidou & Taylor, 2011; Chochon et al., 1999; Grabner, Ansari, et al., 2009; K.-M. Lee, 2000; Stanescu-Cosson et al., 2000). The defining characteristic of multiplication versus addition and subtraction is that the former operation is fast and precise, since it circumvents actual calculation, instead retrieving the result from long-term memory (e.g. 8 × 9 = 72), whereas the latter operations are approximate[2] and thus error-prone (e.g. 23 – 7 = 16). An interesting analogy can be made between this dissociation of the three arithmetic operations mentioned above, and the different stages of arithmetic learning discussed earlier: acquiring competence in arithmetical tasks involving any combination of these operations is typically accompanied by a shift in strategy, from initially slow, effortful and error-prone ones requiring calculation to fast, effortless and exact ones requiring retrieval of linguistically-stored facts. Evidence for this comes from both developmental studies (Barrouillet & Fayol, 1998; Lemaire & Siegler, 1995; Siegler, 1988) and adult training studies (Anderson, Fincham, & Douglass, 1999; Logan & Klapp, 1991; Rickard, 2004). Efficient use of arithmetic involves an appropriate balance of retrieval and calculation strategies (as well as other strategies; see Dowker, 2005) and also involves the ability to transfer trained knowledge to new problems, which is why sometimes the transfers itself is of interest (Delazer et al., 2005).

[2] The term 'approximate' does not imply that the answer itself is approximate - an exact result must typically be provided in any addition problem. Rather, it implies that these operations (addition and subtraction) rely on calculation and on the magnitude system (which is approximate), even though the end result itself is exact.

Another way in which numerical cognition is investigated in the brain is by examining the mutual causality by virtue of which expertise in higher level cognitive faculties involving numerical processes (either existing or acquired through learning) is influenced by (and changes) the function and structure of the brain. Even though – as discussed elsewhere in the thesis – the direction of the causality, i.e. untangling cause from effect, is rarely straightforward, further examination of this complex mutual interaction can still provide meaningful answers. Such examination be done at different levels:

- behaviourally, for instance by examining the nature of the relationship between such higher-level abilities and more basic ones relating to number sense;
- by imaging the brain and observing how its function and structure differs with expertise. To my knowledge, studies of this sort are very few (structural: e.g. Aydin et al., 2007; functional: e.g. Fehr, Weber, Willmes, & Herrmann, 2010), and *Study 1 (Chapter 2) set out to fill in some of that gap*.
- by using non-invasive brain stimulation to interfere with normal brain function and observe how the process of acquiring expertise changes any combination of the measures stated above. *This is what Studies 2 and 3 (Chapter 3 and 4) have set out to do.*

1.5. Methodology

1.5.1 Using brain stimulation to boost cognition

Transcranial Electrical Stimulation (tES) is a brain stimulation technique that is receiving an increasing amount of attention due to its ability to improve both impaired and normal performance in a large spectrum of domains including motor (Nitsche & Paulus, 2000, 2001), working memory (Andrews, 2011; Fregni et al., 2005), language (Holland et al., 2011), perceptual learning (Fertonani, Pirulli, & Miniussi, 2011), synesthetic experience (Terhune, Tai, Cowey, Popescu, & Cohen Kadosh, 2011) and basic numerical abilities (e.g., Cappelletti et al., 2013; R. Cohen Kadosh et al., 2010; Snowball et al., 2013). In comparison with the more widely used transcranial magnetic stimulation (TMS), tES is less expensive, more portable and presents a reduced spectrum of possible side effects (Poreisz, Boros, Antal, & Paulus, 2007; Utz, Dimova, Oppenländer, & Kerkhoff, 2010).

There are several forms of tES, depending on the type of current used. Transcranial direct current stimulation (tDCS) uses constant current in a polarised circuit with a positive (anode) and a negative

(cathode) electrode, and manipulates this polarity to achieve both impairment as well as enhancement by using either cathodal (Antal, Nitsche, & Paulus, 2001; Loui, Hohmann, & Schlaug, 2010; Nitsche et al., 2005) or anodal (Fregni et al., 2005; Nitsche et al., 2005) stimulation, respectively. A form of tES that has been in use more recently, and which I have used in Studies 2 and 3 (Chapters 3 and 4), is transcranial random noise stimulation (tRNS). The current in tRNS is a signal consisting of random current intensity values, with frequencies distributed across a certain range (Terney, Chaieb, Moliadze, Antal, & Paulus, 2008). Figure 1 illustrates the typical frequency distribution of the electrical signal used in tRNS, as well as its time-domain waveform, with normally-distributed current intensity values. Like the effects of the more established method of anodal tDCS, the effects of tRNS applied online with a cognitive or motor task have, in many studies so far, proved to also be facilitatory (Cappelletti et al., 2013; Chaieb, Paulus, & Antal, 2011; Prichard, Weiller, Fritsch, & Reis, 2014; Snowball et al., 2013; Terney et al., 2008). In one study directly comparing the effect size of said facilitation between the two methods applied to the visual cortex during an orientation discrimination task, it was even suggested that tRNS had a stronger effect than anodal tDCS (Fertonani et al., 2011). Conversely and, to my knowledge, so far uniquely, in a study where anodal tDCS and tRNS were applied to the DLPFC during a dot pattern prototype distortion task, the methods again agreed in that they both produced similarly *inhibitory* behavioural effects (Ambrus et al., 2011). An instance of anodal tDCS (but not tRNS) enhancing working memory as a result of DLPFC stimulation has been reported, although the authors refer to several methodological shortcomings to explain the discrepancy (Mulquiney, Hoy, Daskalakis, & Fitzgerald, 2011). In summary, the consensus of previous studies comparing the effects of tRNS with those of anodal tDCS is that the former are qualitatively similar and possibly quantitatively superior, which is why it is behavioural *facilitation* that I aimed to achieve with tRNS in Studies 2 and 3. Both tDCS and tRNS are comparable in terms of their aftereffects and in terms of their effects' dependency on the distance between stimulation electrodes under various montages (Moliadze, Antal, & Paulus, 2010). In comparison with tDCS, tRNS has the advantage of having a higher current intensity threshold for perception (Ambrus, Paulus, & Antal, 2010), which makes it logistically better suited to blind controlled experiments.

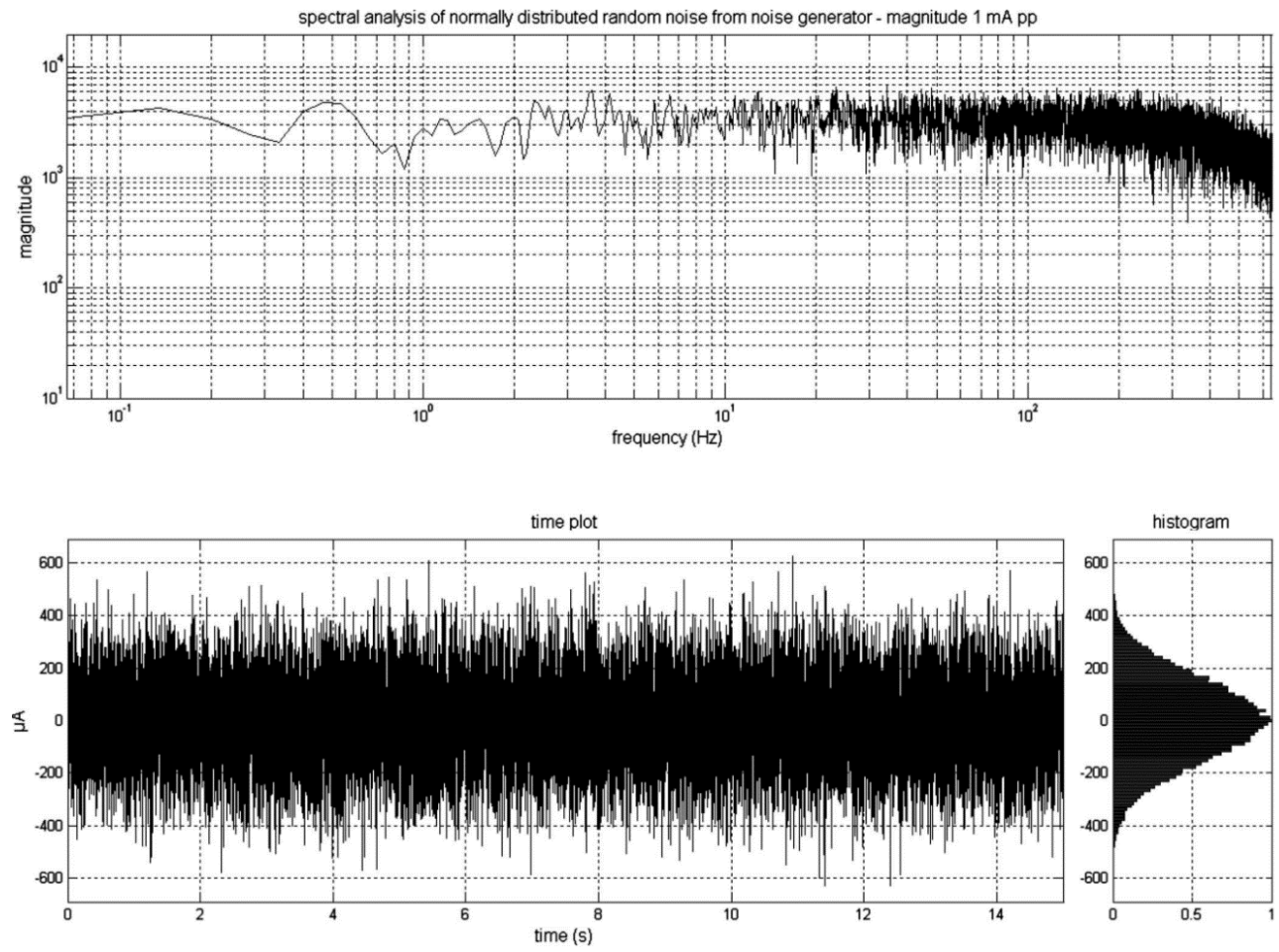


Figure 1: Upper panel: The approximately-uniform frequency distribution of the tRNS signal, resembling that of white noise. Lower panel left: Typical wave form of the tRNS signal, plotted as current intensity (expressed in microamperes, µA) against time (expressed in seconds, s). Lower panel right: Histogram showing the normally-distributed levels of current intensity. Reproduced, with permission, from (Terney et al., 2008).

As illustrated elsewhere in this chapter, stimulating the brain transcranially with electrical currents has the ability to enhance cognition. Previous work in the field has tried to clarify the cellular-level mechanisms by which such behaviourally-observable effects occur. However, these mechanisms are currently only well understood for tDCS, and only partially understood for newer forms of tES such as tRNS. When thought of as an electrical circuit, the cortex is polarised, and it is believed that tDCS acts by either de- or hyper-polarising it – specifically, its resting membrane potential – based on the current polarity – anodal or cathodal – applied (Nitsche et al., 2008). In particular, anodal stimulation lowers the resting potential of neural membranes and depolarises the cortex, bringing it closer to the activation threshold. This increased excitability increases the spontaneous firing rate of cortical neurons and helps facilitate the learning process by inducing the strengthening of synapses and inducing long-term potentiation (LTP) by modifying NMDA-receptor efficacy (Terney et al., 2008). Cathodal stimulation does just the opposite, i.e. it hyper-polarises the membrane, which

inhibits cell firing and decreases the spontaneous firing rate of cortical neurons. tDCS is believed to modulate glutamatergic activity and its behavioural effects are believed to be at least partly due to metaplasticity, i.e. the plasticity of synaptic plasticity (Stagg et al., 2011).

While the facilitatory behavioural effects of tRNS appear so far quite reliable (Cappelletti et al., 2013; Prichard et al., 2014; Snowball et al., 2013; Terney et al., 2008), its cell-level operating mechanisms are only starting to be explained and speculated. The two most likely mechanisms are, at systems level, stochastic resonance, and, at cell level, the repeated potentiation of voltage-gated channels (Chaieb et al., 2009). Stochastic resonance is a phenomenon that occurs when an appropriate measure of information transfer (such as signal-to-noise ratio) is maximised in the presence of a specific level of random noise, thereby lowering the response threshold. In other words, the system resonates at a particular noise level, and so the presence of noise then becomes beneficial to the system. The other putative mechanism is the neuronal activation of sodium channels, which may induce LTP-like cortical plasticity (Kuo & Nitsche, 2012).

In the field of numerical cognition, several studies increased the interest in using tDCS or tRNS to increase basic understanding of number representation in the brain and to aid in maths enhancement and possibly in rehabilitation programs. Hauser, Grabner, & Jäncke (2013) found that parietal tDCS applied as left-anodal right-cathodal (LARC), but not in the opposite configuration (RALC), enhanced performance in number comparison and simple arithmetic tasks. On the contrary, Cohen Kadosh et al. (2010) found that only the RALC configuration, applied to the same parietal site, enhanced the speed of acquisition of artificial digits in a magnitude comparison task. These artificial digits were later used in a numerical learning paradigm with two dyscalculic adults (Iuculano & Kadosh, 2014). tDCS was applied to the posterior parietal cortex (PPC), in LARC or RALC configuration for the two subjects respectively, and performance relating to automaticity of number processing and number line acuity (using the artificial digits) was recorded. Unlike with healthy subjects (R. Cohen Kadosh et al., 2010), it was only the LARC configuration that improved performance on both tasks. Kasahara et al. (2013) asked subjects to perform a mental calculation task both in the scanner – where the laterality index of their parietal activity, representing the relative contributions of the two parietal cortices, was evaluated using fMRI – and outside of the scanner. Parietal tDCS applied as LARC shortened response times in subjects with left-hemispheric parietal lateralization, but not in subjects with bilateral parietal activation. Clemens et al. (2013) asked healthy adults to solve repeated presentations of simple multiplication problems while administering anodal tDCS (or sham, within-subjects) over the right angular

gyrus (AG). Subjects were fMRI-scanned before and after each tDCS condition. Although no improvements in task performance were observed after a single session of tDCS, brain activation was altered in the sense of increased involvement of the bilateral AG when compared to task performed during sham or no tDCS controls (which echoes findings from Snowball et al. (2013), that I will discuss below). Setting out to uncover potential unwanted side-effects of enhancements such as the one described thus far in this section, Iuculano and Cohen Kadosh (2013) trained healthy adults for 6 days in tasks involving learning and automaticity in the processing of the same artificial set of digits. tDCS to the posterior parietal cortex facilitated learning while impairing automaticity, whereas the opposite was true for tDCS to the dorsolateral prefrontal cortex. These results suggested that, as hypothesised by the authors, enhancement of one cognitive function may come at the expense of another. In summary, these studies leave open the question of the polarity and laterality that tDCS is to be used with in order to optimally enhance numerical functioning, ideally without unwanted interference with other types of functioning. This is one of the reasons why researchers have begun employing tRNS towards the same purpose, as this eliminates the need of making a choice between the two main tDCS configurations (LARC and RALC).

Snowball et al. (2013) showed that tRNS of the bilateral dorsolateral prefrontal cortex, applied during a 5-day training of arithmetic learning, improved speed of calculation- and retrieval-based problems. Transfer of the improved performance to new (untrained) problems was observed in a post-training testing session. These group changes remained when tested again after 6 months, suggestive of long-term behavioural and physiological alterations in the stimulated group. Cappelletti et al. (2013) coupled transcranial random noise stimulation (tRNS) applied bilaterally and training aimed at improving the precision of the ANS. Improvement was sharper and longer-lasting (16 weeks post-training) when parietal tRNS was applied during training as opposed to training alone, parietal tRNS without training or motor tRNS applied during training. Only in said condition the increased ANS acuity transferred as increased proficiency in relative judgements of space and time, dimensions proposed to be processed in the same part of the inferior parietal cortex (Walsh, 2003), but not to domain-general control functions such as attention, executive functions, and visual pattern recognition. In summary, these limited number of tRNS studies suggest that this type of stimulation, coupled with cognitive training, can enhance numerical and arithmetic skills, and that these changes can outlive the duration of the stimulation and training.

1.5.2 Using structural MRI to quantify (differences in) cognition

It has, for a while, been well established in cognitive neuroscience that there is a "feedback loop" between brain and behaviour, where changes in either affect the other (for a review of plastic processes in the brain, see e.g. Pascual-Leone, Amedi, Fregni, & Merabet, 2005). Although the way that the brain gives rise to behaviour is most intuitive in visualisations of functional activation, as depicted for instance by fMRI, this function is necessarily dependent on the brain's structure, specifically on the complex patterns of organisation and connectivity that characterise its two main types of tissue, grey and white matter (for a review, see Zatorre, Fields, & Johansen-Berg, 2012). For instance, the efficiency of any of its neural networks is dependent of the strength of synaptic connections, themselves affected by processes such as synaptic pruning, which after stabilisation leads to a certain measurable cell packing density. Also, *functional* connectivity in the brain (i.e., activity between regions that is correlated across time) is dependent on an existing infrastructure of WM bundles (i.e. *structural* connectivity). Given the difficulty of measuring fine-grained neural processes such as synaptic pruning or fibre myelination, an acceptable compromise is offered by neuroimaging methods, which provide us with rather 'holistic' quantifications of these phenomena for a given region or across the whole brain, as detailed in other parts of this thesis.

1.5.2.1 Grey matter

There are two main approaches to analysing anatomical MRI scans in a way that gives information about the structure of the cortex. The first approach is voxel-based morphometry (VBM), which takes the brightness level of each voxel in the image as an indication of the relative concentration of tissue imaged in that voxel, expressed in measures such as grey matter volume or density (GMD). The second approach is surface-based morphometry (SBM), which, rather than working directly on the discrete elements of the 3D anatomical image (i.e. on the voxels), instead takes that information to build a mathematical model (mesh) of the main surfaces that describe the brain, i.e. those lying at the interface between grey and white matter, and between grey matter and the cerebrospinal fluid (CSF). Statistics are then performed on measures derived from these generated surfaces, which can refer to the thickness and degree of gyrification of the cortex.

Figure 2 illustrates graphically the different approach taken by the VBM and SBM methods, each of which is further described in the following subsections.

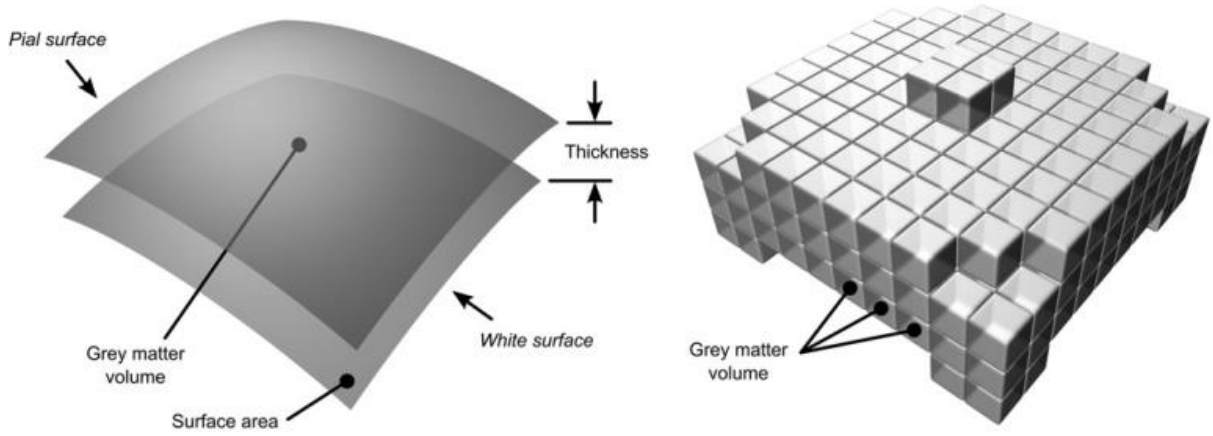


Figure 2: Surface-based (SBM; left) versus voxel-based morphometry (VBM; right). Reproduced, with permission, from (Winkler et al., 2010).

1.5.2.1.1 VBM

I used VBM, implemented in FSL, in Studies 1 and 3, in order to compute and compare regional concentrations of grey matter, i.e. grey matter density (GMD). I followed the optimised VBM procedure (Good, Johnsrude, Ashburner, Henson, Friston, et al., 2001), implemented in FSL (FMRIB's Software Library, Oxford, UK). Briefly, in this procedure, all structural images are brain-extracted, segmented according to tissue type (white matter (WM), grey matter (GM), cerebrospinal fluid (CSF)), aligned to standard space, nonlinearly registered and modulated (divided by the Jacobian of the warp field used for registration). The 'optimised' adjective in the procedure's name refers to this Jacobian modulation, which compensates the spatial transformation by making appropriate compensations in terms of image intensity. The images thus processed are then smoothed in order to increase signal-to-noise ratio. Finally, voxel-wise general linear model (GLM) is applied using FSL's permutation-based non-parametric testing, with corrections for multiple comparisons across space. For group comparisons, group is entered into the design matrix as a categorical predictor, while other factors that are not directly of interest but that nevertheless have to be controlled for, such as gender and age, are entered as predictors of no interest (also known as nuisance variables). For brain-behaviour correlations, a behavioural score is additionally entered as a continuous predictor. The entire VBM processing pipeline is summarised schematically in Figure 3.

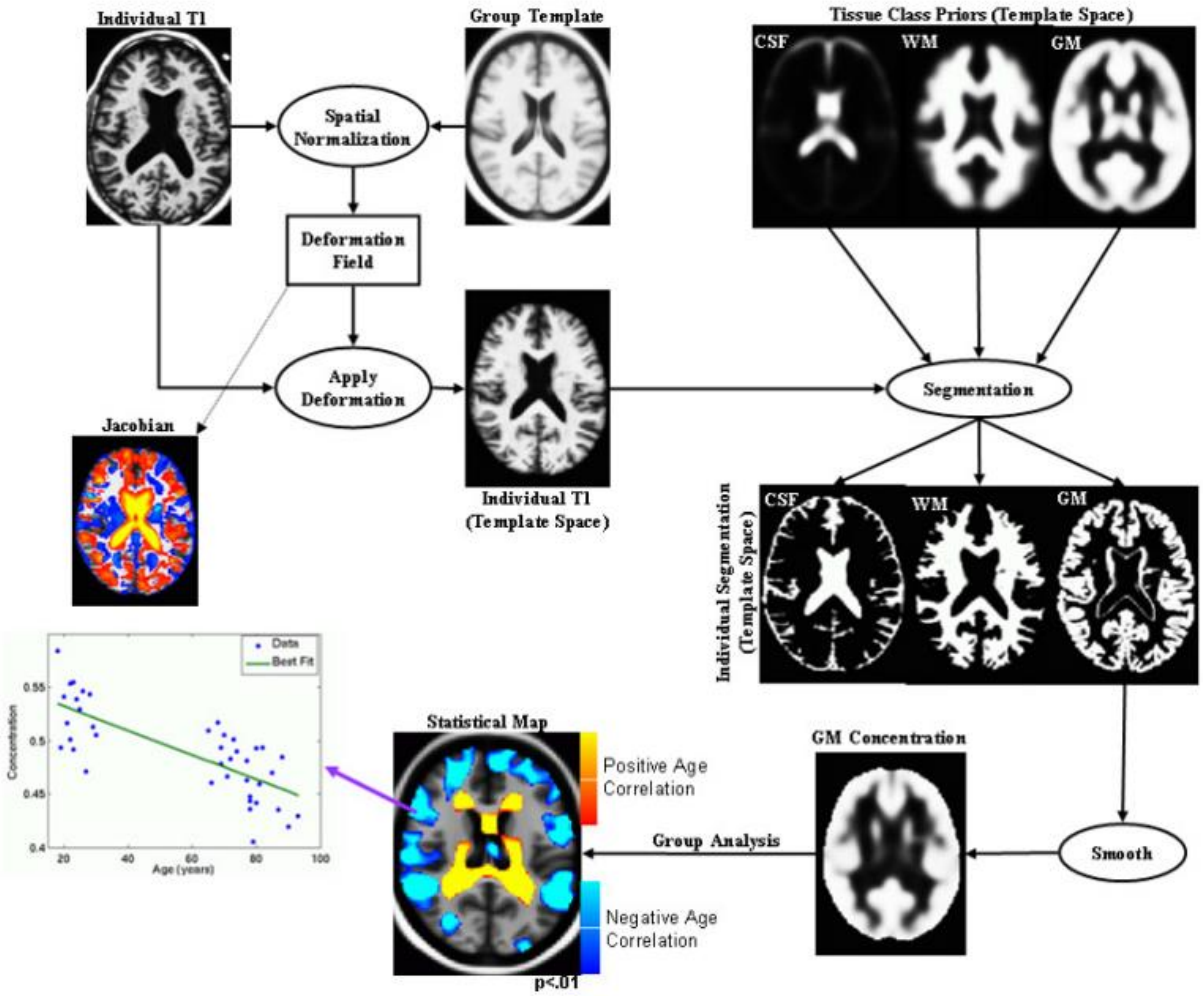


Figure 3: Schematic of the steps involved by the VBM procedure. Reproduced, with permission, from (Greve, 2011).

1.5.2.1.2 SBM

In Studies 1 and 3 (Chapters 2 and 4 respectively) I used the SBM approach, implemented in FreeSurfer (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999; Fischl, Sereno, Tootell, & Dale, 1999), in order to complement the VBM measure of GMD with additional measures such as cortical thickness and surface area. SBM involves an initial number of pre-processing steps, similar to VBM. Briefly, these includes motion correction, removal of non-brain tissue, automated Talairach transformation, tissue segmentation, intensity normalization, tessellation of the grey/white matter boundary, automated topology correction, and surface deformation following intensity gradients to optimally place the grey/white and grey/CSF borders at the location where the greatest shift in intensity defines the transition to the other tissue class. The following operations are then carried out on the cortical models: surface inflation, registration to a spherical atlas using individual cortical folding patterns to match cortical geometry across subjects, parcellation of the cerebral cortex into units based on gyral and sulcal structure using a Bayesian procedure (Desikan et al., 2006), and

creation of a surface-based data including maps of curvature and sulcal depth. This method uses both intensity and continuity information from the entire three dimensional MR volume in segmentation and deformation procedures to produce representations of cortical thickness, calculated as the closest distance from the grey/white boundary to the grey/CSF boundary at each vertex on the tessellated surface.

1.5.2.1.1 Comparison between VBM and SBM

VBM and SBM produce metrics (e.g. GMD and cortical thickness, respectively) that are often strongly correlated (Bermudez et al., 2009; Foster & Zatorre, 2010b; Schaer & Eliez, 2009). The two approaches are thus complementary in the sense that a VBM result of altered GMD can essentially be explained by altered thickness, altered surface area, or a combination of the two. Unlike VBM, which provides an indirect estimation of the relative concentration of cortical tissue, SBM provides a biologically meaningful metric that is easier to verify against "gold standard" manual measures of cortical thickness from post-mortem studies (Fischl & Dale, 2000; Rosas et al., 2002). The differences between them motivate a multianalytical approach to the same data set, with the aim of possibly strengthening the reliability of an effect.

The two methods are each reliable predictors of how cortical degeneration contributes to various neurodevelopmental conditions (Chung et al., 2005; and e.g. De Lange et al., 2008; Fletcher & Witoonpanich, 2013 for GMD; see e.g. Shaw et al., 2007; Thompson et al., 2004 for cortical thickness), as well as being predictors of expertise in various domains (Aydin et al., 2007 for GMD; and Bermudez et al., 2009 for both; see e.g. Elmer, Hänggi, Meyer, & Jäncke, 2013 for cortical thickness; Maguire et al., 2000).

Image registration from native to standard space is more accurate in SBM, as smoothing follows anatomical boundaries more accurately than the three-dimensional smoothing used in VBM. SBM group analyses employ inter-subject alignment based on the patterns of sulci and gyri, as opposed to Talairach registration, which often ignores sulcal and gyral landmarks (A. Klein et al., 2010; Tucholka, Fritsch, Poline, & Thirion, 2012). Also, SBM has increased sensitivity: with approximately 20 times fewer vertices in a cortical mesh model than voxels in a typical 1mm-resolution T1 image used in VBM, the problem of multiple comparisons in a whole-brain analysis is diminished.

1.5.2.2 White matter

1.5.2.2.1 Diffusion pre-processing and measures

The network of structural (synaptic) connections linking sets of neurons in the brain is what concerns the field of anatomical connectivity. In animal models, invasive tract-tracing techniques are considered the gold-standard for making direct inferences about direct axonal connections and overall connection strength (for a review, see Johansen-Berg & Rushworth, 2009). Researchers trying to relate individual difference in the human brain structure to cognitive functions, however, have for the moment to be content with only whole-brain in-vivo markers of temporal changes in fibre tracts provided by diffusion weighted imaging techniques, such as diffusion tensor imaging (DTI). Due to the limited spatial resolution of this technique (order of millimetres), measures computed voxel-wise from the diffusion tensor are necessarily averaging over a large number of axons included in that voxel. For instance, the fractional anisotropy (FA) measure exploits the fact that water diffuses more readily along versus across axons, and gives a normalised (range: 0 to 1) measure of this diffusion, with values closer in absolute value to 1 representing more diffusion constrain and thus higher WM integrity; see Figure 4. It is assumed that increased integrity of WM structures linking cortical regions relevant to a specific domain contributes to increased behavioural performance in that domain by strengthening the functional connectivity between the connected cortical areas (Johansen-Berg & Rushworth, 2009).

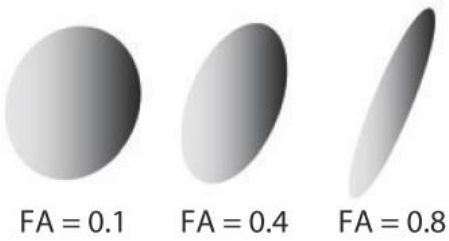


Figure 4: Fractional anisotropy (FA), a measure of the directionality of water diffusion in white matter. Voxels having an FA value closer to 1 correspond to a more coherent section of the white matter bundle. Reproduced, with permission, from (Johansen-Berg & Rushworth, 2009).

A diffusion data set consists of several diffusion-weighted images, each image being acquired after the scanner has been sensitised to water diffusion along a particular direction in space. Typically, 60 diffusion directions are used; a non-diffusion image is also acquired for each subjects, for aligning purposes. Before being used for voxel-wise statistics, each subject's diffusion images are first pre-processed in a number of steps. Correction for Eddy currents and for head motion is done by aligning each subject's data to the non-

diffusion volume. Images are then brain-extracted in order to eliminate non-brain voxels. The diffusion tensor model is then fitted to each voxel, and produces a vector indicating the dominant direction of diffusion within that voxel. It also produces scalar (numerical) descriptions of that average diffusion properties in that voxel; the most important of these measures is FA but other measures of the average diffusion in a voxel can also be computed, that are complimentary to FA – these include axial (AD), radial (AD) and mean (MD) diffusivity[3]. The complete set of tensors across the entire brain is thus indicative of the directions of fibres connecting the different parts of the brain. These anatomical fibres can be probabilistically reconstructed using a tractography procedure, described below.

1.5.2.2.2 TBSS

Subsequent to pre-processing, diffusion images are processed according to the steps of the Tract-based Spatial Statistics (TBSS) procedure (Smith et al., 2006). TBSS projects each subject's FA data onto a group-mean spatially restricted image called the *skeleton.* This allows voxelwise analyses across subjects which in turn enable group comparisons of measures derived from the tensor model, such as FA and MD.

1.5.2.2.3 Tractography

The probabilistic tractography implementation of FSL, ProbTrackX, allows the user to select specific connections to investigate, from the total number of tracts computed during the previous steps of the diffusion pipeline (explained above). Specifically, an area of the brain (close to the GM/WM border) is chosen as the "seed", and from there paths are sought to surrounding areas, based on the diffusion profiles of voxels adjacent to the seed voxel(s). Essentially, a high FA value in a voxel indicate the presence of a WM tract passing through that voxel's principal diffusion direction, and the tractography path is therefore more likely to proceed along that direction than along those of adjacent voxels. Waypoints can be defined that guide the tracking process through certain parts of the brain (usually based on a prior hypothesis). ProbTrackX samples from the distributions of voxel-wise principal diffusion directions, each time computing a streamline through these local samples. By taking many such samples, the algorithm is able to build a histogram of the streamline location or the connectivity distribution. The output of tractography is a statistical map whose voxel intensity represents the relative likelihood of that voxel being "reached" when starting the tracking process from the given seed, while passing through any desired waypoints.

[3] More technical details about these diffusion measures can be found in Study 1's Methods (section 2.3.4.2, page 1)

1.5.2.3 Previous numerical cognition results using this methodology

Assuming, as explained above, a link between brain function and brain structure, it is all the more surprising that, despite the last decade's surge in studies investigating the *functional* correlates of mathematical cognition (e.g. Andres, Pelgrims, Michaux, Olivier, & Pesenti, 2011; Ansari, Grabner, Koschutnig, Reishofer, & Ebner, 2011; Clemens et al., 2013; R. Cohen Kadosh et al., 2005, 2011; Delazer et al., 2003, 2005), there is a clear gap in the complementary literature, that of *structural* correlates underlying this type of cognition. This is evident, for instance, from the fact that the number of brain structures associated functionally with maths, i.e. the number of fMRI studies, was, even a few years ago, sufficient to allow Arsalidou and Taylor (2011) to use meta-analysis methodology such as activation likelihood estimation (ALE) in order to compute a normative fMRI atlas for both number and calculation tasks. Structural investigations are, on the other hand, limited enough in number that such a meta-analysis with structural coordinates is (to date, and as far as I am aware) yet to be performed. Instances of investigations into the structure of the mathematical cortex include the study by Aydin et al. (2007), where mathematicians and non-mathematicians were compared in terms of differences in local grey matter density examined at the whole-brain level, and the (parietal and frontal) differences found were correlated against number of years spent as a mathematician.

On the other hand, from the white matter perspective, studies aiming to unravel the connectivity of the mathematical brain are often motivated by the assertion of the triple-code model (Dehaene et al., 2003) that numerical cognition relies on multi-modular and distributed processing. It is likely that the existence of this testable hypothesis has led to relatively more numerous studies investigating connectivity and WM integrity as compared to cortical organisation. Many of these studies have found FA in several fibres of the brain to be positively correlated with performance in either exact or approximate arithmetic, suggesting a role for the efficient propagation of signals along these particular fibres for proficiency in mental arithmetic. For instance, Barnea-Goraly et al. (2005) asked children and teenagers with velocardiofacial syndrome to perform mental arithmetic in the scanner, and found that their scores correlated with FA in left parietal WM structures adjacent to the left IPS cortex. Further studies – discussed below – later addressed doubts about the generalisability of these results by searching for FA correlates of mental arithmetic in typically-developing children. Tsang et al. (2009) found that mean FA extracted from the left (but not the right) anterior portion of the superior longitudinal fasciculus (aSLF), a dorsal stream bundle connecting parietal and frontal cortex,

positively correlated with individual variability in the children's arithmetic approximation skill. The correlation was not explained by shared variance with more broad number and general-cognitive abilities. Also, van Eimeren et al. (2008) found that scores from the mathematical components of the Wechsler Individual Achievement Test (WIAT; Wechsler, 2005) correlated positively with FA in left hemisphere bundles, even after accounting for variance explained by chronological age and individual differences in general RTs. Specifically, FA in the left inferior longitudinal fasciculus predicted Numerical Operations scores, while FA in the left superior corona radiata predicted Numerical Operations as well as Mathematical Reasoning scores. Finally, Qin et al. (2014) found that the procedural-to-memory shift in arithmetic strategy taking place in childhood and adolescence is mediated by an inverse U-shaped increase in functional connectivity between the hippocampus and number-relevant parts of the neocortex. Overall, these studies suggest that quantifying structural parameters of relevant parts of the brain may be predictive of individual differences in strategy use and also overall maths competency.

1.6. This thesis

The main aims of my DPhil thesis are twofold. On the one hand, to fill in the existing gap between the knowledge that we have from magnetic resonance imaging on how the expert mathematical brain *behaves* differently, and the more limited current understanding of how the brain is *changed* by acquired expertise. On the other hand, it aims to understand how transcranial electrical stimulation can enhance basic numerical and arithmetic abilities in the healthy adult. The first three studies address these two aims in a complementary fashion. These main aims are then supplemented by a final neuropsychology study that examines the way that the building block of mathematics – number – interacts with two other dimensions that are likely to share processing resources in the parietal lobe, namely space and pitch. The approach can be seen as a different type of intervention from TMS studies seeking to find causal relationships between brain and behaviour by inducing virtual transient "lesions" in the brain. Here, patients with lesions whose immediate impact is numerical cognition were used, and the more generalised impact on potentially related areas of cognition, like music, was investigated.

2 Study 1: A cross-sectional snapshot into the structure of the mathematical brain

2.1. Abstract

Much more is known about the neural correlates of mathematical deficits (such as dyscalculia) than those of extensive experience in mathematics. Studies in several domains of expertise have established that experience-dependent plasticity is associated with not just functional but also anatomical changes. However, the exact nature and direction of these practice-induced changes remains unclear. Here, I aimed to identify cognitive and structural differences (in terms of grey matter, GM, and white matter, WM) that are associated with long-term training in mathematics. Mathematicians and non-mathematician academics, matched for age, gender and number of years of education, underwent structural and diffusion MRI scans and were tested on a broad range of cognitive dimensions, including working memory, attention, IQ, numerical and social skills. Mathematical expertise was associated with better performance in a diverse range of behavioural tasks, even for dimensions that are not purely mathematical, such as fluid intelligence, logic, and verbal reasoning. Anatomically, for mathematicians relative to controls I found increased GM density in the right superior parietal lobule (SPL) but decreased GM density in the right intraparietal sulcus (IPS) and left inferior frontal gyrus (IFG), and decreased surface area in the left fusiform gyrus (FG). For controls, behavioural performance correlated negatively with GM density and local gyrification index in the left superior frontal gyrus (SFG). In terms of WM, there were no significant differences in fractional anisotropy, however tractography revealed a stronger right SFG to left SFG tract in mathematicians than in controls. In sum, these results suggest a non-linear relationship between mathematical expertise and GM metrics in parietal and prefrontal regions previously implicated in numerical cognition. Although the cross-sectional

nature of this study prevents causal inferences, it can be speculated that a bidirectional feedback loop exists between experience and structural brain changes, and that the nature of this loop changes across the lifespan.

2.2. Introduction

With the ever-growing presence of numerical information in our environment, those endowed with good abilities of understanding, evaluating and manipulating numbers and mathematical information have a greater chance to do well in life (Estrada, Martin-Hryniewicz, Peek, Collins, & Byrd, 2004; Parsons & Bynner, 2005). Furthermore, it has been suggested that the contribution of numeracy to different indices such as employment, education, and physical and mental wellbeing is even more substantial than that of literacy (Beddington et al., 2008; Bynner & Parsons, 1997). The exact ways in which our brain determines these abilities is, however, not yet completely understood. Additionally, much more is known about the neural correlates of mathematical deficits than about those of extensive experience in mathematics.

2.2.1 The neuroanatomy of mathematical expertise

Several neuroimaging studies with healthy volunteers have consistently identified a number of brain regions that are involved in the performance of arithmetic tasks (Andres et al., 2011; Grabner, Ansari, et al., 2009; Kong et al., 2005; Rivera, Reiss, Eckert, & Menon, 2005). There is evidence (reviewed in meta-analysis by R. Cohen Kadosh, Lammertyn, et al., 2008) that the intraparietal sulcus (IPS), and in particular its horizontal segment (HIPS), is the region that is mostly reported in tasks that involve numerical representation. The network of areas activated during number processing is, however, much wider, and is distributed across the parietal lobe (Delazer et al., 2003; Grabner, Ansari, et al., 2009; Wu et al., 2009), as well as involving the prefrontal cortex (as evidenced by Rivera et al., 2005 in tasks using mental arithmetic). The degree to which the frontal and parietal lobes complement each other in number processing varies with development, with the latter gradually taking on more functions (Ansari, Garcia, Lucas, Hamon, & Dhital, 2005; Cantlon, Brannon, Carter, & Pelphrey, 2006; Rivera et al., 2005).

There is inconsistency in the functional neuroimaging literature as to whether deficient – or, on the other hand, exceptional – performance in a task is linked to *more* or to *less* functional activation of the areas relevant to expertise (see review by Kelly & Garavan, 2005). In the number domain, studies that have looked

at children with mathematical difficulties have found both *less* activation of the IPS than in normal children during a number comparison task (Price et al., 2007; using functional magnetic resonance imaging, fMRI) and *more* activation of the right inferior and right superior parietal regions during a simple arithmetic task (Simos et al., 2008; using MEG). On the other hand, two studies that investigated the morphological brain structure changes associated with mathematical training both seem to suggest that "more is better". First, Hu et al. (2011) found that children who started early intensive training in using the abacus show enhanced integrity of white-matter tracts related to motor and visuo-spatial processes, in comparison with children not trained with the abacus. Second, the only study so far – to my knowledge – to investigate brain structure in mathematicians (Aydin et al., 2007) found increased grey matter density (GMD) in the left inferior frontal gyrus and bilateral inferior parietal lobule for mathematicians vs controls; these increases correlated with the number of years spent as a mathematician. However, this study has critical limitations, as the level of mathematical education was not controlled for in the two groups, and subjects were not cognitively tested, which did not enable the authors to draw any conclusions about behavioural group differences, or about brain-behaviour correlations.

2.2.2 Supporting functions for mathematics

Psychologists have long been trying to find which cognitive abilities are associated with exceptional mathematical ability. A 19th century report (Scripture, 1891) that made a psychological "meta-analysis" of the characteristics of several mathematical prodigies pointed to a few common traits, which included good working memory capacity, an episodic memory loaded with a plethora of arithmetic facts, and good visual imagery abilities; the preference for the visual modality in the mathematician's creative process is also confirmed in other qualitative reports (Hadamard, 1949). Despite the anecdotal nature of these conclusions, they are confirmed by much more recent studies (reviewed in Pesenti, 2005). The role of supporting (auxiliary) functions – such as working memory, attention and mental imagery – to numerical competence is evident both from studies in which these functions are impaired (Passolunghi & Siegel, 2004; K. M. Wilson & Swanson, 2001) and from studies suggesting that their good functioning plays an active role in mathematical reasoning (O'Boyle, 2008; Pesenti et al., 2001). It is for this reason that, in this study, I used not only numerical tasks to investigate mathematical abilities, but also non-numerical tasks that tap into these general cognitive abilities.

I also aimed here to investigate the link between high-level vs basic numerical abilities. The latter is embodied in the concept of the 'approximate number sense' (ANS), a cognitive faculty believed to represent the cardinality of sets of objects in an approximate fashion (Butterworth, 2010; Halberda, Mazzocco, & Feigenson, 2008). The acuity of the ANS can be determined by using a paradigm that requires discriminating the numerosity (cardinality) of two sets of dots (Halberda et al., 2008; Piazza et al., 2010); the ratio formed by the two numerosities determines the difficulty of the task: people find the task easy when the difference between the numerosities is large (such as 2 vs 8), but the task becomes difficult when the numbers are closer to each other and thus the ratio closer to 1 (such as 6 vs 8). Response time and accuracy data from such a task typically demonstrates the reliable numerical ratio effect, whereby numerosities that form a larger ratio, such as 6:8, are responded to slower and less accurately than those that form smaller ratios, such as 3:6. The extent of this difference in terms of response times (e.g. the slope of the RT as a function of ratio curve) quantifies the size of the numerical ratio effect – a metric of the ANS. By varying the ratio across trials, it is also possible to determine the precision of an individual's ANS, indexed by the Weber fraction, *w*, a number that quantifies the amount of noise in the ANS; smaller *w* values indicate more precise representations, leading typically to more a pronounced numerical ratio effect as measured by the first metric. Intuitively, *w* is the proportional difference that is reliably detected in dot arrays (non-symbolic numerosity) discrimination tasks. It determines how rapidly performance rises from chance (50%) to near-asymptotic (100%) in such tasks. Specifically, *w* is defined as WR-1, where WR is the Weber Ratio given by the larger/smaller set. On average, the *w* of adults is approximately 0.11, corresponding to successful non-verbal discrimination of arrays differing by as little as a 9:10 ratio (WR=1.11). The exact procedure used for computing *w* here is described at the end of section 2.3.2.

Previous studies have found that ANS acuity (the slope in RTs or the *w*) correlates with either past or future higher-level mathematical performance, such as scores in standardised tests of arithmetic, in children but not in adults (Halberda et al., 2008; Inglis, Attridge, Batchelor, & Gilmore, 2011; Libertus, Feigenson, & Halberda, 2011; Mazzocco, Feigenson, & Halberda, 2011b). Moreover, deficits in measures of ANS have been found in children with developmental dyscalculia (Mazzocco, Feigenson, & Halberda, 2011a; Mussolin, Mejias, & Noël, 2010; Piazza et al., 2010; Price et al., 2007). While such results speak of the potential of ANS acuity to predict performance in higher-level maths, some authors have argued that differences in ANS acuity in fact predominantly reflect differences in non-numerical factors such as inhibition rather than reflecting an improved number sense acting as a building block for general maths ability (Szucs et al., 2013); while other

studies have not found a link between the two altogether, in both adults (Price et al., 2012) and children (Soltész, Szűcs, & Szűcs, 2010).

2.2.3 The current study

The current study aimed to uncover whether and how mathematicians are different from non-mathematicians controls – both in terms of behavioural performance and in terms of the brain structure underlying this performance. The latter was quantified by using diffusion measures for white matter, and complementary voxel- and surface-based morphological measures of grey matter. The former was assessed using a variety of tasks; this allowed me to tease apart whether mathematical skills are associated with superior numerical abilities only or also to non-numerical cognitive functions such as intelligence, working memory and mental imagery. Based on the literature discussed earlier, I expected to find cortical differences between mathematicians and controls in prefrontal and parietal cortices, as well as differences in the integrity of the white matter tracts radiating from these structures. Moreover, I validated whether these changes are mathematically related by correlating the different brain indices with behavioural performance.

2.3. Methods

2.3.1 Subjects

Thirty eight subjects (all right handed; 10 female) were recruited from the University of Oxford. All were academics either at doctoral or post-doctoral level. The mathematicians group consisted of 19 subjects (14 males, μ=25.7, σ=1.6, range=23.2 to 32.9) working in various areas of mathematics (such as Algebra, Logic and Number Theory) at the Mathematical Institute of the University of Oxford. The control group also consisted of 19 subjects (14 males, μ=26.3, σ=2.4, range=24.2 to 30.4) from the field of humanities (Departments of English, Modern Languages, Classics and History). Subjects were reimbursed at £10 per hour. The research has been approved by the Berkshire Research Ethics Committee.

2.3.2 Behavioural tasks

I used an extensive battery of tests to map possible group differences in several cognitive abilities. Due to word limitations, these are only briefly summarised in Table 2. More details are included about the number acuity task, in the section below.

Table 2: Summary of the battery of cognitive tests used.

Cognitive category	Test name	Reference
Intelligence	IQ test[4]	(Wechsler, 1999)
Working memory	Digit span (backward&forward)	(Wechsler, 1997)
	Letter span (forward)	
Attention	Attention networks task (ANT)	(Fan, McCandliss, Sommer, Raz, & Posner, 2002)
Mental imagery	Mental rotation task (MRT)	(Peters et al., 1995)
Numerical skills	Number acuity	(based on Piazza et al., 2004)
	Number line task	(based on Siegler & Opfer, 2003)
	Numerical Stroop	(Henik & Tzelgov, 1982)
	Numerical agility[5]	(designed in the lab)
	Numerical strategies	(based on Levine, 1982)
	Arithmetic task	(based on Wechsler, 1997)
Logic	Wason logic task	(Wason, 1966)
Verbal reasoning	Verbal reasoning task	(Association of American Medical Colleges, 2007)
Social skills	Emotion recognition task	(K. Cohen Kadosh, Henson, Cohen Kadosh, Johnson, & Dick, 2010)
	Gaze task	(K. Cohen Kadosh et al., 2010)
	Face recognition task	(Benton, Van Allen, Hamser, & Levin, 1978)
	Autism spectrum quotient (ASQ)	(Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001)

Additionally, all subjects completed a questionnaire asking them to rate (on a scale from 1 to 10) the degree to which they use the following different modalities (strategies) when solving arithmetical problems: visuospatial (trying to visualise the numbers or equations), outer verbalisation, inner verbalisation or kinaesthetic (finger counting).

In order to test the acuity of the ANS, I used a nonsymbolic number comparison paradigm, with paired simultaneous presentation of two arrays of dots. Trials contained comparisons between 16 or 32 dots in one array (the "fixed" numerosity, or n1) and 12, 13, 14, 15, 17, 18, 19, 20 dots, or 24, 26, 28, 30, 34, 36, 38, 40 dots respectively in the other (the "variable" numerosity, or n2), thus leading to 8 possible ratios between the two arrays, shown in Table 3. There were thus 2 combinations for each ratio, and there were 5 repetitions for each combination, for a total of 80 trials, the order of which was randomised between subjects. The left/right position of n1 and n2 was balanced.

[4] Wechsler Abbreviated Scale of Intelligence (WASI), First Edition

[5] In this in-house task, subjects had to creatively apply the basic arithmetical operations to four given numbers in order to arrive at another given number. There were a total of 5 problems, of increasing difficulty.

Table 3: Number of dots compared, and corresponding ratio.

Number combinations	16:12 32:24	16:13 32:26	16:14 32:28	16:15 32:30	16:17 32:34	16:18 32:36	16:19 32:38	16:20 32:40
Ratio	0.75	0.8125	0.8750	0.9375	0.9411	0.8888	0.8421	0.8

The dot diameter in n1 was randomly assigned within a certain range centred around a mean value; for n2, on half of the trials, dot diameter was fixed at the mean value, and in the other half, it was randomised in the same way as for n1. Dot dispersion (distance from the dot to the centre of the white disc) was manipulated in the same way. Subjects were asked to choose the disc that contains the larger number of black dots, a choice made using the keys Q (left) and P (right) of a computer keyboard. The task was administered using a script written in MATLAB (Mathworks Inc.). Reaction times and accuracy of responses was recorded. A sample trial is pictured in Figure 5.

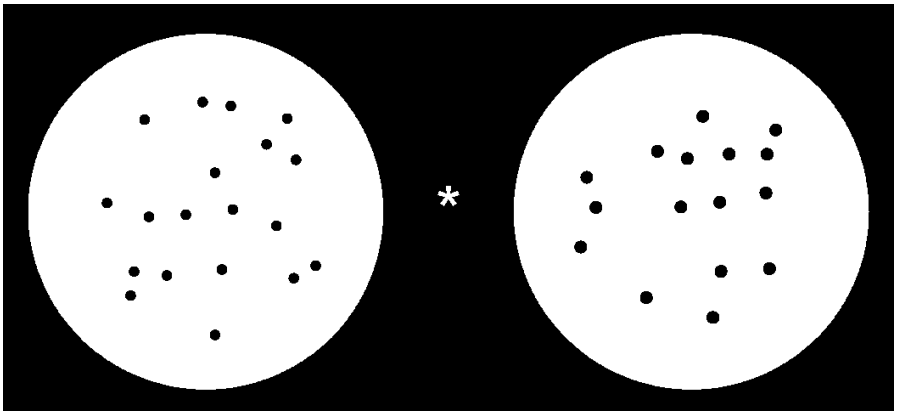


Figure 5: Number acuity task; subjects were asked whether the higher number of dots was contained in the left or in the right array.

The slope of the RT=f(ratio) line was determined, for each subject, as the unstandardised linear regression coefficient (β). RTs for correct trials were then averaged across subjects for each of the 8 ratios.

The accuracy data were used to determine a percentage of correct responses, at each ratio. The Weber fraction, *w*, was then determined as the value that minimised the difference between the observed percentage and the percentage predicted by the formula below (Halberda et al., 2008):

$$\mathrm{ER} = 1 - \frac{1}{2}\,\mathrm{erfc}\left(\frac{n_1 - n_2}{\sqrt{2}\,w\sqrt{{n_1}^2 + {n_2}^2}}\right),$$

where ER is the error rate and erfc is the complementary error function related to the integration of the normalised Gaussian distribution. The model fits percentage correct as a function of the Gaussian

approximate number representations for the two sets displayed on each trial (n_1 and n_2) with w as the single free parameter.

2.3.3 MRI acquisition

All subjects were scanned using the Siemens 3T scanner at the Oxford Centre for Functional Imaging of the Brain (FMRIB). T1-weighted anatomical images were acquired in the sagittal plane using an MPRAG sequence (TR=15.7 ms; TE=4.6 ms; flip angle=8, inversion time=900 ms, voxel size=1 x 1 x 1 mm). In addition, whole brain diffusion weighted volumes were acquired (60 directions; b=1000 $s.mm^{-2}$; 65 slices; voxel size 2 x 2 x 2 mm ; repetition time (TR)=9.6 s; echo time (TE)=87 ms) plus eight volumes without diffusion weighting (b=0 $s.mm^{-2}$).

2.3.4 Analyses

2.3.4.1 Grey matter

The grey matter analysis was performed using two separate approaches: voxel-based morphometry (VBM), conducted using the FMRIB Software Library, FSL (Oxford Centre for Functional MRI of the Brain, University of Oxford, UK), and surface-based morphometry (SBM)[6], conducted using FreeSurfer (Massachusetts General Hospital, Harvard University, Cambridge, MA, USA).

2.3.4.1.1 VBM

In order to compare GMD[7] across groups, I used Voxel-Based Morphometry (VBM), which compares local grey matter tissue volumes at the voxel level. For the whole-brain level analysis, I followed the optimised VBM procedure (Good, Johnsrude, Ashburner, Henson, Friston, et al., 2001). Namely, all structural images were brain-extracted, segmented according to tissue type (white matter (WM), grey matter (GM),

[6] Note that while VBM is the "official" name of the specific volume-based technique described in (Good, Johnsrude, Ashburner, Henson, Fristen, et al., 2001; Good, Johnsrude, Ashburner, Henson, Friston, et al., 2001), "surface-based morphometry" (SBM) is, at least at the time of writing, rather an umbrella term referring to techniques that extract measures such as cortical thickness and surface area not directly from the volumetric T1-weighted structural images but from the mesh models derived from those images, representing the cortical surfaces. It is for the sake of simplicity of comparing between the two approaches that I am using the acronyms VBM and SBM.

[7] The dependent variable used in VBM is a measure of GM concentration, referred to as either GM volume (GMV) or GM density (GMD). Originally, the distinction between the two terms had to do with the presence or absence (respectively) of normalisation by the Jacobian, which is normally performed, in accordance to the predominant (optimised) VBM algorithm (Good, Johnsrude, Ashburner, Henson, Fristen, et al. 2001). This distinction is, however, more nominal than meaningful, and since the latter term is used predominantly in the literature while maintaining the suggestion of relative concentration, it is this term that I will be using here. The term 'density' has to be understood as relative amount of GM and does not represent cytoarchitectonic measure such as cell packing density.

cerebrospinal fluid (CSF)), aligned to MNI space, nonlinearly registered and modulated (divided by the Jacobian of the warp field used for registration). The images thus processed were smoothed using a Gaussian kernel with σ=4 mm (corresponding to a full-width at half-maximum of 9.4 mm), and voxel-wise general linear model (GLM) was applied using FSL's permutation-based non-parametric testing, with corrections for multiple comparisons across space. For group comparisons, group was entered into the design matrix as a categorical predictor; gender and age were entered as categorical and continuous (respectively) predictors of no interest (nuisance variable). For brain-behaviour correlations, behavioural score was additionally entered as a continuous predictor; the design was also repeated within each group separately, after eliminating group as predictor.

As there were no significant results at the whole-brain level, I selected regions of interest (ROIs) known to be involved in the mathematical brain network, based on a recent meta-analysis of studies in the field of numerical cognition (Arsalidou & Taylor, 2011). These included: the bilateral superior parietal lobule (SPL), the bilateral intraparietal sulcus (IPS), the bilateral inferior frontal gyrus (IFG). A sphere of radius 8 mm was built around the centres of each of these ROIs in order to create masks from which the average GMD was extracted using FSL's fslstats tool. All statistical analyses were subsequently performed in STATISTICA (StatSoft Inc.).

2.3.4.1.2 SBM

The technical details of the cortical reconstruction and volumetric segmentation procedures are described in prior publications (Dale et al., 1999; Fischl, Sereno, & Dale, 1999; Fischl, Sereno, Tootell, et al., 1999) and are summarised in the General Introduction's Methodology subheading (section 1.5.2.1.2, page 18). Here, based on SBM, cortical thickness, surface area and volume (the latter of which is product of the previous two measures) were calculated vertex-wise from the surface models. A final measure computed vertex-wise was local gyrification index (lGI), a measure of cortical folding calculated as the ratio of the total folded cortical surface over the total perimeter (Schaer et al., 2008). Higher lGI values indicate a higher degree of sulcal convolutions. Individual subject measures were smoothed using a full width half maximum (FWHM) kernel of 5 mm for lGI and 10 mm for thickness, area and volume. Clusters were corrected using a Monte Carlo null-z simulation, controlling the family-wise error rate at $p<.05$. The GLM designs used were similar to those used for VBM (see previous section).

2.3.4.2 White matter

2.3.4.2.1 Fractional anisotropy

Distortions in the DTI images due to subject motion or Eddy currents were corrected by applying affine alignment of each image to the non-diffusion image (b_0), using FMRIB's diffusion toolbox (FSL, version 5.0.5; www.fmrib.ox.ac.uk/fsl, (Smith et al., 2004)). The diffusion tensor elements were then estimated by solving the Stejskal and Tanner equation (Basser, Mattiello, & Lebihan, 1994), and eigenvalues λ_1, λ_2 and λ_3 were obtained by diagonalising the tensor matrix. FA (fractional anisotropy) was then defined as

$$\mathrm{FA} = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}$$

All eigenvalues and FA maps were calculated using FSL's DTIFit tool. After aligning each subject's FA map to standard space (MNI152), the tract-based spatial statistics (TBSS) method was carried out in FSL for all FA images, using the procedure described in (Smith et al., 2006). In short, a mean FA image and a "skeletonised" FA image is created from all subjects' aligned FA maps, and each subject's FA map is then projected onto the skeleton. Finally, voxel-wise statistics are made across subjects for all skeletonised FA images. As with the grey matter analyses, I included age and gender as covariates.

For the ROI analysis, the masks included the white matter regions that underlie the ROIs chosen for the grey matter analysis. The masks had to be defined as spheres of a larger radius (12 mm), in order for the sphere to overlap the FA skeleton. I extracted voxel intensities (which in this case represented FA values) from the skeletonised FA images.

2.3.4.2.2 Mean diffusivity

Mean diffusivity (MD) was also computed from the diffusion data. MD is defined as the average of the three eigenvalues:

$$\mathrm{MD} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$$

and represents a measure of the average diffusion in a voxel that is complimentary to that provided by FA. Increased MD is associated with decreased barriers to water diffusion in white matter, and its decrease has been associated with various diseases (e.g. obsessive–compulsive disorder, Lochner et al., 2012).

2.3.4.2.3 Tractography

In addition to computing voxel-wise FA values, I also used probabilistic tractography (Behrens, Johansen-Berg, et al., 2003; Behrens, Woolrich, et al., 2003), a technique for identifying fibre pathways from diffusion-weighted imaging, to reconstruct the anatomical networks linking the brain regions that are relevant to mathematics. Both exploratory and hypothesis-driven avenues were pursued in an attempt to further interrogate the diffusion imaging data.

In a study on children, Tsang (2009) used tractography to delineate the anterior portion of the superior longitudinal fasciculus (aSLF), a pathway connecting parietal and frontal cortex. They found that FA in the core of the left (but not the right) aSLF correlated with performance in a mental arithmetic task of approximate (but not exact) calculation, even after co-varying out age and scores on other more general cognitive tasks. Here, I first created a mask of the bilateral aSLF by identifying it in the JHU White-Matter Tractography Atlas and setting a population threshold of 20%; that meant that, in order for a voxel to be considered as part of the tract, it had to be identified as the aSLF in at least 20% of the individual brains that make up the probabilistic atlas. From this tract, I then kept the same seven coronal planes used in (Tsang et al., 2009), i.e. the middle of the aSLF, from Y=-21 to Y=-27. The average FA and MD were then extracted from this portion of the aSLF, and correlations between those averages and the aggregated measure of behaviour (see section 2.3.4.3) were computed.

The strength of homotopic connections (i.e. cross-hemispheric connections between anatomically equivalent regions) can be investigated with tractography, and varies considerably between cortical regions, as well as between individuals in the general population. For instance, tracer studies in the macaque demonstrate that the hand area of SMA has homotopic transcallosal connections that are more dense than those between primary motor hand areas (Rouiller et al., 1994). On the other hand, Putnam et al. (2009), attempting to delineate the complete cortical projection topography of the human splenium, found inter-individual differences in splenial connectivity, suggested by an inconsistent incidence (one third of the sample) of homotopic connections between the primary visual cortices, through the splenium.

Although to my knowledge there is no evidence suggesting differences in strength of homotopic connections as a function of expertise, I nevertheless attempted to compare my two groups in terms of the strength of homotopic connections between regions where significant effects were obtained in the VBM analyses. A seed was thus defined for the following bilateral regions: SFG, IPS, SPL, IFG, and M1 as a control

region. The seed was defined as a cube of edge length equal to 10 voxels, centred around the region's centre of gravity (see VBM results; or around atlas coordinates in the case of M1) and then smoothed down to a sphere of volume 8,000 mm^3. Probabilistic tractography (FSL's probtrackx) was then run from that seed, and the number of tracts reaching the termination mask of the contralateral equivalent was counted for each region. Consequently, for each region, the corresponding output image consisted of voxels whose intensity values represented the number of samples passing through that voxel and, therefore, the probability of a connection between the current voxel and the seed voxel.

2.3.4.3 Brain-behaviour correlations

Due to the extensive numbers of tests and ROIs, when looking at the brain-behaviour correlations I used Discriminant Function Analysis (DFA). DFA allowed me to create a single score for all the relevant behavioural measures, and similarly one for the relevant GMD and FA measures. DFA is a statistical technique that predicts[8] group membership based on a set of measures. These measures then act as predictors, with each one of them receiving a weight (similar to a multiple regression coefficient). Note that in contrast to logistic regression, which could not have been performed in the current study due to the relatively low number of subjects, DFA does not require a minimal number of subjects (Tabachnick, Fidell, & Osterlind, 2001). This approach is similar to previous studies that have chosen a single measure to integrate several behavioural measures based on sensitivity to between-group differences (Hyde et al., 2009; Steele, Bailey, Zatorre, & Penhune, 2013).

First, I used DFA to assess which set of measures[9] (both behavioural and brain-related) manage to significantly discriminate between the mathematicians group and the control group. I then took the standardised canonical coefficient as a weight for each measurement (i.e., for each behavioural test or for each ROI). These weights were then used to aggregate all relevant scores (z-scores) into a single score. This effectively reduced each set of measures down to a single aggregated measure: one for behaviour, one for GMD and one for FA, enabling me to correlate these DFA scores without running into a multiple comparisons problem. The aggregated behavioural score was used as a regressor in both VBM and SBM analyses seeking brain-behaviour correlations.

[8] NB: here I used DFA not to predict group membership but to obtain a meaningful single measure. In order to predict group membership, one would need validation on an independent group.

[9] I used the standardised (z-)scores of these measures, i.e. the sample mean (across both groups) was subtracted from each subject's score and the result was divided by the sample standard deviation.

2.4. Results

2.4.1 Behavioural results

I observed differences between the mathematicians and controls in numerical tasks, as well as non-numerical tasks. There was a tendency for mathematicians to perform better in most of the numerical tasks, in the verbal reasoning and in the logic tasks, as well as in some of the working memory tasks. This was not the case in the tasks relating to attention and social skills. The results are summarised in Table 4.

Table 4: Summary of behavioural results. Significant group differences are in bold. Non-parametric tests were used when statistical assumptions (e.g. distribution normality) were not met.

Cognitive category	Test	Controls μ±σ	Mathematicians μ±σ	Score range	Statistical test (when significant)
Intelligence	**IQ test (PIQ section)**	114.8±8.4	126.2±6.3	μ=100, σ=15	t(36)=4.71; p< 0.00005
	IQ test (VIQ section)	130.0±5.2	124.1±12.9	μ=100, σ=15	n.s.
Working memory	Digit span (forward)	11.2±2.2	11.1±2.7	0 to 16	n.s.
	Digit span (backward)	7.6±2.4	9.3±2.2	0 to 14	t(36)=2.26; p<0.05
	Letter span (forward)	9.2±1.5	10.3±1.3	0 to 16	t(36)=2.16; p<0.05
Attention	ANT:alerting	18.1±27.3	19.4±18.0	-∞ to ∞	n.s.
	ANT:orienting	61.0±34.7	58.1±26.6	-∞ to ∞	n.s.
	ANT:executive	100.7±34.8	98.9±35.0	-∞ to ∞	n.s.
Mental imagery	**Mental rotation task (MRT)**	10.4±4.3	14.8±4.2	0 to 24	t(36)=3.20; p<0.005
Numerical skills	Number acuity (*w*)	0.20±0.10	0.21±0.13	0 to ∞	
	Number line (positive numbers)	62.2±25.8	40.8±14.3	0 to ∞	Mann-Whitney Test: Z=2.86; p<0.005
	Number line (negative numbers)	59.7±29.7	54.0±23.7	0 to ∞	n.s.
	Numerical Stroop	0.09±0.08	0.10±0.04	-1 to 1	n.s.
	Numerical agility	2.3±2.6	6.5±0.9	0 to 10	Mann-Whitney Test: Z=4.16; p<0.00005

	Numerical strategies	19.4±7.6	31.7±5.2	0 to 40	t(36)=3.50; p<0.005
	Arithmetic task	13.2±2.1	15.4±2.0	0 to 22	t(36)=3.29; p<0.005
Logic	**Wason logic task**	0.1±0.3	0.6±0.5	0 or 1	Fisher's exact p=0.01, two-tailed
Verbal reasoning	**Verbal reasoning task**	5.5±1.4	5.9±2.7	0 to 10	F(1,35)=5.51, p<0.05
Social skills	Emotion recognition task	0.94±0.05	0.93±0.06	0 to 1	n.s.
	Gaze task	0.93±0.10	0.94±0.06	0 to 1	n.s.
	Face recognition task	48.5±5.0	47.9±2.8	0 to 54	n.s.
	Autism spectrum quotient (ASQ)	18.3±7.3	18.1±4.7	0 to 50	n.s.
Arithmetic strategies questionnaire	Visuo-spatial	5.21±3.31	7.67±2.74	1 to 10	t(24)=2.04, p=.053
	Inner verbalisation	3.29±2.46	2.08±1.31	1 to 10	n.s.
	Outer verbalisation	6.29±2.58	6.00±2.17	1 to 10	n.s.
	Kinesthetic	4.93±2.87	2.67±1.50	1 to 10	t(24)=2.46, p=.022

There were no significant effects of gender on any of these measures (all *ps*>.05), and since this factor was balanced between groups, I did not use gender as a covariate. On the other hand, since the proportion of native English speakers was different in the two groups (10 out of 19 mathematicians and 17 out of 19 controls; $\chi^2(1)=6.27$, p<0.05), I used this factor as a covariate when testing for group differences in VIQ and verbal reasoning; a main effect of group was only found for the latter. For the number line task, there were no significant main effects of group (mathematicians vs controls) or polarity (negative vs positive numbers), however there was a significant two-way interaction between the two factors (F(1,36)=7.17; p<0.05); this interaction was due to a better performance (i.e. more accurate mapping) for mathematicians with positive numbers (Mann-Whitney Test: Z=2.86, p<0.005, as per the above table), but not for negative numbers (Mann-Whitney Test: Z=0.53, p>0.1). For the numerical Stroop task, I observed a main effect of congruity (F(2,70)=52.66; p<0.00001) and distance (F(2,70)=4.32; p<0.05), however there was no main or interaction effect involving group. Cumulatively, the results indicate that the mathematicians exhibited superior performance on a wide range of cognitive tasks, not necessarily relating to numerical and mathematical skills with regard to controls.

Figure 6 shows the RTs corresponding to each of the 8 ratios used in the number acuity task. The β values across all subjects, obtained via linear regression, were 2194 ± 837 ms (μ ± σ). As can be seen in the figure, although the general trend is for RTs to increase as ratio increases, as one would expect given the reliability of the numerical ratio effect (Price et al., 2012), the trend is non-monotonous, with three data points (for ratios=0.8421, 0.8888 and 0.9411) taking exception to it.

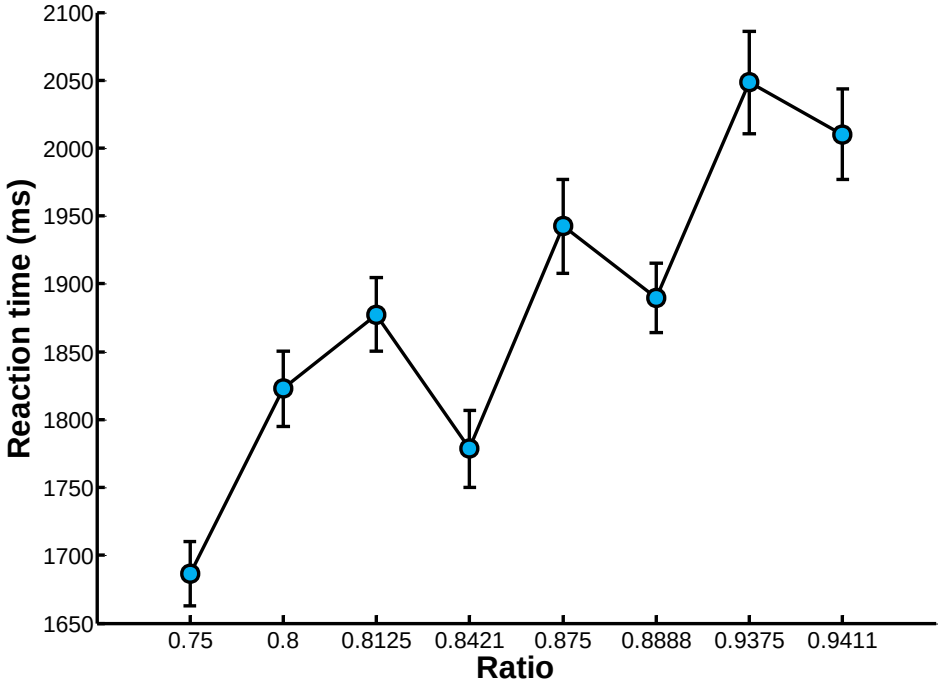


Figure 6: RTs for each ratio in the number acuity task (smaller to larger numerosity). Error bars show ±1 standard error of the mean of between-subject variability.

The two measures of ANS acuity, w and β, were significantly correlated ($r=-0.58$, $p<0.005$), as depicted in Figure 7. This means that subjects who manifest a pronounced numerical ratio effect, as indexed by a large β, have less noise in their internal number representation, as indexed by a small w. The average w is approximately 0.11 in the general adult population (Halberda & Feigenson, 2008).

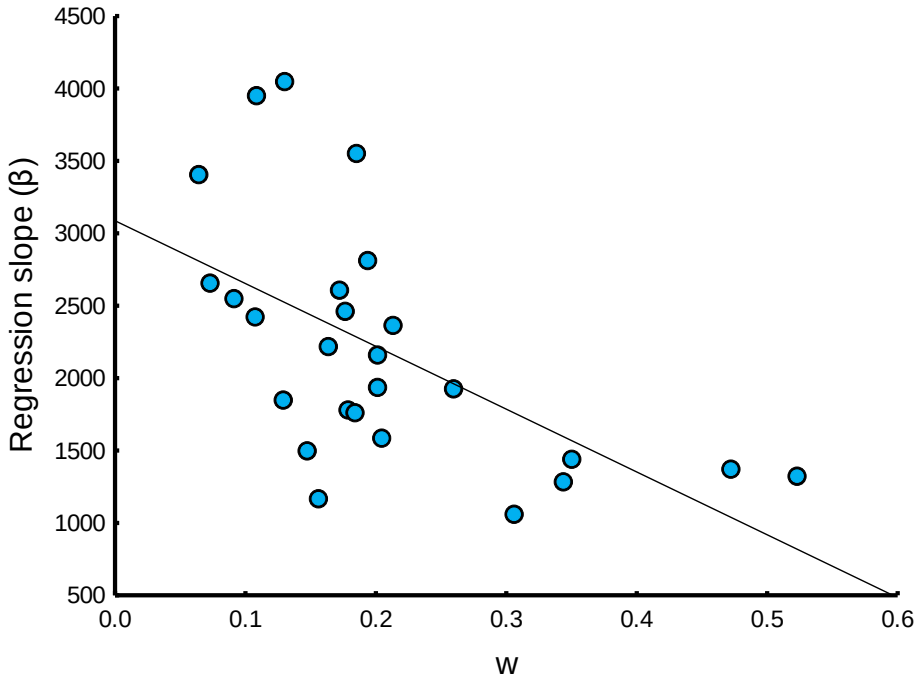


Figure 7: Correlation between β and w.

There was no group difference in either of the two measures (independent samples t-test: both ps>.4). When testing for the existence of correlations between the two ANS measures and measures of higher-level maths, specifically arithmetic and numerical strategies, it was found that, only for mathematicians, w was negatively correlated with arithmetic ($r(9)=-.65$, $p<.05$) and β was negatively correlated with numerical strategies ($r(9)=-0.72$, $p<0.05$), as per panels **a** and **b** respectively of Figure 8. The latter result is surprising, as it means that mathematicians who obtained high scores on the numerical strategies test had weaker numerical ratio effects.

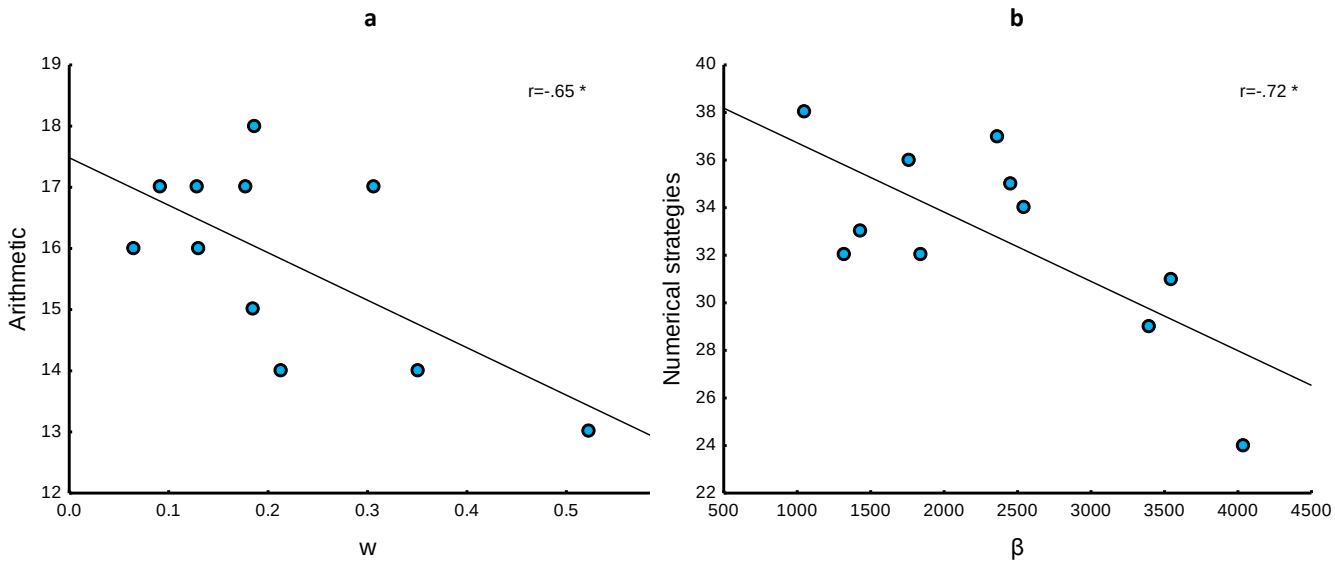


Figure 8: Correlations between ANS measures and measures of higher-level maths. Both correlations are significant (*: p<.05) and only mathematicians are included.

2.4.2 Brain results

2.4.2.1 Group differences

2.4.2.1.1 VBM

When contrasting the groups to assess GMD differences at the whole-brain level, no clusters survived the significance level (p<0.05, corrected for multiple comparisons).

For the ROI analysis, I tested for group differences in fronto-parietal areas known to be involved in number processing and abstract mathematical reasoning (R. Cohen Kadosh, Lammertyn, et al., 2008; Dehaene & Cohen, 1995; Dehaene et al., 2003; Pesenti et al., 2001). GMD values were submitted to a 3-way mixed-design ANOVA with group (mathematicians vs controls) as between-subjects factor and hemisphere (left vs right) and region (SPL vs IPS vs IFG) as within-subject factors. The analysis revealed a main effect of region ($F(2,72)=23.47$, $p<0.0001$), as well as a region×group interaction ($F(2,72)=4.28$, $p<0.05$) that was qualified by a region×hemisphere×group interaction ($F(2,72)=5.59$, $p<0.01$). I decomposed the 3-way interaction for each hemisphere by performing separate region×group ANOVAs for each hemisphere. For the right hemisphere, there was a main effect of region ($F(2,72)=11.60$, $p<0.00005$) and a group×region interaction ($F(2,72)=7.56$, $p<0.005$); the interaction reflected higher GMD for mathematicians in the rSPL ($F(1,36)=4.80$, $p<0.05$) but lower GMD for mathematicians in the right IPS ($F(1,36)=6.44$, $p<0.05$). For the left hemisphere, I only found a main effect of region ($F(2,72)=18.86$, $p<0.00001$) and a trend towards a main effect of group ($F(1,36)=2.80$, $p=0.1$); further exploratory analysis revealed, for the left IFG, a main effect of

group ($F(1,36)=4.35$, $p<0.05$, uncorrected), with controls having higher GMD than mathematicians. The coordinates of the regions that showed a group difference are given in Table 5, and the group differences are plotted in Figure 10. The significance of these effects did not change when the analyses were repeated with PIQ as a covariate.

Table 5: Coordinates of ROIs that showed a significant group difference in GMD.

Region	Coordinates in voxels (centre of gravity)		
	x	y	z
rIPS	40	-40	43
rSPL	30	-62	44
lIFG	-42	4	30

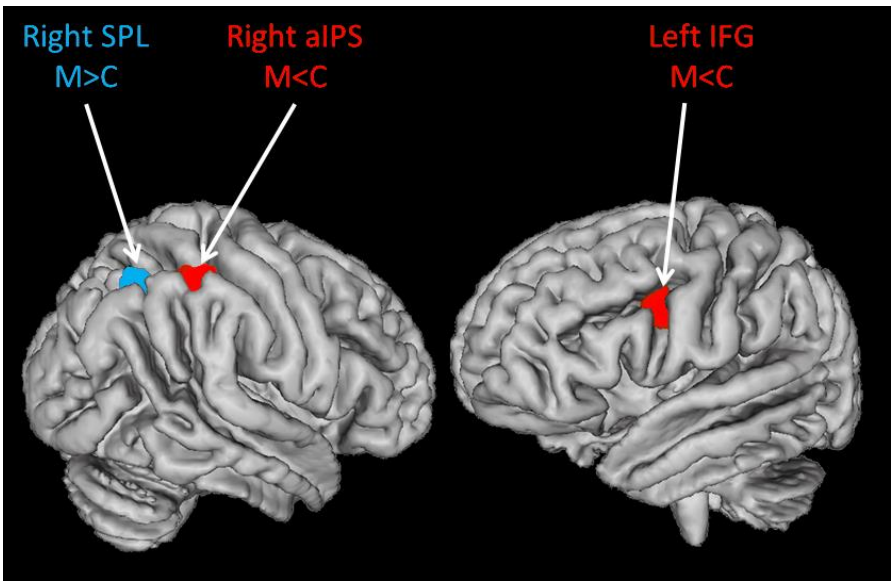


Figure 9: The ROIs that showed a significant group difference in GMD. Increased GMD in mathematicians is shown in blue, while increases in controls are shown in red.

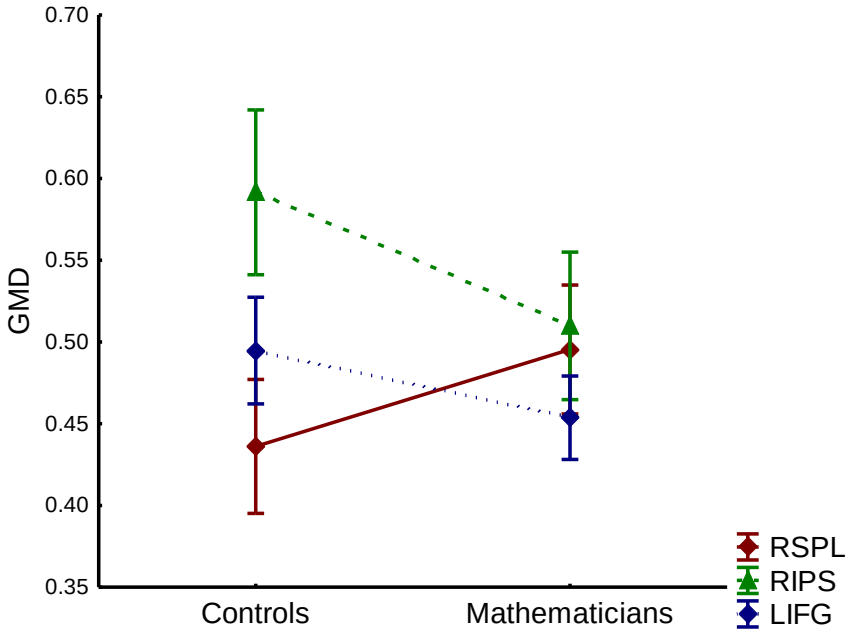


Figure 10: Mean extracted GMD from each of the three ROIs where groups were significantly different.

2.4.2.1.2 SBM

When comparing groups at the whole-brain level, decreased lGI was found in mathematicians relative to controls in a 3,568 mm^2 cluster in the right postcentral gyrus (maximum at Talairach coordinates x=30.0, y=-31.5, z=68.8) extending further into the parietal cortex, depicted in Figure 11a. Also, decreased surface area and volume was found in a 1,546 mm^2 (946 mm^2 respectively) cluster centred in the left fusiform gyrus (maxima at Talairach coordinates x=-42.5, y=-64.6, z=-18.8); the surface area cluster is depicted in Figure 11b.

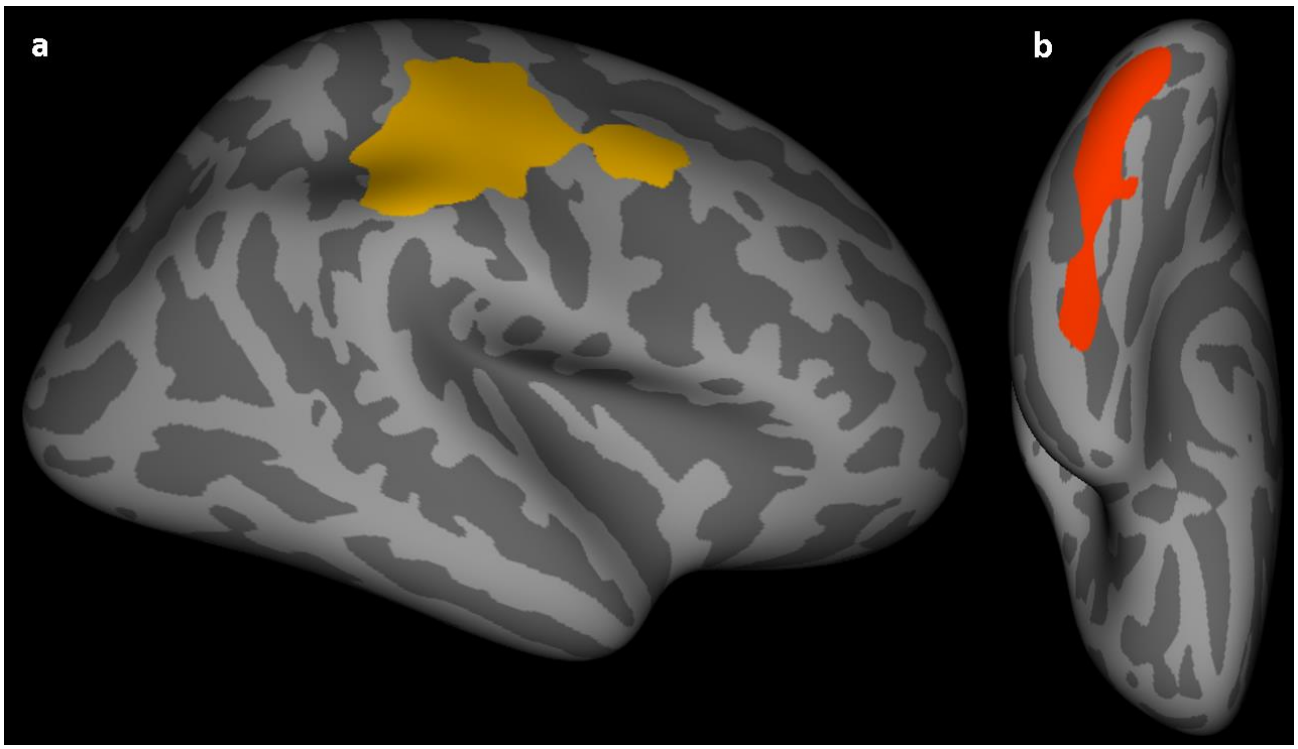


Figure 11: (a) Inflated surface map (dark grey = sulci; light grey = gyri) showing right postcentral gyrus cluster with decreased lGI in mathematicians (lateral view). (b) Left fusiform gyrus cluster with decreased area in mathematicians (ventral view).

2.4.2.1.3 White matter

TBSS performed at the whole-brain level did not identify any significant voxels that differed in FA across groups ($p<0.05$, corrected). There were also no significant group differences in FA for any of the ROIs.

2.4.2.2 Brain-behaviour correlations

2.4.2.2.1 Aggregated (DFA) scores for behaviour, GMD and FA

In order to clarify the relationship between behavioural and brain (ROI) measures, I used DFA to assess which set of measures manage to best discriminate between the groups, for each measurement category (behaviour, GMD, FA). On the basis of the DFA results, I computed an aggregate score for each category, in which individual measures were weighted on the basis of the standardised canonical coefficient observed for the respective measure; this yielded an aggregated behavioural score, an aggregated GMD score and an aggregated FA score (see the Methods section for more details). For behaviour, the only tests that significantly differentiated between the groups ($F(2,35)=26.84$, $p<0.0001$) were the Numerical Strategies and the Numerical Agility tests. For GMD, all three regions that showed a group difference (rIPS, rSPL and lIFG) were also significant predictors in the analysis ($F(3,34)=6.74$, $p<0.005$). For FA, the left IFG and right IPS were the only significantly predicting ROIs ($F(2,35)=3.44$, $p<0.05$).

Table 6: Summary of DFA predictors. *indicates significance at p<.05.

Aggregated score	Predictor [group comparison]	Significance of Wilks' Lambda	Weight
Behaviour	Numerical Strategies [M>C *]	p < 0.05	-0.49
	Numerical Agility [M>C *]	p < 0.005	-0.69
GMD	rSPL [M>C *]	p < 0.01	0.75
	rIPS [C>M *]	p < 0.05	-0.65
	lIFG [C>M *]	p < 0.05	-0.63
FA	rIPS [C>M]	p < 0.05	1.05
	lIFG [M>C]	p < 0.05	-1.17

Taking into account the weight sign and the group comparison associated with each predictor (summarised in Table 6), we can infer that a *lower* (more negative) aggregated behavioural score would correspond to *better* performance; in order to make this score more intuitively interpretable, it was subsequently inverted (multiplied by -1) thus making *greater* (positive) values correspond to *better* performance.

On the other hand, a *higher* aggregated GMD score meant *greater* GMD values in RSPL and *lower* GMD values in RIPS and LIFG; and a *higher* aggregated FA score meant *greater* FA values in RIPS and *lower* FA values in LIFG. A limitation of the analysis is that it is difficult to interpret the aggregated scores for GMD and for FA in absolute terms, i.e. in terms of there being higher or lower GMD and FA in the whole of the *network* of interest defined by the respective regions; this is due to the effects in those regions having opposite directions. There was a significant correlation between the aggregated behavioural and the aggregated GM scores ($r=0.41$, $p<0.05$), and also with the aggregated FA scores ($r=-0.36$, $p<0.05$), both depicted in Figure 12.

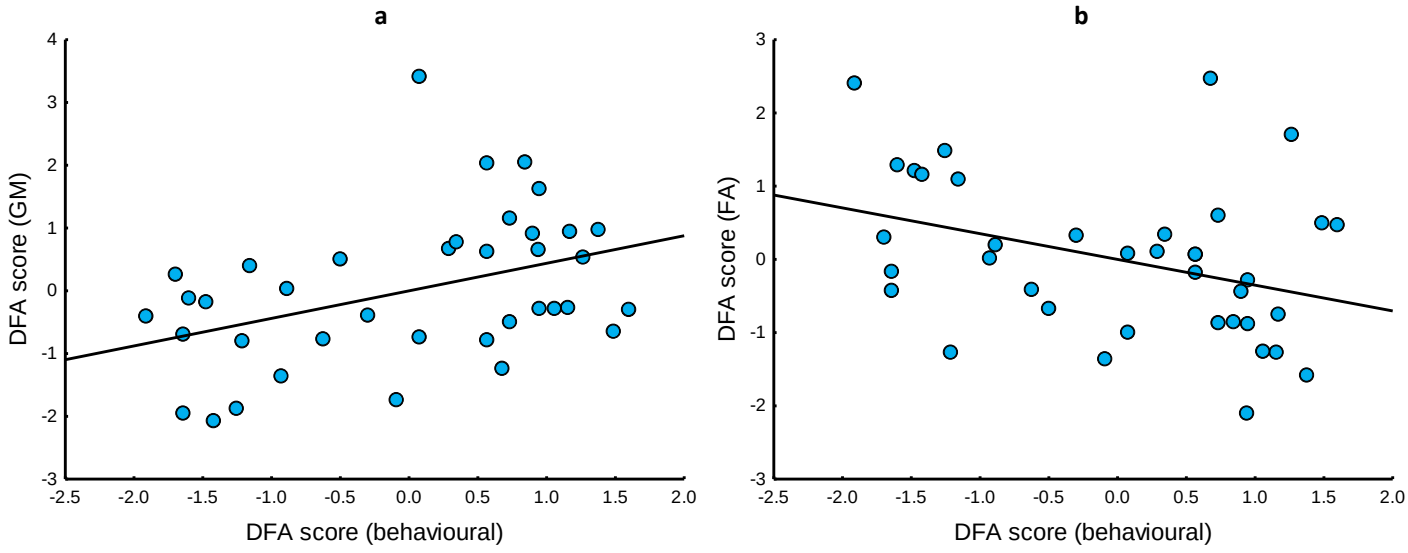


Figure 12: Correlation between the aggregated (DFA) measures of behavioural performance on the X axis, and on the Y axis (a) GMD and (b) FA.

2.4.2.2.2 VBM

The VBM whole-brain analysis was repeated with the aggregated behavioural score as an independent variable. GMD in the *left* superior frontal gyrus (SFG; x=-57, y=58, z=70), showed a double dissociation as it was *negatively* correlated with the aggregated behavioural score in controls (r=-.51, p<0.05), but *positively* correlated with this score in mathematicians (r=.65, p<0.005), as depicted in Figure 13. GMD in the *right* SFG *positively* correlated with behaviour only for mathematicians (r=0.53, p<0.05), but was not significantly correlated for controls (r=0.30, p>0.1).

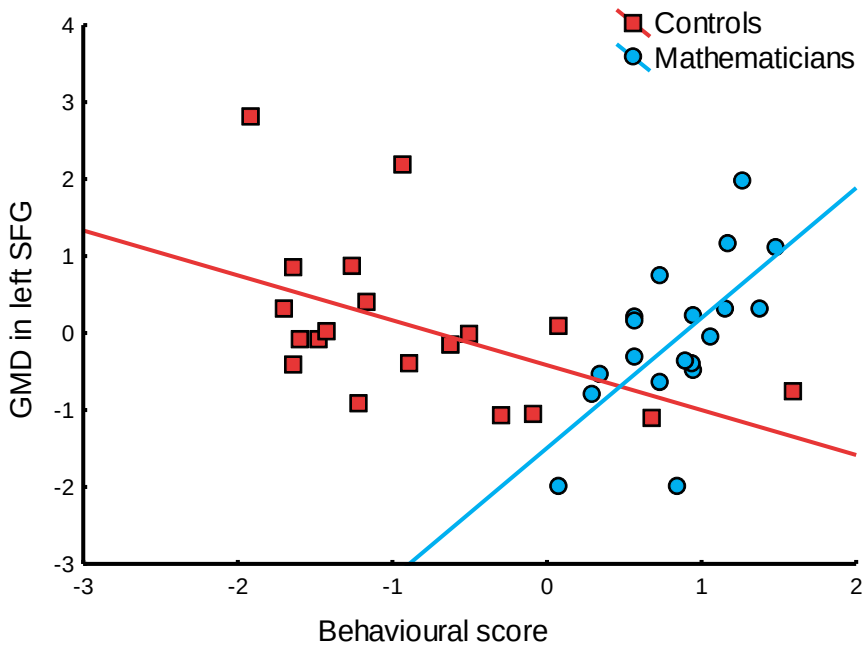


Figure 13: Correlation between GMD in the left SFG and behavioural score, showing a double dissociation.

An additional analysis was performed, whereby correlations were computed between answers to the subjective ratings regarding frequency of use of various strategies in solving arithmetic and GMD in the regions of interest described previously, as well as those for which effects were obtained in the whole-brain correlational analysis (section 2.4.2.2.2, page 45). Significant correlations were found, for controls, between ratings for kinesthetic and GMD in lIFG ($r(12)=-.56$, $p<.05$) and between verbal and GMD in lSFG ($r(12)=.62$, $p<.05$), and for mathematicians between visuo-spatial and rIPS ($r(10)=-.70$, $p<.05$). There was, however, no multiple comparisons correction applied for these correlations.

2.4.2.2.3 SBM

When using the behavioural score as a continuous predictor, group as categorical predictor, and age and gender as nuisance variables, there were no clusters where the correlation between behaviour and either thickness, surface area, volume or lGI differed between mathematicians and controls. When searching within each group separately, for controls only I found a 2,549 mm^2 cluster in the left lateral occipital area (max Talairach coordinates x=-19.9, y=-88.0, z=-7.4) and a 1,698 mm^2 cluster in the left supramarginal gyrus (max Talairach coordinates x=-57.6, y=-26.3, z=27.4), shown in Figure 14a, where lGI correlated negatively with behavioural score.

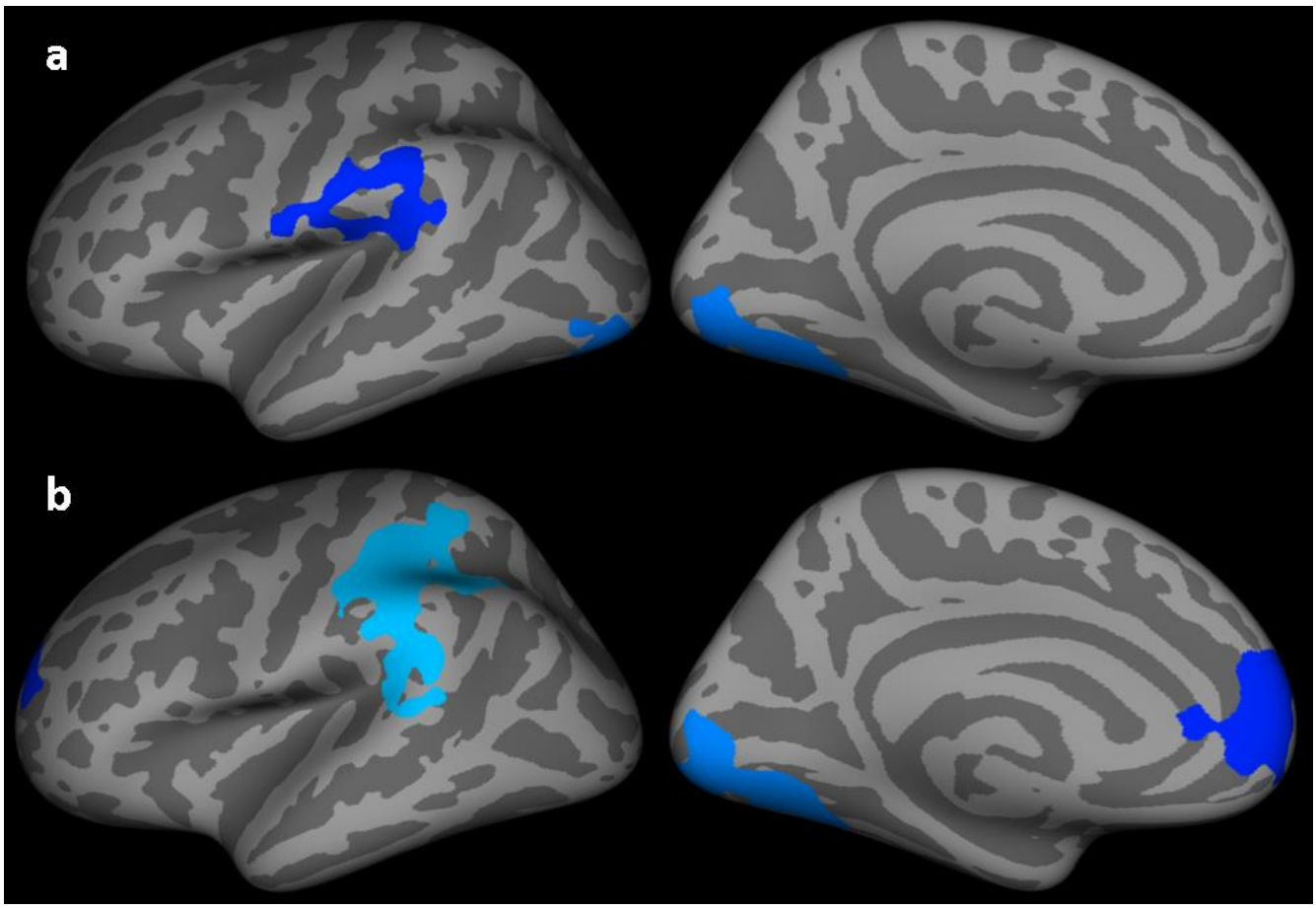


Figure 14: Lateral (left panel) and medial (right panel) views of the left hemisphere of the brain, showing clusters where lGI correlated negatively with behavioural score. Clusters are in (a) the lateral occipital and the supramarginal gyrus, and, when excluding an outlying control subject, (b) an extra cluster in the superior frontal gyrus.

Since one of the control subjects could be considered an outlier in terms of his behavioural score (which was more than 2.5 SDs away from the mean of his group), I also repeated this analysis without him; this changed the previous clusters to 2617 mm^2 (x=-13.4, y=-91.2, z=3.8) and 3705 mm^2 (x=-47.2, y=-27.8, z=52.9) respectively and also added an extra 1665 mm^2 cluster in the left superior frontal gyrus (x=-12.5, y=53.1, z=3.1) – see Figure 14b.

2.4.2.2.4 Tractography

The mean FA and the mean MD in the selected portion of the aSLF, both on the left and on the right, did not correlate significantly with the behavioural DFA measure (all ps>0.6).

Due to the significant link between GMD in the SFG and behaviour, I further assessed whether the tract connecting rSFG to lSFG was stronger in mathematicians than in controls. This was the case ($p<.05$, independent samples t-test). Since Levine's test in this case was significant, an unequal variances t-test was

also used to compare the groups – the effect in this case only maintained marginal significance (p=.06). This group difference between tract strengths was non-significant in the case of the rM1 to lM1 control tract.

2.5. Discussion

In order to understand the neurocognitive bases of mathematical ability, I compared mathematicians and non-mathematician controls on a battery of cognitive tasks and with regard to brain structure. I found group differences in both these measures, as well as in the relationship between them. These results might have important implications for understanding the neurocognitive mechanisms supporting mathematical ability.

2.5.1 Behavioural results

Mathematicians were found to display superior performance in a wide range of (but not all) tasks. They performed better on all higher-level mathematical tasks (e.g., numerical agility and numerical strategies), but only on one of the lower-level tasks (mental number line for positive numbers). The former tasks build on skills acquired through education, whereas the latter type relates to number sense, which is arguably more innate (Dehaene, 2011) – although the *linearity* of the number line has been suggested to also depend upon formal education (Dehaene et al., 2008). Surprisingly, mathematicians also displayed superior verbal reasoning; although performance in this task relies mostly on linguistic skills and inference ability, it is also an exercise of reasoning in a more general sense (Hitch & Baddeley, 1976), and thus may reflect superior reasoning among mathematicians, which is consistent with their superior performance on the Wason logic task. There were no group differences in any of the social skills measures, including the ASQ, which tentatively goes against previous evidence, both anecdotal and scientific (Baron-Cohen, Wheelwright, Burtenshaw, & Hobson, 2007; Baron-Cohen et al., 2001), of mathematicians – indeed scientists in general – scoring higher on autism-related traits.

Mathematicians had superior performance in the mental rotation task. Despite anecdotal reports of mathematicians' and arithmetical prodigies' increased reliance on visuo-spatial strategies (Hadamard, 1949; Scripture, 1891), and despite the robust relationship between number and space, rooted in the brain's organisation for these capacities (Hubbard, Piazza, Pinel, & Dehaene, 2005), there is limited reason to expect a strong relation between (2- or 3-dimensional) mental rotation abilities in particular, and the sort of (1-

dimensional) spatial abilities involved in number representation and arithmetic. Indeed, there have been conflicting findings about this relation. For example, a meta-analysis by Friedman (1995) suggested that verbal tasks predicted arithmetic much more than spatial tasks did; and Delgado & Prieto (2004) found that mental rotation predicted performance in algebra and geometry but not in arithmetic. On the other hand, Cheng & Mix (2014) did find that training 6- and 8-year-old children in mental rotation improved their mathematical performance, while Reuhkala (2001) found that these two scores correlated in 15- and 16-year-old adolescents. The present result may thus suggest a greater relationship between spatial abilities (mental rotation in particular) and number representation than often thought.

In the arithmetic strategies questionnaires, controls gave higher ratings to kinaesthetic strategies (finger counting), while the mathematicians rated the visuo-spatial modality higher, confirming older anecdotal reports of mathematicians' increased relative reliance on visuo-spatial – as opposed to verbal – thinking (Hadamard, 1949). Additionally, mathematicians' ratings of use of visuo-spatial strategies correlated negatively with GMD in rIPS, a result that is congruent with the group difference (M<C) that existed in terms of mean GMD in this region. More surprisingly, the controls' ratings of use of kinesthetic strategies correlated - also negatively - with GMD in the lIFG (where, on average, the same M<C group difference was observed). Although possibly spurious (as multiple comparisons correction was not applied), these correlations suggest that there might exist structural correlates of preferred modality, and that those might interact with expertise (group membership).

Although ratings of verbal strategies did not differ between groups, this result suggests that, for mathematicians, the visuo-spatial form might represent an objectification of numbers. This supports and gives a directional interpretation to the SBM fusiform surface area difference, as it seems, in this case at least, that a *higher* reliance on visuo-spatial number form is associated with *decreased* volume and area of this region, which underlies the visual number form in the triple-code model (Dehaene et al., 2003). The visuo-spatial modality preference also complements the VBM finding of increased GMD in RSPL in mathematicians – a region that, in the triple-code model, plays a similar role to the fusiform gyrus, and that is also involved, in addition to numerical operations (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; K.-M. Lee, 2000), also in a variety of visuo-spatial tasks such as hand reaching, grasping and eye movements (Simon, Mangin, Cohen, Le Bihan, & Dehaene, 2002).

Although the two measures of ANS acuity obtained from the number acuity task, w and β, were correlated, demonstrating yet again the good convergent validity existent between the different ANS metrics (Price et al., 2012), none of them reliably predicted performance in the higher-level mathematical tasks, namely the arithmetic and numerical strategies tasks. For the mathematicians, w was found to negatively correlate with arithmetic and β was negatively correlated with numerical strategies, however these two correlations are essentially pointing in different directions, as an accurate ANS is indexed by a *low* w but by a *high* β; this apparent artefact in β's correlation might be due to the lack of a clear-cut number ratio effect, as suggested by the crooked line depicting the relationship between RTs and ratio, in Figure 6. On the other hand, the lack of correlation between low- and high-level numerical skills in adults is consistent with previous results that have only found such a link in children, but not also in the healthy adult population (Inglis et al., 2011; Libertus et al., 2011; Mazzocco et al., 2011b), and further supports the idea that ANS acuity and arithmetic ability may change over the course of the development.

2.5.2 Brain results

2.5.2.1 Grey matter

2.5.2.1.1 VBM

Several previous reports link expertise in various domains to *increased* GMD in relevant regions (Aydin et al., 2007; Maguire et al., 2000; Scholz et al., 2009; Sluming et al., 2002). There are, however, also reports of expertise being associated with changes in GMD in both directions; for instance, James et al. (2013) found *increased* GMD in musicians vs non-musicians in cognitive areas such as left IFG and left IPS, but *decreased* GMD in sensorimotor areas such as the striatum. Similar dissociations across regions of the brain have been found for on-line gamers (Han, Lyoo, & Renshaw, 2012) as well as in conditions such as autism (Abell et al., 1999) and bipolar disorder (Wilke, Kowatch, DelBello, Mills, & Holland, 2004). Notably, such a structural dissociation is not found in dyscalculia, where *lower* GMD is reliably found in the right IPS as compared to healthy controls (Cappelletti & Price, 2014; Molko et al., 2003; Rotzer et al., 2008; Rykhlevskaia et al., 2009). Here, I found such a decrease in right IPS GMD for mathematicians vs non-mathematicians. This suggests that, at least in the case of right IPS, local GMD cannot reliably be taken as a linear predictor of mathematical ability. In fact, if we were to simplistically consider people with dyscalculia and mathematicians at opposite ends of a single dimension, this would imply an inverted-U shape relationship between mathematical expertise and GMD in right IPS.

Mathematicians also had decreased GMD in the left IFG. Increased GMD in this region was previously observed in musicians versus non-musicians (James et al., 2013; Sluming et al., 2002). While the speech relevance of this region is not obviously linked to either mathematical or musical expertise, its number relevance is well established; for instance, it is involved in calculation (Arsalidou & Taylor, 2011; Menon et al., 2000) and is activated for untrained vs trained items (Delazer et al., 2003), i.e. in conditions of higher working memory demands. It is not implausible (although difficult for any study to prove) that the direction of change in GMD with experience is domain-dependent, and that it might be positive for some domains (e.g. music) but negative for others (e.g. mathematics), at least in the case of left IFG.

In contrast to Aydin et al. (2007), who found increased cortical GMD in mathematicians in the left inferior frontal and in the bilateral inferior parietal lobules, the only region where I observed an increase in mathematicians was the rSPL. The double dissociation found in two closely-located parietal regions (increased GMD for mathematicians in the rSPL, but decreased GMD in the right IPS) is important because these regions have different but complimentary roles in number processing. As suggested by single-cell recordings in monkeys and neuroimaging studies with humans, the IPS is number-*selective*, in that neurons in this region fire during the presentation of a particular numerosity and thus they seem to represent cardinal values or specific places on the 'mental number line' (*place coding*)(Nieder, Freedman, & Miller, 2002; Nieder & Miller, 2004; Piazza et al., 2004); the SPL, on the other hand, is number-*sensitive* (i.e., more objects presented lead to more neural activity) and supports *summation coding* of non-symbolic numerosities such as dots (Roitman, Brannon, & Platt, 2007; Santens, Roggeman, Fias, & Verguts, 2010), and as indicated in the previous section might be linked to greater reliance on visuo-spatial analysis (Dehaene & Changeux, 1993; Verguts & Fias, 2004). The observed effect may therefore reflect the level of mathematical abstraction among mathematicians, and may support the view of more efficient neural organisation in mathematicians.

Overall, the VBM results suggest that, at least in the domain of mathematical expertise, the relationship between expertise and GM metrics in relevant parts of the brain is non-linear. It is possible that this relationship becomes clearer upon breaking down the GMD metric into its main possible macroscopic sources of variation – thickness, surface area and gyrification; it is this that I attempted to do in the SBM analysis.

2.5.2.1.2 SBM

Surface area in the left fusiform gyrus was reduced in mathematicians with respect to controls. Although not part of the "traditional" fronto-parietal network of "number areas", this area might play an indirect but important role in exceptional mathematical talent. For instance, stronger activation was observed here during complex calculation in a calculation prodigy than in controls (Fehr et al., 2010); the authors speculated that, for the prodigy, this area showed neural reorganisation (e.g. superior calculation skills at the cost of decreased social skills). A similar account might explain a recent multivoxel pattern analysis (MVPA) result in which the fusiform gyrus was the region most strongly differentiating a group of autistic children from a control group in terms of arithmetic problem complexity; the groups were IQ-matched, however the autistic group had superior numerical problem solving abilities (Iuculano et al., 2014). The authors proposed that this region might specialise early in development, and that it subsequently responds preferentially to classes of objects for which an individual has perceptual expertise (in that case: numbers). However, given the cross-sectional design of the study, one is not able to conclude whether these structural changes are due to early developmental changes, or due to later changes as a function of education.

A group difference in lGI (decrease for mathematicians) was observed, but only in the right postcentral gyrus, a part of the somatosensory cortex. The sensory areas are among the thinnest in the cortex (Fischl & Dale, 2000), and so a floor effect might explain why a thickness group difference did not accompany the lGI one. To my knowledge, the morphology of the right postcentral gyrus has not been linked to experience. In addition, for controls but not for mathematicians, behavioural score correlated negatively with lGI in a left lateral occipital cluster and in a left supramarginal cluster close to the postcentral gyrus. This suggests that the best performing controls, like mathematicians, had decreased folding in this postcentral region.

Traditionally, lGI has been thought of as a feature that remains relatively stable from the first months of life till old age; recently, it has been shown to decrease with age (Hogstrom, Westlye, Walhovd, & Fjell, 2012), although there is less evidence of it being as prone to experience-related changes as compared to, for instance, cortical thickness. Gyrification is altered in conditions such as schizophrenia (Palaniyappan, Mallikarjun, Joseph, White, & Liddle, 2011) and Williams' syndrome (Gaser et al., 2006), but might also play a role in cognition and experience-dependent plasticity: for instance, differences in cortical folding predict executive processes (Fornito et al., 2004) and discriminate between meditators and non-meditators (Luders et al., 2012).

2.5.2.1.3 Divergence between VBM and SBM

The GMD effects found with VBM would have their most intuitive counterpart in the SBM-produced cortical volume (obtained vertex-wise as the product between surface area and thickness), and it would be reasonable to expect GMD to be proportional to both surface area and cortical thickness individually. SBM's explicit model of the cortical surfaces offers independent measures of these three metrics and thus may offer a more precise interpretation of the VBM-derived density change. However, VBM is sensitive to a combination of changes in grey matter thickness, area and folding, and these relative contributions are confounded, so the VBM-SBM correspondences do not always corroborate.

In the present case, the only such correspondence occurred when excluding a possible outlier in the SBM analysis; in that case, and for controls only, an additional region's lGI correlated negatively with behaviour: the left SFG, an area mainly involved in working memory (Boisgueheneuc et al., 2006) but also in mental calculation (Arsalidou & Taylor, 2011). The same negative correlation with behaviour was obtained in controls with VBM, which suggests that (i) anatomically, denser GM in this area might be due to increased folding, but that (ii) this is associated with poorer behavioural performance. The correlation had the opposite sign for mathematicians in VBM and was not significant in SBM, which may mean (i) a different physiological basis for individual differences in GMD in that area, and (ii) an inversed functional relationship in terms of behaviour (i.e. *increased* GMD predicts *better* performance).

There are relatively few studies that have applied both approaches on the same data set. Using VBM, Voets et al. (2008) found regions of GMD reduction in schizophrenia; in the SBM approach, only some of these, however, corresponded to thinned cortex, whereas the remaining ones corresponded to decreased surface area. VBM density changes in schizophrenia not supported by cortical thinning were thus attributable to altered surface area in these regions, prompting the authors to suggest that a neurodevelopmental basis to the disease exists, rather than alterations occurring later, as would have been suggested by a predominantly thickness-explained GMD reduction. On the other hand, several studies have found consistent convergent effects between GMD and thickness, when comparing groups of musicians with non-musicians (Bailey, Zatorre, & Penhune, 2013; Bermudez et al., 2009; Foster & Zatorre, 2010b), which enables a more experience-dependent explanation of the VBM finding as opposed to an early developmental one.

2.5.2.2 White matter

My lack of finding of a group difference with regard to FA is somewhat puzzling, given that previous research has found that performance in a variety of domains correlates *positively* with FA in relevant brain areas (Bengtsson et al., 2005; M. X. Cohen, Schoene-Bake, Elger, & Weber, 2009; Della-Maggiore, Scholz, Johansen-Berg, & Paus, 2009; Flöel, de Vries, Scholz, Breitenstein, & Johansen-Berg, 2009; Fuentemilla et al., 2009; Kim & Whalen, 2009). Less frequently, expertise has been found to *negatively* correlate with FA (Imfeld, Oechslin, Meyer, Loenneker, & Jancke, 2009; Schmithorst & Wilke, 2002). Decreased FA is, however, more reliably associated with disease; for instance low FA in the corpus callosum has been linked with schizophrenia and multiple sclerosis (Foong et al., 2000; Roosendaal et al., 2009). Performance in these behaviours can be attributable to either inter-individual differences (i.e. genetic factors) or expertise (i.e. environmental factors), and studies that have found a correlation between effect of magnitude and number of years of experience suggest that the differences in those cases might be environmental rather than innate (see review by Johansen-Berg, 2010); on the other hand, it was found (Chiang et al., 2009) that the majority of the variance in FA in large portions of the brain's white matter, particularly in parietal and frontal lobes, is explained by genetic rather than environmental factors.

The fact that the tract connecting rSFG to lSFG was stronger in mathematicians than in controls adds to the evidence in support of homotopic tracts displaying a high degree of variability between regions and between individuals (Putnam et al., 2009; Rouiller et al., 1994). However, the case for expertise potentially being a factor in this variability is not strong: although the homotopic control tract connecting the primary motor cortices did not show a similar group difference, the same can be said of other homotopic tracts between regions that *are* relevant to mathematical expertise, and which were also highlighted in the present brain results – regions such as the IPS or the SPL. Also, the result lost significance once (the appropriate) non-parametric statistics were used. In the case of the SFG, however, the effect directions of both GM and WM results seem to agree for mathematicians: for the former, GMD was *positively* correlated with behaviour for both left and right SFG, while for the latter, the rSFG-lSFG tract was *stronger* in mathematicians than in controls. Convergent directions of change between cortical metrics and integrity of underlying WM was previously found, both as a result of disease (Rykhlevskaia et al., 2009) and of experience (Scholz et al., 2009).

The fact that, unlike in the Tsang et al. study (2009), the mean FA in the aSLF did not correlate with behavioural performance, might be related to subjects here being adults rather than children as in the case

of Tsang et al., and also could well be due to the different measures of arithmetical performance used in each case. Tsang et al.'s own data, for example, shows no significant correlation in the aSLF between FA values and *written* math, only with *mental* arithmetic.

2.5.3 Limitations & conclusions

In summary, this study found cognitive and structural brain differences between mathematicians and non-mathematicians; conclusions regarding the correlation between these measures are limited by the interpretation of the aggregated scores for GMD and FA. An important difference between this and Aydin et al.'s (2007) study is that they attributed the observed differences in brain structure between mathematicians and controls to the extensive and specific training that the former group benefited from. The wider age range in their study enabled the observation of a positive correlation between GMD and number of years spent as a mathematician, which prompted Aydin et al. to suggest neuroplasticity as a mechanism that justifies the observed brain structure differences; in turn, MRI-detectable neuroplasticity could come about via a number of cellular mechanisms such as neurogenesis, gliogenesis or synaptogenesis.

With the exception of the voxel-wise statistics, multiple comparisons corrections (such as the Bonferroni) were not applied when brain-behaviour correlations were computed. Given the high number of structural measures that were defined across all methods used, applying these corrections would have caused most effects to lose significance. This shortcoming limits the amount of trust one can place upon the brain-behaviour correlations.

Although the cross-sectional nature of this study prevents causal inferences, it can be speculated that a bidirectional feedback loop exists between experience and structural brain changes, and that the nature of this loop changes across the lifespan; for instance, early (congenital) differences in the surface area of the fusiform gyrus might enable a predisposition for mathematics, which later – with acquired experience –might affect GMD. This type of study presents a clear difficulty in untangling cause and effect, and it is only longitudinal studies that will be able to untangle the relative contributions of experience and predisposition in creating the macroscopic changes I observed here cross-sectionally. Also, the direction of the observed increases and decreases of cortical metrics can only be understood in relative terms with regard to my two random samples of mathematicians and non-mathematicians. There is no absolute reference that can be taken for these changes, and one cannot know if those areas represent, for instance, areas of increased

plasticity induced by studying humanities rather than decreased plasticity induced by studying mathematics. Although the direction of changes across all measures is far from being in unison, a clear conclusion is that "more is not always better".

Finally, in interpreting the results of this study, it is important to address the question of specifically what type of expertise the observed effects are linked to. Undoubtedly, mathematics is a vast field, and so there are many types of "mathematicians". In order to investigate the brain-behaviour link associated with expertise in maths, perhaps the most unequivocal expertise group would be one formed of experts (prodigies) in mental calculation, arguably maths' purest form. However, performing group analyses of the sort presented here would require a suitably large sample of such rare experts, which would have been impractical to recruit in this instance. Although arithmetic itself is also vast in terms of the variety of components and supporting functions it relies on, the breadth of the term "mathematical expertise" would have been greatly reduced, and consequently the interpretation of the effects found would have been less equivocal. Although, due to their educational background, mathematicians such as the ones recruited in this study are likely to be well above average in arithmetic, they are not truly arithmetic experts; it is probably fairer to assume that the mathematicians here excel in logic, as the results of the Wason task have indeed shown. This caveat must be borne in mind when interpreting the present effects as being linked with "mathematical expertise". Also, any of the findings reported here could ultimately be viewed from the vantage point of either of the two groups of expertise, i.e. the effects might reflect exceptional characteristics of the humanities academics (controls) just as well as they can reflect those of the mathematicians. Such an argument might be suggested, for instance, by the controls' strong discrepancy between Verbal- and Performance-IQ scores.

2.6. Acknowledgements

I thank Darko Odic for help with the Weber fraction analysis; Marie Schaer for invaluable guidance with my surface-based analyses in FreeSurfer; Rogier Mars for crucial help and insight with the DTI analyses; Adam Thomas for help with the VBM analyses; and Dick Passingham for valuable overviews into some of my results.

3 Study 2: Enhancing arithmetic learning using tRNS

3.1. Abstract

Proficiency in arithmetic learning can be achieved by using a multitude of strategies, the most salient of which are rote learning (making use of long-term memory) and procedural learning (applying a certain set of computations). Here I investigated the effect of transcranial random noise stimulation (tRNS) on both types of learning. tRNS (as 'real' and 'sham' stimulation) was applied to the bilateral prefrontal cortex for the first 3 days of a 5-session training phase, and to the posterior parietal cortex for the last 2 days; these regions were previously shown to be implicated in learning in this order. The training involved learning to solve arithmetic problems by applying an algorithm (Calculation) or by memory retrieval (Drill); both trained and untrained problems were used in a post-training testing session. Measures of attention and working memory were acquired before and after testing. tRNS decreased overall reaction times (RTs) and increased the rate of learning for Calculation problems during the training phase; it also decreased Calculation RTs for both trained and untrained (new) problems during the testing phase. Task difficulty was manipulated between groups by having subjects train on either a small or a large set of problems; all facilitative effects reached significance only for the more difficult condition. There were no effects of tRNS on attention and working memory, however tRNS' facilitative effects on learning arithmetic correlated negatively with performance on standardised arithmetic tests. These results confirm the ability of transcranial electrical stimulation to produce specific facilitative effects on numerical cognition and suggest that task difficulty is an important factor that modulates such effects.

3.2. Introduction

3.2.1 The neuroanatomy of mental arithmetic

Numerical competence is an increasingly good predictor of career success and general well-being in today's world (Parsons & Bynner, 2005). Conversely, poor numerical skills have adverse effects on the quality of life of the individual and on the economy of nations (Beddington et al., 2008). Learning arithmetical skills is therefore an important part of an individual's quantitative education, and understanding just how this learning takes place in the brain is of interest not only for basic research but also for understanding conditions such as developmental dyscalculia, where this learning is impaired (Butterworth, 2005a; R. Cohen Kadosh & Walsh, 2007). One of the more salient distinctions that can be made among the different types of arithmetic learning is that between learning by *drill* (memorisation) and learning by *calculation*. The former strategy involves committing arithmetic facts (such as multiplication tables) to long-term memory, whereas the latter involves applying a known algorithm (such as long division) in order to determine the result of a mathematical operation. The two strategies are not mutually exclusive, and are in fact typically used in a complementary fashion when arithmetical procedures are either taught (as part of mathematics education) or applied in everyday calculations. At the neural level, however, drill and calculation strategies appear to be subserved by independent mechanisms. For instance, retrieval of memorised arithmetic facts is largely associated with verbal representations in the left angular gyrus (Delazer et al., 2003; Grabner, Ansari, et al., 2009; Ischebeck et al., 2009), whereas calculation strategies during mental arithmetic have been shown to recruit the bilateral intraparietal sulci (Delazer et al., 2003; Stanescu-Cosson et al., 2000) as well as frontal regions such as the middle and inferior frontal gyri (Delazer et al., 2005; Pesenti et al., 2001).

One functional meta-analysis of studies involving both number tasks (e.g. number comparison, counting) and calculation tasks of mental arithmetic (Arsalidou & Taylor, 2011) further clarifies the relative contributions of frontal and parietal lobe in numerical cognition by suggesting that, whereas the posterior parietal cortex (specifically the intraparietal sulcus and the inferior and superior parietal lobules surrounding it) is involved with virtually any number-related task, frontal regions seem to be comparatively more specialised in arithmetical calculation. Functional activations of these regions, and in particular of the dorsolateral prefrontal cortex (DLPFC), in tasks involving arithmetical calculations have been found using both magnetic resonance (Kawashima et al., 2004; Zamarian, Ischebeck, & Delazer, 2009) and optical (Near-

infrared spectroscopy, NIRS) functional imaging methods (Pfurtscheller, Bauernfeind, Wriessnegger, & Neuper, 2010; Tanida, Sakatani, Takano, & Tagai, 2004).

FMRI investigations have shown that, ontogenetically, frontal functions are predominant in numerical cognition earlier in development, and are gradually complemented by recruitment of parietal areas such as the bilateral intraparietal sulcus. This observation holds both for basic processing of symbolic number (Ansari et al., 2005; Holloway & Ansari, 2009a) as well as for mental arithmetic (Kucian et al., 2008; Rivera et al., 2005). Cognitively, this suggests decreased reliance on domain-general resources such as attention and working memory during numerical cognition, and, in the case of mental arithmetic, it corresponds to a transfer from methodological (and computationally inefficient) strategies of calculation, to faster and more effortless strategies of memory retrieval (Grabner, Ansari, et al., 2009).

3.2.2 Brain stimulation and behavioural performance

The facilitatory effect found in studies using non-invasive brain stimulation (e.g., R. Cohen Kadosh et al., 2010; Snowball et al., 2013) is often expressed as a reduction in the mean or median reaction time required to perform a certain task. Distributional analyses of RTs, however, provide more detailed descriptions than just measures of central tendency, which can obscure finer aspects of performance. The shapes of these distributions can be used to make inferences about underlying processes, allowing for the testing of hypotheses that are indistinguishable when only comparing mean RTs (for a review, see Balota & Yap, 2011). Among the different distribution analyses, the *ex-Gaussian* function is a popular choice as it provides good fits for empirically obtained RT distributions, with parameters that are intuitively interpretable on the distribution histogram. The ex-Gaussian distribution is defined as the convolution between a normal and an exponential distribution (Heathcote, 1996; Heathcote, Popiel, & Mewhort, 1991; Lacouture & Cousineau, 2008), and its right-hand tail makes it a good fit to RT distributions, which are often positively-skewed. Fitting the ex-Gaussian to an RT distribution provides, for each subject, an estimate of μ and σ (the mean and standard deviation of the normal component) and of τ (the mean of the exponential component, i.e. the positive skew). The ex-Gaussian results from the convolution of the Gaussian and the exponential distribution, and the mean RT is approximated by the sum of μ and τ. μ reflects the position of the distribution along the x axis. An increase in μ reflects a uniform rightward shift of the main body of the distribution, whereas increases in τ corresponds to a rightward-skew of the distribution, which can reflect a slowing on some trials due to "later" processes than those contributing to μ. For tasks relying on working

memory, an increase in μ has been suggested to reflect a global slowing of memory retrieval, with reliance on reasoning contributing predominantly to τ (Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). In numerical cognition, ex-Gaussian parameters μand τ have been used to identify patterns of individual differences in simple mental arithmetic strategies (LeFevre, DeStefano, Penner-Wilger, & Daley, 2006).

An important empirical question in brain stimulation studies is not just *which* region to stimulate but also *when* to deliver the stimulation. This decision is informed by knowing when exactly a certain region becomes actively involved in the task. There is convincing evidence that stimulating an irrelevant "control" area produces no benefit above and beyond the cognitive training itself (e.g., Cappelletti et al., 2013), however even with a reduced subset of relevant candidate regions, the question of time of stimulation is appropriate; it might be that if an area is stimulated before or after it actually becomes involved, then the stimulation will not produce an optimal effect, even if that stimulation is "online" in relation to the task. Ischebeck, Zamarian, Egger, Schocke and Delazer (2007) "zoomed in" on the learning process and found that, as training in a mental arithmetic task progressed, the effect of learning consisted in a relative decrease of activation in frontal areas such as the precentral and inferior frontal gyri, and a relative increase of activation in parietal areas such as the angular gyrus. Similarly, a study that recorded ERPs during mental arithmetic training (Pauli et al., 1994) found that it is predominantly the prefrontal regions that are involved in the first part of the learning (when there is little automaticity), while parietal areas only start to be involved later in the process, imitating the early developmental stages. Another study (Delazer et al., 2003) had found arithmetic learning to be accompanied by a shift of activation within the parietal lobe, from the intraparietal sulcus to the left angular gyrus, suggesting a shift in strategy from calculation to retrieval (Grabner, Ansari, et al., 2009).

In the present study, I sought to probe the idea that cortical recruitment during arithmetic learning shifts from frontal to parietal. I did this using a mixed-model design in which subjects learned drill or calculation arithmetic for 5 days, while receiving (sham or real) frontal and subsequently parietal stimulation. I expected that real (as compared to sham) stimulation would facilitate learning, and that the amount of facilitation would be modulated by task difficulty, which was manipulated between-subjects.

3.3. Methods

3.3.1 Subjects

Thirty-two right-handed volunteers (14 female, 18 male; mean age = 22.38 years, SD = 3.37 years) were recruited. All subjects had normal or corrected-to-normal vision, and met the safety requirements of participation in a tES experiment, i.e. I excluded subjects who had a history of neurological conditions including seizures, implanted metal objects, heart problems or past fainting spells. Subjects were randomly assigned to receive either tRNS or sham stimulation with the constraint that age and gender distributions were similar in the two stimulation groups (tRNS group: 7 females, 9 males; mean age = 21.94 years, SD = 3.55 years; sham group: 7 females, 9 males; mean age = 22.81 years, SD = 3.23 years). The two groups did not differ significantly in terms of age ($F(1, 30)=0.532$, $p>.1$) or gender (Yates corrected $\chi^2=.13$, $p>.1$).

The experiment was double-blind, with neither the subject nor the experimenter aware of the type of stimulation received. This was achieved by having another experimenter start the stimulator in the appropriate condition (real or sham stimulation) at the start of the session, unbeknownst to the main experimenter. Informed consent was obtained before the start of each experimental session, and each subject received £50 compensation. Ethical approval was obtained from the local research ethics committee.

3.3.2 Training phase

Subjects completed a five-session training phase followed by a testing phase (Delazer et al., 2005). The training involved daily sessions from Monday to Friday, during which subjects performed two types of arithmetic learning tasks: *Calculation*, where two numerical operands were separated by the operator §, and *Drill*, where the operands were separated by the operator #. The result was always a double-digit number, which the subjects were asked to enter using a standard computer keyboard. Both conditions were presented in blocks of 18 trials. Within each session, there was an equal number of calculation and drill blocks, and their order was alternated. In keeping with the original paradigm (Delazer et al., 2005), the total number of blocks per session was different on each day (10 blocks on day 1, 12 on day 2, 14 on day 3, 16 on day 4 and 14 on day 5), in an attempt to have daily training sessions of equal duration. In order to manipulate task difficulty between subjects, only one set of operand pairs was used for the training session in half of the sample (the "easy" condition), meaning there were 6 calculation problems and 6 drill problems in total, whereas for the other half (the "difficult" condition), two sets of operands were used (i.e. 12 possible

combinations for each type of problem). The task was implemented using E-Prime (PST Inc. , Pittsburgh, USA) on a desktop PC. Accuracy and reaction time (RT) for each response was recorded.

Subjects were randomly assigned to one of four groups defined by the existence of 2 algorithms for the calculation problems and 2 different sets of numerical operands used in all problems (see Table 7). This assignment was done for the purpose of balancing the parameters of the task for each group, and was independent of the assignment of subjects to either the sham or the tRNS group. It was only the latter grouping that was taken as a factor in analyses.

Table 7: The combination of 2 calculation algorithms + 2 operand sets formed 4 groups, to which subjects were assigned randomly and independently of the stimulation group. This grouping was not used in subsequent analyses.

	Set 1	*Set 2*
Algorithm 1 (2×R-L+1)	**Group 1**	**Group 3**
Algorithm 2 (2×R+L-10)	**Group 2**	**Group 4**

3.3.2.1 Drill

Drill trials involved remembering associations between pairs of operands and a given result. Subjects were instructed to try to remember these associations throughout the training week. The result of each drill problem was computed using a certain algorithm; subjects were not given this algorithm (which was irrelevant once each pair of operands was coupled with its corresponding result), and were instructed to not try and guess it but instead to learn the associations by "drill". The problem and associated result were displayed on the screen for a length of time that halved daily throughout the week (500 ms in session 1, 250 ms in session 2, 125 ms in session 3, 62 ms in session 4 and 31 ms in session 5). A mask formed of hash symbols was then displayed for 250 ms, followed by a repetition of the problem, this time without the result, and with a limited time window for subjects to enter their response (3000 for session 1, 2500 ms for session 2, 2000 ms for session 3, and 1500 ms for sessions 4 and 5). The decreasing times of the first display and response windows aimed to encourage subjects to produce the result by relying increasingly on memory retrieval and decreasingly on observing the presented result. Feedback on the accuracy of the response was then displayed on the screen for 500 ms, and subjects could proceed to the next trial only after giving the correct response (see Figure 15).

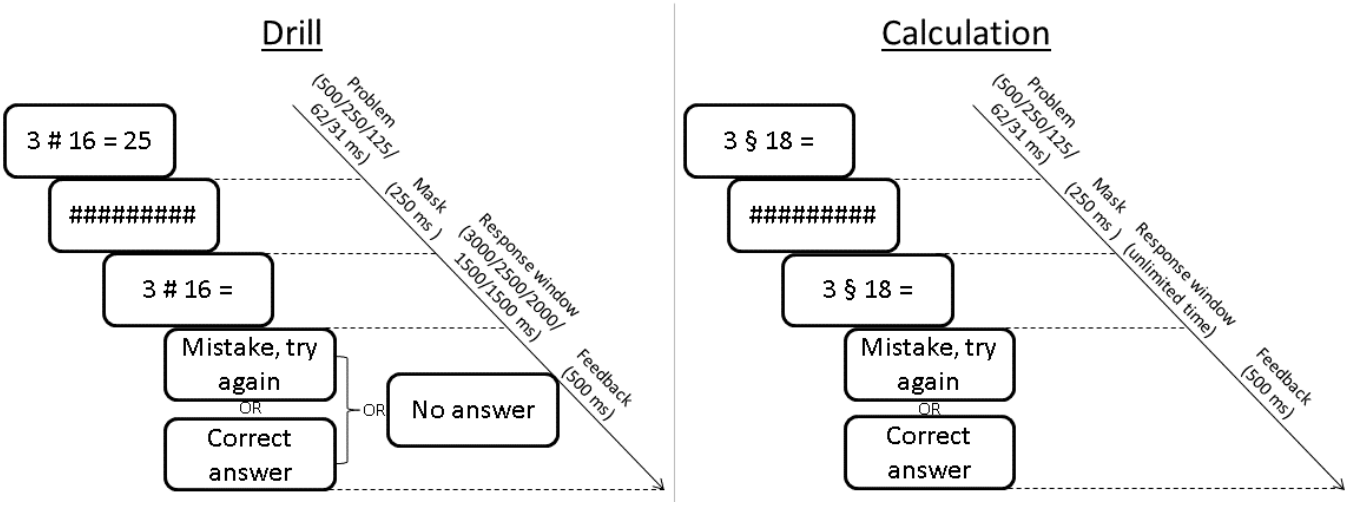


Figure 15: Structure of a trial for each type of problem. Left panel: example of a Drill trial; the result ("25") is presented for a brief duration, allowing the subject to memorise the association between the operands and the result, and then respond within a given time window. Right panel: example of a Calculation trial using algorithm 1 (2×R-L+1); the correct answer is "34".

3.3.2.2 Calculation

For each trial, subjects were shown a pair of operands and asked to calculate the corresponding result using an algorithm that involved multiplication, addition and subtraction. The two algorithms used were "2×R-L+1" and "2×R+L-10" (where L and R respectively denote the operands to the left and to the right of the "§" operator). Each set contained 6 pairs of operands; difficulty was manipulated between-subjects by having the training material consist of either one set ("easy") or two sets ("difficult"). Subjects were instructed to always apply the algorithm rather than making use of any memorised results. Parameters relating to display durations, feedback and proceeding to the next trial were the same as for Drill trials, except that here the response window never timed out (see Figure 15).

3.3.3 Testing phase

Following the final training session on day 5, subjects underwent a testing session, which aimed to assess whether the skills acquired for the two types of problems during training would transfer to new problems (pairs of operands). This session had the same structure as the training sessions, except (i) the operands included both "old" pairs (used during the training phase) and "new", untrained pairs; (ii) feedback was not given and subjects continued to the next trial regardless of the accuracy of their response.

In keeping with the original paradigm (Delazer et al., 2005), in the easy condition (with one training set), 3 sets were used for the testing session: one set of 6 old problems and 2 sets of new problems (6

problems for each set). For the difficult condition (2 training sets), the testing session used 2 sets: one set of old problems (6 problems that were randomly selected of the total of 12 trained) and one set of (6) new problems. Subjects were asked to give their best guess for "new" drill problems (whose results they had not been learning) in order to check whether the underlying algorithm had been deduced. A total of 96 trials were presented for each problem type, half of which were old and half new.

3.3.4 Control tasks

Subjects also completed a series of control tasks, in order to enable testing the specificity of the effect of tRNS. Before the start of training, the Numerical Operations and Mathematical Reasoning components of the Wechsler Individual Achievement Test (WIAT, Wechsler, 1997) were administered, in order to have a baseline measure of subjects' mathematical abilities (relating to arithmetical computations and language-based problems, respectively). On the other hand, subjects were administered the (forward and backward) digit span task and the attentional networks task (ANT) before the first and after the final (fifth) experimental session, in order to examine whether the brain stimulation affected verbal working memory or attention, respectively. The digit span is a subtest of the WAIS-III (Wechsler, 1997) and requires subjects to listen to a string of digits and then reproduce them in the same (forward digit span) or in reverse (backward digit span) order. To test attentional networks, I used the short version of the ANT, originally developed by Fan et al. (2005), which includes 3 blocks of 24 trials each. The ANT produces reaction time measures that can be used to quantify the efficiency of the three networks proposed to form the attentional system: alerting, orienting, and executive attention. In the ANT, subjects respond to the direction (left or right) of an arrow flanked horizontally by two other arrows on either side, with a warning cue appearing either instead or vertically on either side of a fixation cross (see Figure 16). The efficiency of the three attentional networks is assessed by measuring how reaction times are influenced by varying the warning signal interval, the validity of the spatial cue and the contrast between congruent and incongruent flankers (respectively). Specifically, the three resulting scores are computed as follows:

$$\text{Alerting effect} = RT_{\text{no cue}} - RT_{\text{center cue}}$$
$$\text{Orienting effect} = RT_{\text{center cue}} - RT_{\text{spatial cue}}$$
$$\text{Executive attention effect} = RT_{\text{incongruent}} - RT_{\text{congruent}}$$

For all tasks, subjects were seated in front of a 19" monitor, at a distance of approx. 90 cm.

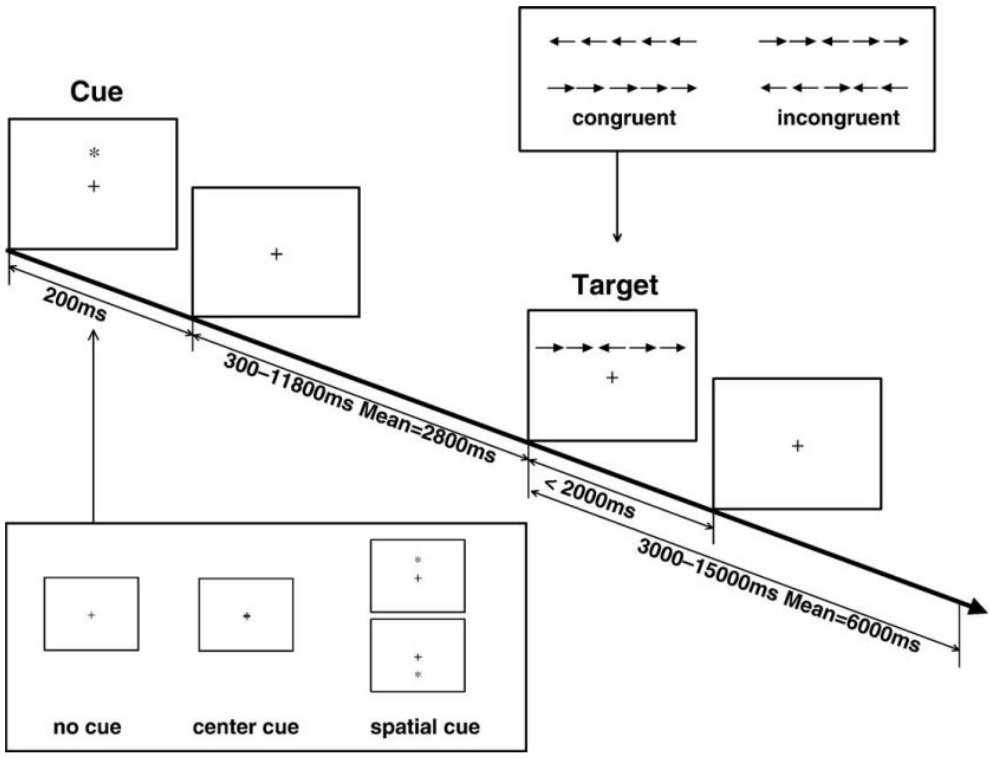


Figure 16: Structure of the attentional networks task (ANT). The two conditions – presence of spatial cue and flanker-arrows congruency – are depicted together with an example trial. Subjects pressed a button to indicate whether the central arrow pointed left/right. Reproduced, with permission, from (Fan et al., 2005).

3.3.5 Brain stimulation

All subjects received tRNS to the bilateral DLPFC (corresponding to scalp positions F3 and F4 in the international 10/20 EEG system) for the first three sessions of the training phase. For the last two sessions (days 4 and 5) of the training phase, the stimulation site was moved to the posterior parietal cortex (positions P3 and P4). The localisation of these sites was determined by using tape to measure the subject's head along the left-right and anterior-posterior dimensions, and then using a soft-tip pen to mark the locations corresponding to the measured percentages, in accordance to the 10/20 system diagram. Although of limited spatial resolution, this method is nonetheless suitable given the low level of current focality typically characteristic of tES electrodes (Nitsche et al., 2008).

Current in the form of high frequency noise (100-640 Hz) was delivered by a battery-driven current stimulator (DC Stimulator Plus, from Magstim, UK). The current intensity was 1mA peak-to-peak, with each sample being drawn from a normal distribution with mean 0 μA, and with 99% of all generated amplitude values lying between -500 μA and +500 μA. The stimulation duration was set to 20 minutes for the tRNS group and to 30 seconds for the sham group, both flanked by a 15 second upward and downward ramp. The stimulator was always started at the same time as the onset of the task and was turned off when the task

ended. The current was delivered through two 4x4 cm electrodes, which were secured in place using an elastic strap and were placed in saline-soaked sponges to improve electrical contact with the scalp and reduce the risk of skin irritation.

All subjects were asked, at the end of the experiment, whether they thought they were in the real stimulation or in the sham ("placebo") group, judging by the sensation that they felt underneath the electrodes during the stimulation.

3.3.6 Analyses

The RTs used in all analyses were measured from the onset of the problem display and until the final (second) digit of the response was pressed[10]. RT outliers (defined as further than 2.5 SDs away from the mean) were removed before calculating the mean RT of each session, separately for each type of problem and for each subject. On average, 0.7% of trials were removed as outliers.

A number of subjects had inordinately low accuracy scores for old drill problems in the testing phase, suggesting that, during training, they had not learnt the associations between operand pairs and their corresponding result, as instructed, instead just "transcribing" the isolated result. The "guessing range" was defined as 2.5 SDs around the mean accuracy of new drill problems, which resulted in the 0.50 mean accuracy value as a cut-off for defining outliers for old drill problems. This led to 11 subjects (i.e. 34% out of the total of 32) being excluded from all analyses of drill problems, for both training and testing; 4 of these were from the tRNS group and 7 from the sham group; 3 of them were in the easy condition and 8 in the difficult condition.

Mean accuracy and mean RTs on correct trials for the two different types of problems (Drill and Calculation) were submitted to separate 4-way mixed model analyses of variance with stimulation group (tRNS vs sham) and difficulty (easy vs difficult) as between-subjects factors, and Session (1 through 5) and Block (1 through 5) as within-subject factors. For each problem type, only the first 5 blocks of each session were taken into consideration for these factorial analyses (Snowball et al., 2013). In all ANOVAs having within factors of at least three levels, in case of violations of the sphericity assumption (checked with Mauchly's test), the Greenhouse-Geisser correction was applied..

[10] The second digit was used rather than the first because it was considered that the subject would only at that stage have the complete answer in mind. Nonetheless, all RT analyses were also repeated with the RT measured from the onset of problem display until the *first* digit; the statistical significance of all results stayed the same.

3.3.6.1 Modelling learning curves

Curves described by power laws have been shown to be a good model (the "power law of practice") for the improvement in RTs that accompanies practice-based learning in several types of cognitive domains, including mental arithmetic as well as memory recall of arithmetic facts (Newell & Rosenbloom, 1981; Rickard, 1997).

Using a power law, RT is expressed as a function of time (session number) as follows:

$$RT = B \times N^{-\alpha},$$

where *N* is session number, *B* is a constant signifying initial performance (RT) for N=1, and *α* is the learning rate; a positive α is indicative of decreasing RTs, with higher values reflecting faster learning. I fitted this power law to Calculation RTs and looked for an effect of stimulation group on α, a dependent variable also used in previous training studies (Iuculano & Cohen Kadosh, 2013). I used non-linear estimation, using the least squares method, to estimate the unknown parameters B and α for each subject, across the five sessions.

3.4. Results

3.4.1 Training phase

3.4.1.1 Mean accuracy and RT

The only significant effect for all analyses was a group × difficulty interaction for Calculation RTs ($F(1, 28)=7.79$, $p<.01$, $\eta_p^2=.22$), depicted in Figure 17a; all other effects were $p>.1$. Planned comparisons for this interaction revealed that, in the difficult condition, the tRNS group was significantly faster than the sham group ($F(1,28)=6.78$, $p<.05$). By contrast, for the sham group only, the difficult condition was slower than the easy condition ($F(1, 28)=11.64$, $p<.005$). The session means for each difficulty condition are depicted in Figure 17b. The same two plots for Drill trials are depicted in Figure 18.

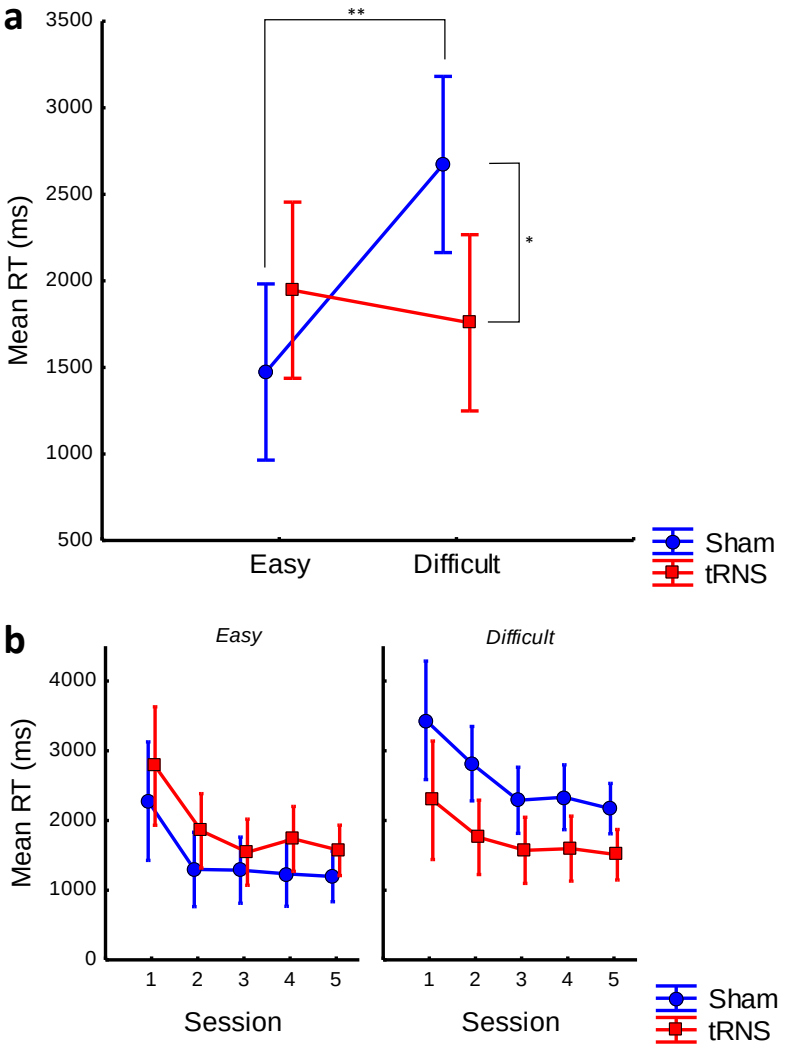


Figure 17: RTs on Calculation trials: (**a**) Overall means, as a function of difficulty and stimulation group. Significant differences are marked with * ($p<0.05$) and ** ($p<0.01$); (**b**) Session-level means for each difficulty condition, as a function of stimulation group. Error bars represent 95% confidence intervals.

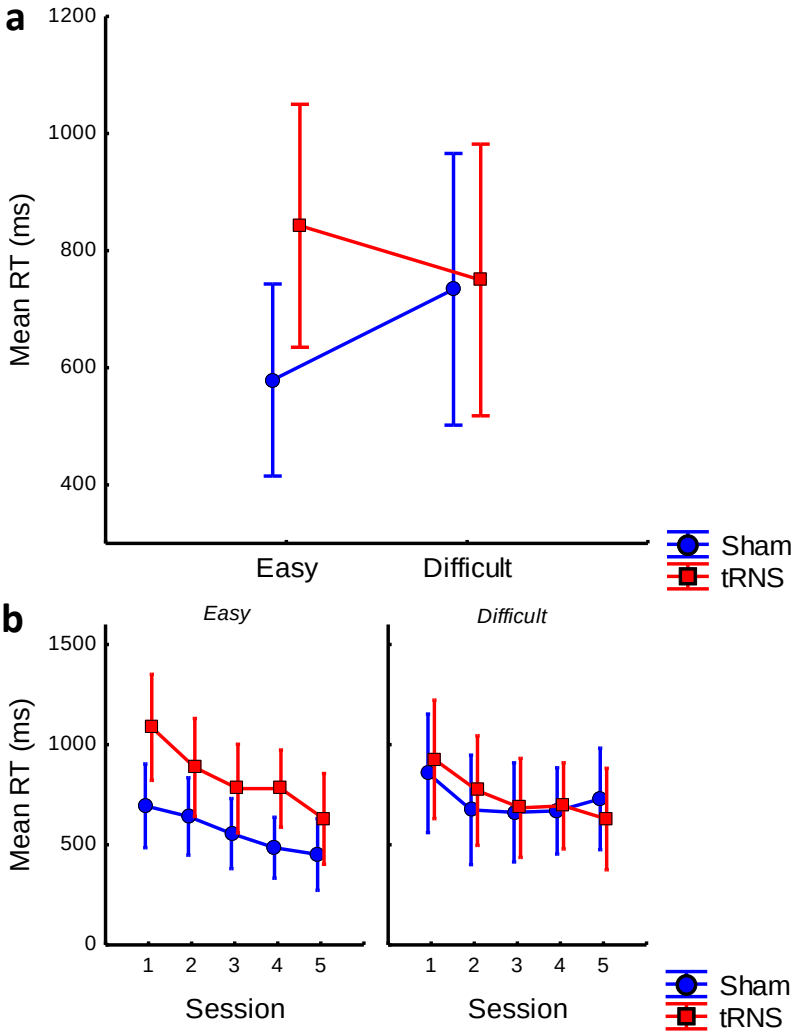


Figure 18: RTs on Drill trials: (**a**) Overall means, as a function of difficulty and stimulation group; (**b**) Session-level means for each difficulty condition, as a function of stimulation group. Error bars represent 95% confidence intervals.

Since the effect of electrical brain stimulation has been shown to manifest as early as a few minutes after current onset (Nitsche & Paulus, 2000), it could be that the lack of a significant interaction between group and either session or block masked an effect occurring during the blocks of the first session, with the two groups possibly diverging early on in the session. Additional ANOVAs were therefore run only for the blocks of session 1. There were no significant main or interaction effects for either RTs or accuracy. It is therefore plausible that the main effect of group reported above was due to an initial difference in RTs between the two groups, that was simply maintained throughout the training rather than attenuated or amplified by tRNS. This is indeed what Figure 17b, which shows RTs for calculation problems across sessions, suggests at a descriptive level.

3.4.1.2 Modelling learning curves

A 2-way between-subjects ANOVA for the dependent variable B, with stimulation group and difficulty as factors, revealed no significant main effect of group, nor an interaction (all ps>.2), indicating no group difference in initial performance. However, in order to control for inter-subject variability in this parameter, I used B as a covariate in the analysis of covariance (ANCOVA) done for dependent variable α (Snowball et al., 2013), where stimulation group and difficulty were between-subjects factors. There was a significant group × difficulty interaction ($F(1, 27)=4.60$, $p<.05$, $\eta_p^2=.15$), depicted in Figure 19. Planned comparisons revealed a non-significant effect of group for either of the two difficulty conditions, but a significant effect of difficulty for the sham condition ($F(1,11)=14.72$, $p<.001$, $\eta_p^2=.32$), showing decreased learning rate for this group when difficulty was increased.

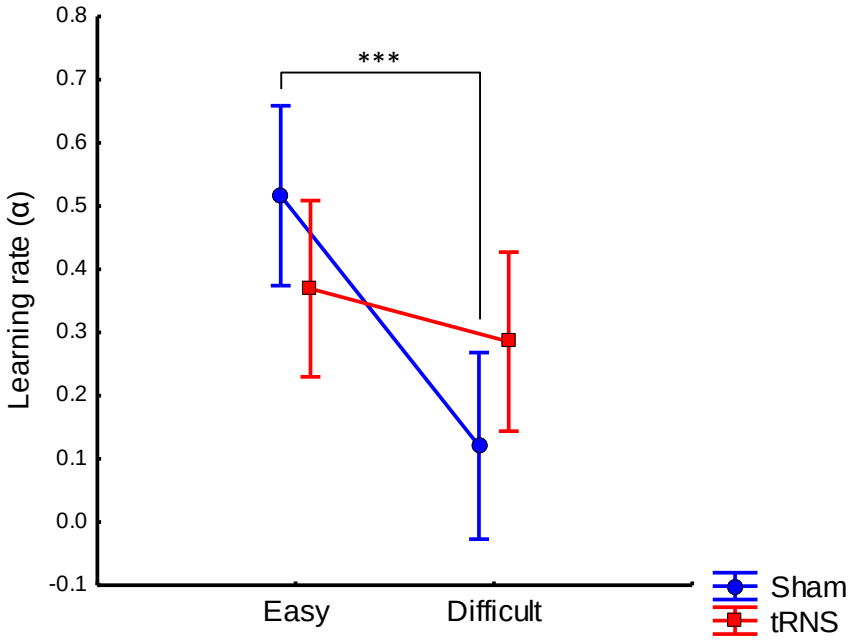


Figure 19: Calculation learning rate (α) for the two stimulation groups, under the two different difficulty conditions.

Using the power law to model *drill* RTs would not have been reliable, as one assumption of the power law is that the task difficulty stays constant throughout the learning phase (Newell & Rosenbloom, 1981), whereas the difficulty of the drill problems here increased slightly with every session, due to the decreased display and response window times.

3.4.1.3 Reaction time distribution analysis

For each session[11] of each subject, the RT distribution for each type of problem was submitted to an analysis which estimated the ex-Gaussian parameters μ, σ and τ (Lacouture & Cousineau, 2008). Each of these was then used as a dependent variable in separate mixed-model ANOVAs with group and difficulty as between-subjects factors, and session as a within-subjects factor. The only significant effects were for Calculation: a group × difficulty interaction for μ ($F(1,28)=4.76$, $p<.05$, $\eta_p^2=.15$) and a trend towards the same effect for τ ($F(1,28)=3.45$, $p=.074$, $\eta_p^2=.11$), as per Figure 20**c**.

[11] The RT distributions were calculated at session level and not at block level because in order to fit the ex-Gaussian to empirical data, at least 100 observations per distribution are typically needed (Ratcliff, 1979).

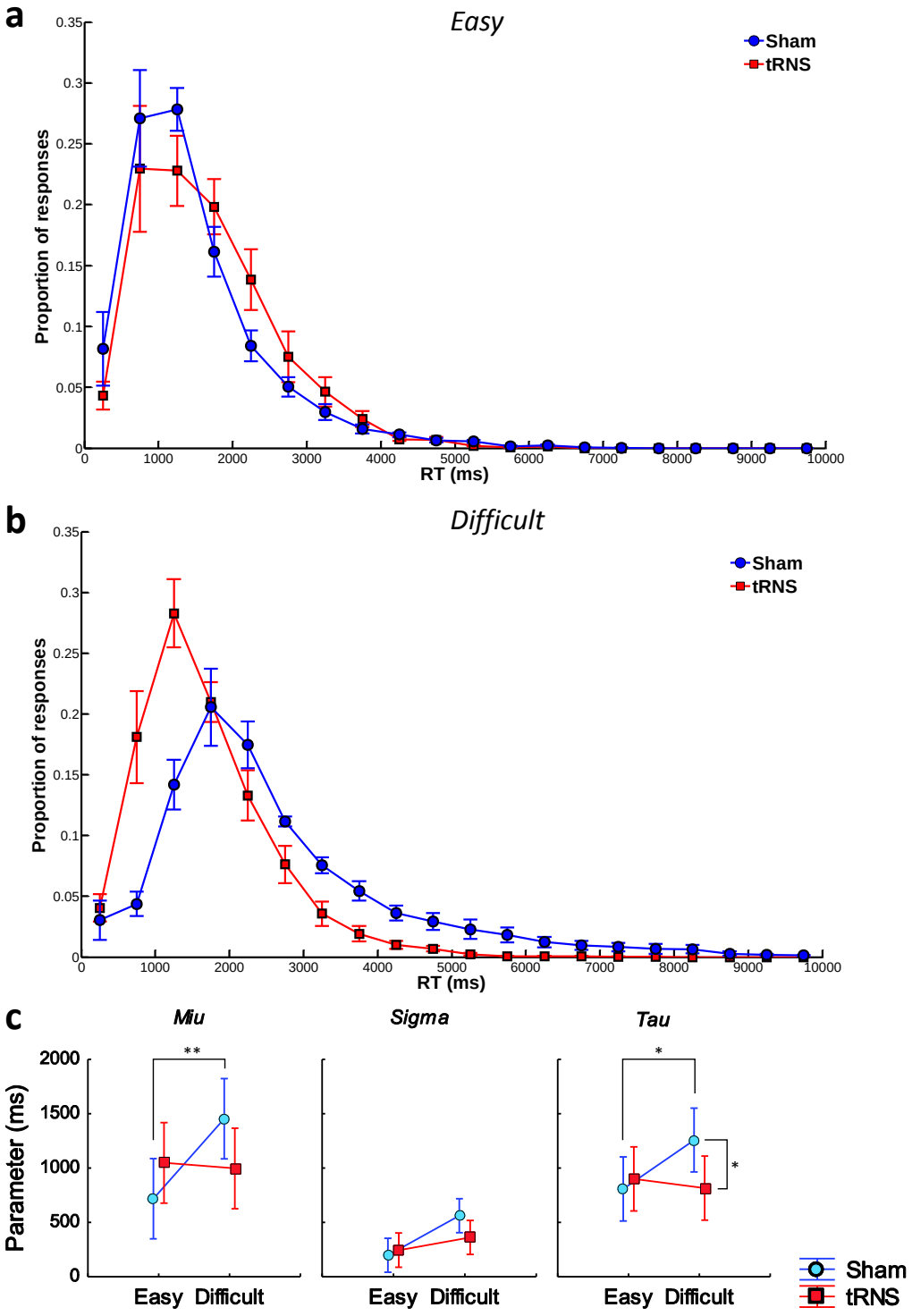


Figure 20: Distribution of Calculation RTs for the easy (**a**) and the difficult (**b**) conditions, separated for the two stimulation groups; the τ parameter of the ex-Gaussian is significantly different between the groups, for the difficult condition only; each data point represents the average proportion of responses across subjects, within 500 ms-wide bins. (c) The ex-Gaussian parameters that describe these distributions, as a function of difficulty and stimulation group.

Planned comparisons revealed that, for the difficult versus the easy condition, the sham group had significantly higher values of both μ ($F(1,28)=8.33$, $p<.01$) and τ ($F(1,28)=4.88$, $p<.05$). These effects were not

present in the tRNS condition (p>.1). Additionally, the sham group had significantly higher τ values than the tRNS group within the difficult condition (F(1,28)=4.73, p<.05); the same contrast approached significance for μ (F(1,28)=3.32, p=.083). Panels **a** and **b** of Figure 20 show the RT distribution for Calculation, for each difficulty condition. In the second case, the tRNS group's earlier peak is indexed by the (descriptively) smaller μ, while the smaller τ reflects the shallower exponential tail.

3.4.1.4 Correlation between numerical abilities and learning rate

In order to see whether tRNS affects learning rate differently for people of different maths abilities, I computed the correlation between WIAT's Numerical Operations and α, separately for the two stimulation groups. This correlation, depicted in Figure 21, was non-significant for the sham group (r=-.12, p>.1; α=0.34±0.30) but significant for the tRNS group (r=-.52, p<.05; α=0.31±0.22), suggesting that people of a lower arithmetical skill benefit more, i.e. improve faster during tRNS-assisted training, as compared to people of a higher arithmetical skill – in other words, that there is "more room to improve". For the sham group, said correlation reflects how likely people of different arithmetical skill are likely to benefit from *training* alone, without the *stimulation*. Since the (tRNS group) subject having the lowest score in Numerical Operations could be considered an outlier (as the score is more than 2.5 standard deviations away from the mean), a jack-knife ("leave-one-out") resampling procedure was additionally conducted within this group. In this analysis, significance was maintained in only 11 out of 16 cases (69%), thus qualifying the robustness of this effect. Also, the correlation coefficients were not significantly different for the groups (p>.1), which does not enable the stated question to be answered in a way that teases apart the effect of tRNS from the effect of training. Finally, it should be noted that, as can be seen in Figure 21, it could be argued that a ceiling effect was present for the Numerical Operations score across both groups, which might have contributed to the decrease of the statistical robustness of the correlation.

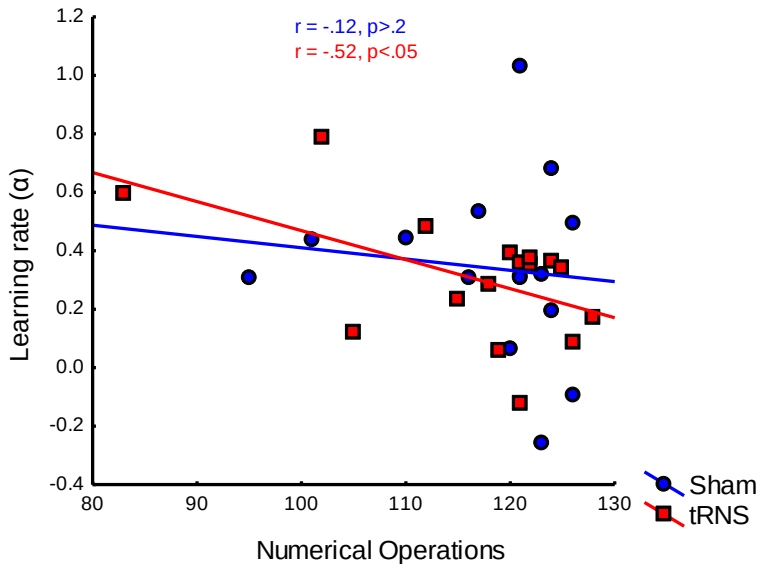


Figure 21: Relationship between learning rate and Numerical Operations score as a function of stimulation type.

3.4.2 Testing phase

Mean RT and accuracy scores for each type of problem were submitted to a 2×2×2 mixed-model ANOVA with stimulation group and difficulty as between-subjects factors, and problem novelty (old or new) as within-subjects factor. Numerical Operations was taken as a covariate for Calculation analyses.

There were no significant effects for Drill.

For Calculation RTs, there was a significant group × difficulty interaction (p=.032), depicted in Figure 22. It is noteworthy that this effect is similar to the one obtained for training Calculation RTs (Figure 17). Planned comparisons revealed lower RTs in the tRNS than the sham group within the difficult condition ($F(1, 28)=4.70$, $p<.05$), and also higher RTs for the difficult than for the easy condition within the sham group only ($F(1, 28)=5.23$, $p<.05$). This effect was across the problem novelty factor, which did not interact with group (group × difficulty × problem novelty: $p>.1$).

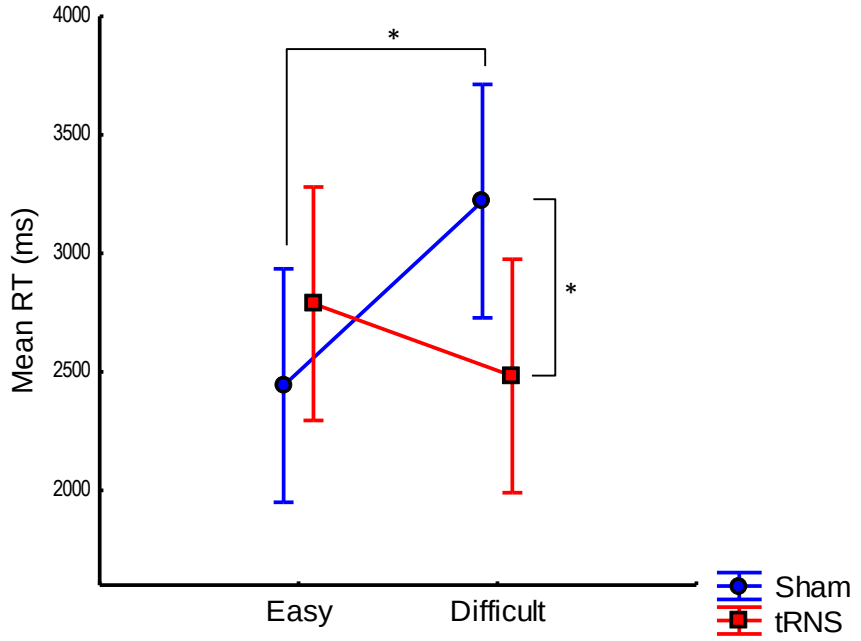


Figure 22: Mean RTs on Calculation problems during the testing phase, as a function of group and difficulty.

For Calculation accuracy, there was a significant group × difficulty × problem novelty interaction ($F(1, 24)=5.30$, $p<.05$), depicted in Figure 23. The group × difficulty interaction was significant for new problems only ($F(1,24)=5.73$, $p<.05$). Planned comparisons revealed lower accuracy in the sham than the tRNS group, in the easy condition for new problems ($F(1,24)=7.45$, $p<.05$), meaning that the increased difficulty presented by new (untrained) problems was only in the case of the sham group reflected in a lower mean accuracy; planned comparisons also revealed lower accuracy in the easy than in the difficult condition, in the sham group for new problems ($F(1,24)=6.89$, $p<.05$), meaning that, when the trained problems were few enough to learn by recall, the subjects in the sham group performed worse with untrained material as compared to when a higher training load (as in the difficult condition) made learning by recall harder.

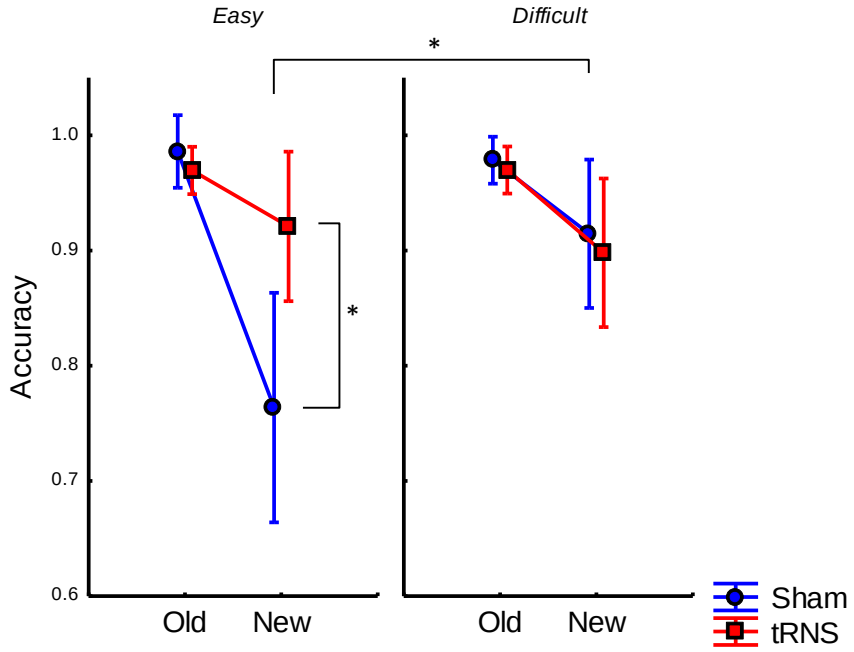


Figure 23: Mean accuracy on Calculation problems during the testing phase, as a function of group, difficulty and problem novelty.

3.4.3 Control tasks

Scores were calculated separately for the three attentional networks (alerting, orienting and executive attention), and each score was submitted to a mixed-model 2-way ANOVA with session (1 vs 5) as within-subjects factor and stimulation group (tRNS vs sham) as between-subjects factor. There were no significant main effects or interactions for any of the three dependent variables (ps>.2; see Table 8). Stimulation did not differentially affect performance on the forwards and backwards digit span tasks either, with no main effect of group and no interaction between group and session (ps>.2).

In terms of baseline mathematical ability, the stimulation groups did not differ on either of the WIAT measures, Numerical Operations and Mathematical Reasoning (independent samples t-tests, all ps>.2).

Table 8: Scores from ANT, digit span and WIAT.

Test	Component	M ± SD (TRNS group)		M ± SD (sham group)	
		Session 1	Session 5	Session 1	Session 5
ANT	Alerting	23.06 ± 15.72	28.19 ± 16.24	24.82 ± 23.56	28.72 ± 10.69
	Orienting	46.62 ± 31.08	47.2 ± 19.86	49.71 ± 25.84	53.89 ± 22.16
	Executive attention	72.38 ± 13.08	64.39 ± 12.33	79.8 ± 14.65	69.95 ± 17.84
Digit span	Forward	11.31 ± 2.70	12.13 ± 3.22	10.81 ± 2.51	11 ± 2.22
	Backward	9.06 ± 3.30	8.75 ± 3.71	7.56 ± 2.68	7.19 ± 2.66
WIAT	Numerical Operations	116.44 ± 11.43		118.13 ± 8.88	
	Mathematical Reasoning	116.44 ± 9.67		116.19 ± 9.36	

3.4.4 Stimulation perception

Table 9 shows the subjects' answers when asked, at the end of the experiment, whether they thought they were administered real or sham stimulation, based on their sensation during the training sessions.

Table 9: Contingency table with subjects' reflections on stimulation sensation.

	Response: "tRNS"	Response: "Sham"	Response: "I do not know"	TOTAL
tRNS group	2	8	6	16
Sham group	6	7	3	16
TOTAL	8	15	9	32

A chi-squared test on this contingency table revealed that subjects were unable to correctly identify their stimulation group (Yates corrected χ^2=.75, p>.1), which is in line with previous studies suggesting that tRNS offers good blinding control (Ambrus et al., 2010; Fertonani et al., 2011; Prichard et al., 2014).

3.5. Discussion

I found that tRNS improved performance on one type of arithmetic learning process (Calculation), where subjects had to learn to apply a given formula to different sets of operands, but not on another (Drill), where given results had to be memorised and later retrieved. This improvement was observed as overall faster reaction times for the tRNS group during training as well as during the testing phase, but only in the more difficult of the two between-subjects conditions. For that condition versus the less difficult one, learning rate decreased only for the sham group but not for tRNS. Cumulatively, these results suggest that tRNS-assisted training, here using mixed frontal-parietal stimulation, mitigates the effect of increased difficulty or that, equivalently, its effect only comes into play when the task is sufficiently challenging.

3.5.1 Training phase

During training, stimulation group did not interact with either block or session, giving rise to the relatively constant difference in RTs between the groups. It is, nevertheless, possible that the significant stimulation group difference found in the difficult condition was due to the stimulation and not to a pre-existing baseline difference between the groups, because: (i) the subjects were randomly assigned to stimulation groups; (ii) the groups did not differ in terms of the baseline of calculation ability (as measured

before the start of stimulation by the Numerical Operations component of the WIAT), or in their measure of initial performance (B), the latter of which was also co-varied out of the learning rate (α) analysis; and crucially, (iii) α itself was higher for tRNS than for sham.

Stimulation group differences were only observed when the task was more difficult. At the same time, the 'cost' for increased difficulty (i.e. increased RTs) was incurred only by the sham group and not by the tRNS group. These results are consistent with the fact that tES selectively increases excitability in neural populations active at the time of stimulation, and, in particular, with a previous study that found the effectiveness of tES increased with task difficulty (Pope & Miall, 2012). Thus, it could be that, for subjects trained in the easy condition, the beneficial effect of the tRNS was not fully "exploited", and that, for the other group, increased task difficulty perhaps acted to "amplify" the facilitation beyond a certain critical threshold. Or, alternatively, that due to the small number of problems, subjects simply switched to rote retrieval for all Calculation problems[12], a strategy that was not facilitated by tRNS. It is conceivable that there is also an upper threshold for task difficulty, and that once this threshold is exceeded (e.g. if I also had conditions for 3 and 4 trained sets), then tRNS would no longer mitigate the effect of increased difficulty.

It is possible that the null results for Drill trials are due to the fact that a large proportion of subjects had to be excluded from the Drill analyses due to their evidently incorrect strategy for performing the task. As compared to Calculation at least, this may have decreased the statistical power and thus increased the probability of a Type II error. Having mentioned this, however, the present study is in agreement with the study by Snowball et al. (2013), which also did not find any group effects (in terms of RT or accuracy) for the Drill testing phase, in either their frontal-only or parietal-only conditions. The shift in stimulation site (with its link to learning-related shifts discussed earlier) in the present study could have been expected to favour a facilitatory effect of memory recall more than of calculation strategy. Such an effect would likely be more visible during the Drill *testing* phase, when subjects had to recall all answers from memory, than during its *training* phase, when the presented result allowed circumvention of memory recall. Since a group difference was not observed during testing, it is possible that either (i) the shift in stimulation site did not, in fact, facilitate memory recall, or (ii) subjects' strategies for Drill never really made sufficient use of memory recall for the facilitation of this strategy to lead to a detectable effect in mean RTs.

[12] It has to be noted here that, although it was my aim here to dissociate the two strategies, there is most likely an element of 'Drill' in 'Calculation' problems for both the 1set and the 2sets training conditions; this is suggested for instance by the large RT difference between old and new Calculation problems.

The group difference effect in mean Calculation RTs in the difficult condition is complemented by the effect based on the ex-Gaussian fit to that particular RT distribution; specifically, tRNS reduced the exponential component (τ) as compared to sham. Greater values of τ imply a stronger rightward skewing of the RT distribution, which in turn has been suggested (Luce, 1986) to underlie later cognitive processes, as opposed to earlier or more perceptual ones such as the time required to register the problem and physically make a response. In a study where subjects solved addition problems in different formats (Arabic numbers or number words) and reported their strategy (calculation or memory retrieval) for each trial, an ex-Gaussian analysis suggested that, for each format, retrieval trials contributed predominantly to μ and calculation trials to τ (Campbell & Penner-Wilger, 2006). Similarly, LeFevre et al. (2006) found that τ discriminated well between the different self-reported strategies (such as retrieval, transformations or counting) used in a simple subtraction task, having small values for retrieval and large values for counting. Applying such an interpretation - based on the assumption that RTs on calculation trials will be sampled from a different distribution than RTs on retrieval trials - to the present study, the smaller τ of the tRNS group might reflect faster or more efficient calculation. Alternatively, it could suggest that tRNS promoted a transfer from slow calculation strategies to fast memory-based ones (Grabner, Ansari, et al., 2009), when the task more difficult. However, the lack of group differences for Drill, in either RTs or accuracy, for both training and testing, leaves enhanced calculation, rather than a switch from calculation- to memory-based strategies, as the more likely explanation.

3.5.2 Testing phase

During Calculation testing, accuracy for new problems was, for the sham group, lower in the easy than in the difficult condition, suggesting that perhaps in the former condition, where relatively few Calculation problems were involved in the training, these were being solved by memory recall, thus making new problems during the testing phase - where calculation actually had to be applied - more difficult simply because a previously unrehearsed strategy (namely, that of mental calculation) had to be used. This effect did not occur for the tRNS group, which might suggest that here, perhaps due to a facilitation of calculation strategies during training (Snowball et al., 2013), such a strategy was used more consistently during both training and testing.

Problem novelty did not interact with stimulation group and task difficulty for Calculation RTs, suggesting that the tRNS facilitation acted equally across both trained and untrained material. This is similar

to the results of Snowball et al. (2013), where transfer of RT facilitation was also found to occur across old and new Calculation problems. In both studies, however, this effect did not also hold for Calculation accuracy. The training literature has mixed results – both positive (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Klingberg, 2010) and negative (H. Lee et al., 2012; Owen et al., 2010) – on the transferability of skills across- or even within-domain, and there are several possible reasons for unsuccessful transfer of trained abilities, including an inappropriate training task (or insufficient practice on it) and an inappropriately "new" task being used to test whether transfer has occurred. Another possible explanation for the lack of accuracy facilitation and transfer could be speed-accuracy trade-off (Pachella, 1974), in other words if subjects focused on accuracy at the expense of speed or vice-versa. This would be indicated by a positive correlation between accuracy and RT – such a correlation was, however, not significant for either of the groups. The likely remaining explanation might then simply be a ceiling effect for the accuracy of old Calculation problems – indeed, its average across the groups was 98%, which might have obscured the effect of tRNS in terms of accuracy, only leaving it visible in terms of RT. It is also possible that the lack of transfer to new problems in the difficult condition in particular (i.e. the lack of a group difference in accuracy) was due to the mixed stimulation site paradigm that I used here, although the mechanisms by which this might have occurred remain elusive even as a speculation. Unfortunately, a direct comparison with the testing phase results of Snowball et al. (2013), where a single-site stimulation paradigm was used, is not possible, as in that study a difficulty manipulation did not exist. It is, in general, not possible for the current study to ascertain whether stimulating a single brain region, as opposed to a succession of brain regions (according to a hypothesis about the order in which these are recruited), is more or less efficient in terms of enhancement of behavioural performance. Future studies are needed in order to further explore the potential of such mixed-stimulation paradigms to perhaps increase effect size in the domain of cognitive enhancement.

3.5.3 tRNS selectivity

Stimulation did not affect performance on control (attention and working memory) tasks, suggesting that, as in other tES studies of numerical cognition (Cappelletti et al., 2013; R. Cohen Kadosh et al., 2010; Snowball et al., 2013), the effects were specific to the trained material and do not extend to domain-general abilities. It is, however, very difficult to rule out the possibility that some hidden cost exists for a brain region other than the one stimulated, or with regard to cognitive tasks other than the ones tested (Iuculano & Cohen Kadosh, 2013). Such an assumption is supported by yet unsubstantiated – but speculatively plausible –

physiological mechanisms of tES that act as a "law of conservation", whereby metabolic resources are simply reallocated within the brain rather than selectively augmented. A limitation of the current study is that it does not enable me to test whether such a cost exists and, if it does, how it is manifested.

3.5.4 Conclusion

Overall, the results of this study confirm the ability of tRNS to selectively improve aspects of behavioural performance, and in particular performance in high-level cognitive training, provided that the level of cognitive load during that training is appropriate. Together with already-established effects of tDCS (R. Cohen Kadosh et al., 2010; Hauser et al., 2013; Iuculano & Cohen Kadosh, 2013), this proven reliability of tES techniques makes it an affordable and promising tool to alleviate symptoms of specific learning disabilities such as dyscalculia or dyslexia, as well as improve normal functioning in those domains. In future studies, it would be interesting to further explore the link between following strategy shifts that are known to occur at the behavioural level with shifts in the specific brain site that is being stimulated at any one time. It would also be relevant to see how the timing of the stimulation (online/offline) interacts with the length of time for which each region is stimulated. Numerical cognition, and mental arithmetic in particular, should continue to provide a fertile medium for exploring such new ground.

4 Study 3: The structural correlates of tRNS-assisted arithmetic learning

4.1. Abstract

In this study I aimed to investigate how learning of calculation- and retrieval-based arithmetic problems can be enhanced by combining tRNS with cognitive training, and what are the structural changes that are associated with such enhancement. A similar training paradigm – involving retrieval- or calculation-based arithmetic problems – was used as for Study 2. tRNS was applied as either frontal, parietal or sham stimulation (between-groups) for all 5 days of training. Structural and diffusion MRI scans were acquired from 72 subjects before and after training. Behaviourally, tRNS decreased RTs on the Calculation task for both frontal and parietal stimulation as compared to sham. Unlike in Study 2, these enhancements were not accompanied by an increase in learning rate. Group differences in the average within-session (online) and between-session (offline) learning magnitudes were found. The two types of learning were negatively correlated both within and across stimulation groups, suggesting that online learning takes place at the expense of offline learning and vice-versa. Anatomically, at a whole-brain level there were no group differences in the pre-to-post-training change in grey matter density and fractional anisotropy (FA), and no correlations between those measures and baseline and learning-related behavioural scores. Null results were also obtained for these measures in further analyses where regions of interest (ROIs) were defined based on the stimulating electrodes and on significant clusters obtained in Study 1. Pre-to-post change in FA along homotopic tracts based on these ROIs was also not different between groups. Several group effects were, however, found in terms of the pre-to-post change in surface-based measures (volume and thickness) of frontal, parietal and temporal clusters. These effects consisted of both increases and decreases in the frontal

and parietal groups as compared to sham. Also, these clusters were not proximal to the stimulation site, which suggests that it is not straightforward to predict either the regions of the brain receiving the maximal tRNS effect, or the direction of change in structural parameters of these regions as a result of tRNS. The results of this study on one hand partly replicate those of Study 2 and others that indicate the ability of tRNS to enhance behavioural performance in arithmetic. They also take our understanding of the mechanisms of action of tRNS further, by showing that such behavioural enhancements are accompanied by altered structural parameters in regions of the brain that, while relevant to numerical cognition, are not necessarily proximal to the stimulation site. Overall, these results represent the first longitudinal investigation into the effects of tRNS-assisted cognitive training on not only behavioural performance but structural brain changes.

4.2. Introduction

4.2.1 Strategies of mental arithmetic in the brain

Previous empirical chapters of this thesis, as well as the General Introduction, contain broader background information on how number understanding and manipulation is carried out in the brain. In this chapter I will focus on aspects that are specific to the current study, namely on the distinction between procedure- and retrieval-based arithmetic, and on longitudinal evidence of structural brain changes as a result of training.

Numerical skills are crucial predictors of an individual's quality of life and of nations' economies (Beddington et al., 2008; Butterworth, 2005b). One of the first steps in life towards acquiring advanced numerical competence such as mathematics is learning arithmetical skills. Such learning is disrupted in children suffering from developmental dyscalculia (DD), a specific learning disability affecting on average about 5% of the general population (Butterworth, 2005a). Perhaps the most important knowledge that normative psychology can confer on DD rehabilitation programmes is an improved understanding of the typical acquisition process taking place in arithmetic learning. This same knowledge can also help optimise maths educational strategies for typically developing children.

Within the domain of mental arithmetic, declarative knowledge of overlearned arithmetic facts (e.g. 8 × 5 = 40) can be distinguished from procedural knowledge (e.g. knowing the steps for long multiplication of 34 × 67). In fact, educationalists have long debated which of the two types of strategies should receive more emphasis in the process of learning arithmetic (Thorndike, 1922). While both strategies can play a role across

all four basic mathematical operations (addition, subtraction, multiplication and division of integers), declarative knowledge is mostly useful as a basis for simple multiplication, division and addition, while procedural knowledge is mostly useful in long multiplication, addition and subtraction. As mentioned in the General Introduction (Chapter 1), from the perspective of the triple-code model (Dehaene et al., 2003), multiplication relies on fast retrieval of arithmetic facts from verbal long-term memory, whereas subtractions are solved by directly manipulating the magnitudes involved in the operation.

Learning by drill and learning by procedure also dissociate at the level of the associated brain regions. For example, the use of retrieval strategies has consistently been associated with involvement of the left angular gyrus (Delazer et al., 2003; Grabner, Ansari, et al., 2009; Zamarian et al., 2009), an area whose activity seems to be parametrically modulated by individual differences in arithmetic performance, specifically in mathematical-numerical intelligence (Grabner et al., 2007) and in performance in previously-untrained multiplication problems (Grabner, Ischebeck, et al., 2009). On the other hand, when new arithmetic material is contrasted with material that was previously trained, a different part of the parietal lobe, the IPS, is activated instead (Delazer et al., 2003). An interesting role in calculation-based arithmetic is also played by the superior parietal lobule (SPL), the visuo-spatial component in Dehaene's et al. triple-code model (2003). In a more recent study from that same group, Knops et al. (2009) trained a multivariate classifier algorithm to infer the direction of eye movements (left/right), based on fMRI activity in regions of the posterior parietal cortex, identified as being involved in saccadic eye movements. This training, using data recorded from the posterior SPL (PSPL) alone, also enabled the classifier to generalise to an arithmetic task, in which it was able to distinguish subtraction from addition trials, regardless of notation (digits or dots). When the same algorithm was applied to the neighbouring hIPS (horizontal segment of the intraparietal sulcus), no generalisation from saccades to calculation was found. These results suggest that this role of the PSPL – the ability to generalise eye movement classification to arithmetic – is quite unique within the parietal lobe. It also explains why the PSPL is involved in mental arithmetic problems on both symbolic and non-symbolic notation, and, more broadly, also explains its role in a large variety of tasks involving, as a common denominator, the representation of and attention to spatial locations.

Procedure- and retrieval-based strategies do not only dissociate at the level of cortical activation. More recently, further dissociations have been identified at the level of the white matter connectivity between these cortices. For instance, Klein et al. (2013) asked healthy adults to perform addition problems relying on

either verbally-mediated fact retrieval or procedural manipulations. Using probabilistic tractography, the authors found that these two types of processing are subserved by two largely separate networks, involving both dorsal and ventral connections (superior and inferior longitudinal fasciculi, respectively). Also, a complex and innovative recent study by Qin et al. (2014) set out to investigate the functional and connectivity changes taking place in the developing brain as cognitive strategies in a given domain shift from procedure-based to retrieval-based. Given the wealth of existing knowledge documenting how this strategy shift occurs in the special case of arithmetic, the authors used an addition task as a paradigm to study functional activation and connectivity. Longitudinal scans of children were acquired at two time points (between the ages and 7 and 9), and, independently, cross-sectional scans were acquired from adolescents and adults. As opposed to the prefrontal-to-parietal shift in functional focus that has been well described in the developmental literature (and reviewed in the General Introduction, specifically section 1.2), the authors instead found, within the longitudinal children scans, a shift from prefrontal and parietal areas to medial-temporal areas (and in particular the hippocampus). Functional activity in this region followed an inverse U-shaped pattern, first increasing through childhood and then decreasing again through adolescence and adulthood. This provides support for certain models of long-term memory consolidation, which assert that the role of the hippocampus in the early phase of knowledge acquisition is limited to a brief time window (McClelland, McNaughton, & O'Reilly, 1995; Tse et al., 2007). This pattern of developmental change also explains why the hippocampus is not among the regions reliably activated during arithmetic (see, e.g., meta-analysis by Arsalidou & Taylor, 2011). Even more interestingly, in the Qin et al. (2014) study, children's gains in fact-retrieval fluency scores were positively correlated with the strength of hippocampal connectivity with several prefrontal and parietal cortical areas, including left and right DLPFC and left IPS. In a different investigation, increased intrinsic connectivity of the hippocampus with parts of the neocortex predicted performance gains after 8 weeks of maths tutoring in children within the same (7-9) age range (Supekar et al., 2013). Overall, these results suggest that it is the changes in hippocampal-neocortical functional circuits, rather than the levels of hippocampal activation per se, that underlie individual gains in memory-based problem solving.

4.2.2 Evidence of plasticity from longitudinal correlates of training

The first chapter in my thesis contains a review of the evidence from cross-sectional studies using structural MRI favouring the argument that neuroplasticity still occurs in the adult human brain. The chapter

also discusses some of the limitations of the cross-sectional design, and how the longitudinal MRI design – with scans at multiple time points – can help mitigate those limitations. The next section discusses more detailed evidence of brain plasticity coming from studies that have trained people in various cognitive tasks and have related the pre-to-post-training brain changes to measures of behavioural performance. More than just *longitudinal* studies – a term that simply implies the existence of two or more assessment (time) points – these are in fact *training* studies that make use of the longitudinal design.

Studies seeking to identify structural changes in the brain as a result of cognitive or motor training will typically compare structural scans acquired after training (and sometimes a period of time subsequent to the end of training) to a baseline scan acquired before the start of training. The structural parameters identified have to do with either the density or thickness of the cortex or to the integrity or connectivity of the underlying fibre bundles. One important parameter that such studies are trying to determine is the exact time-scale of usage-dependant structural changes occurring in grey or white matter structures. For instance, increased grey matter density (GMD) in the visual motion area bilaterally was found after people had learned to juggle over a 3-month period (Draganski et al., 2004). The same group, however, later replicated their own findings and suggested that these changes are apparent after as little as 7 days of training (Driemeyer, Boyke, Gaser, Büchel, & May, 2008). Experience-dependent structural changes can, however, theoretically also occur on an even shorter time scale, e.g. hours (Feldman, 2009). Selective elimination of synapses, leading macroscopically to cortical thinning, therefore underlies not just development and learning but also day-to-day usage-dependent plasticity.

Subsequent to those earlier studies, structural changes have been identified as a result of training periods of various durations. Draganski et al. (2006) made use, as a training paradigm, of the intense 3 months period during which German medical students study for their preliminary medical exams. The control group consisted of people scanned at the same three time points (before and after the study period, and 3 months later), but who were not studying for any exams. When comparing the post- to the pre-study scans, an increase in GMD was found in the posterior and inferior parietal cortices bilaterally, as well as in the posterior hippocampus. Moreover, GMD in the posterior hippocampus continued to increase 3 months after the study period had ended. Takeuchi et al. (2011) subjected adults to high-intensity training of working memory for five days, using mental calculations. The training was adaptive in that task difficulty was modulated to adapt to the performance of subjects, while an active control group received non-adaptive

training, and a passive control group did not receive training. As compared to the combined control groups, the adaptive training group showed a decrease in the GMD of the bilateral DLPFC, bilateral parietal regions and left superior temporal gyrus, areas critical in working memory. As expected, the training improved letter span and complex arithmetic ability. Performance change in the former task correlated negatively with change in GMD in the left superior temporal regions, suggesting that people who improved the most did so at the expense of a loss in GMD in this area. Other training studies have also reported alterations in other cortical metrics such as cortical thickness (CT). For instance, Engvig et al. (2010) asked middle-aged and elderly healthy subjects to undergo an intensive 8-week memory training program. Subsequently, CT in right insula was increased in the training group compared to a no-training control group. Moreover, CT changes in the right fusiform and lateral orbitofrontal cortex correlated positively with improvement in memory performance, independent of initial performance.

Evidence of changes in white matter (WM) structure as a result of experience dependent plasticity started to appear about a decade after this evidence was found for grey matter. Building on the juggling-training paradigm first used by the Draganski et al. group (2004; Driemeyer et al., 2008), Scholz et al. (2009) showed that after a 6-weeks period of juggling training, structural changes occur not only in terms of GMD (in the medial occipital and parietal lobe), but also in terms of fractional anisotropy (FA) of the WM underlying these cortical regions. However, no correlations were found between the change magnitude in grey- and white matter across subjects, despite the close spatial proximity of these regions. Given the different maturational trajectories of the two types of tissue, the authors suggested that the data reflect that structural experience-dependent changes in the grey and white matter of adult brain represent relatively independent processes. The same group later tested whether individual differences in GMD at baseline can predict performance subsequent to a similar 6-week juggling training period (Sampaio-Baptista et al., 2014). Training intensity was manipulated between-groups: half of the subjects received training for 15 minutes daily while the other half received 30 minutes daily. For both groups, a steeper learning slope was associated with greater GMD in medial occipito-parietal areas at baseline. Moreover, subjects with better performance at the end of the training had higher increases in GMD in dorsal parietal cortex and M1, when assessed after another 4 weeks. Also, the intensity of training modulated the relationship between performance and GM volume change in dorsolateral prefrontal cortex.

Using a shorter (5-day) training on a temporal discrimination task, Bueti et al. (2012) found that both GMD and FA in the right cerebellar hemisphere increased with training. These changes correlated with behavioural indexes of learning across subjects, while GMD in post- and precentral regions at baseline (before training) predicted individual learning abilities as indexed after training. Decreasing training duration even further, Sagi et al. (2012) examined human subjects before and after a spatial learning and memory task. MD (mean diffusivity), a measure of the average diffusion in a voxel, was computed across the entire brain. MD changes in limbic structures (hippocampus and parahippocampal gyrus) became significant after only two hours of training. These findings provide an important indication about the order of magnitude of the time course over which structural change is detectable with DTI following training. Lending credibility to these results, and by extension to previous DTI studies of neuroplasticity, is that similar changes in the same limbic structures have been detected at the histological level, in a control experiment with rats, as part of the same study (Sagi et al., 2012).

In summary, these results suggest that plasticity in the brain can occur reliably, at the level of both grey and white matter, after only a few days, or even a few hours, of intensive training in either motor or cognitive skills. Given the wealth of possible domains of training, and researchers' desire to study brain plasticity across an increasingly wide range of these domains, it follows that a limitation of these studies is the rare degree of replication existing in this literature. For instance, it is difficult to predict to what degree the specific patterns of GMD modifications found after training in domain A would overlap and differ from those found after training in an adjacent domain B. To my knowledge, the current study is the first that examines the structural changes that occur in the brain as a result of arithmetic training, undertaken daily for five days.

4.2.3 Using brain stimulation to affect maths cognition via consolidation (offline) effects

The General Introduction of this thesis (Chapter 1), as well the previous empirical chapter (Study 2 of Chapter 3) contain more general background information on the tES method and how it can and has been used to enhance cognition in general and numerical cognition in particular. In this section, I focus only on aspects that are particular to the current study – specifically, on the distinction between online (within-session) and offline (between-session consolidation) effects.

The excitability changes induced by tES resemble neuroplastic alterations of cortical function that are thought to naturally accompany the cognitive processes of learning and memory formation (Kuo & Nitsche,

2012). This similarity, and the fact that neuroplastic alterations of cerebral connectivity are thought to be the major physiological contributor to learning and memory formation, makes plasticity modulation with tES a suitable candidate for enhancing such cognitive processes (Krause & Cohen Kadosh, 2013). When using non-invasive brain stimulation to enhance cognition, one salient distinction is between stimulating online (concurrently with the task) or offline (e.g. priming the brain at rest before the task to be performed). Facilitatory effects on, for instance, motor learning, are more reliably being found by online studies (e.g., Antal et al., 2004) in comparison with offline procedures (e.g., Kuo et al., 2008; Stagg et al., 2011). In fact, in some cases, stimulation delivered offline, prior to task performance, has even worsened performance (Antal, Begemeier, Nitsche, & Paulus, 2008). In numerical cognition, facilitatory effects have been found using both online (Cohen Kadosh et al., 2010; Snowball et al., 2013) and offline stimulation (Hauser, Grabner, & Jäncke, 2013) with tDCS or tRNS. Interestingly, however, even when delivering tDCS stimulation online with the task, improvement in performance was still found – in one study – to occur offline, i.e. between consecutive sessions (Reis et al., 2009). This suggests the existence of a consolidation mechanism in the brain, which can be targeted with transcranial electrical brain stimulation, and which contributes to offline but not to online effects. The current study was carried out with this distinction in mind. To the best of my knowledge, so far no other studies have investigated the structural changes taking place in the brain (in terms of grey and white matter) after repeated applications of tRNS. This study thus has the potential to enhance our understanding of the mechanisms of tRNS by observing structural changes in the brain longitudinally.

4.2.4 The current study

Study 2 (Chapter 3) leaves outstanding several questions:

1. Assuming stimulation is only done to one brain site for the entire duration of training, is it the prefrontal or the parietal site that is more optimal in the sense of e.g. an increased learning rate? A previous study (Snowball et al., 2013) has examined effects of tRNS stimulation to both sites and found no effect following parietal stimulation. However, this might be due to small sample size (N_{tRNS}=14, N_{sham}=12). In addition, that study examined learning rate but not also other potentially relevant measures such as consolidation (Reis et al., 2009).
2. By virtue of which structural brain changes do the behavioural improvements of tRNS-assisted cognitive training come to be?

3. Which structural parameters of the brain at baseline (or their percentage change pre-to-post training) are able to predict a subject's behavioural performance?

The current study was performed in an attempt to complement Study 2 and thus answer these unanswered questions. I aimed to accomplish this by combining a longitudinal design that involved structural and diffusion MRI scans before and after the training period. The behavioural paradigm of drill- and calculation-based arithmetic learning that was used during training was the same as in Study 2, with the exception that difficulty was no longer manipulated. Instead, the "easy" level was used for the Drill task, and the "difficult" level was used for Calculation; this was based on the Study 2's observation of poor retention, during the testing phase, of results for trained problems in Drill's difficult condition, and non-significant effects in Calculation's easy condition. Also, the same prefrontal and parietal locations were used for tRNS stimulation, however each site was assigned to one group for the entire duration of training as opposed to shifting them mid-training, as was the case in Study 2.

This study also complements Study 1 (Chapter 2) from an inference and experimental design point of view. While no causal claims linking brain structure and maths performance could be made in that study when comparing mathematicians and non-mathematicians cross-sectionally, here the mixed-model design, with stimulation-defined groups and longitudinal scans, theoretically allows direct links to be drawn between the brain-level manipulation and the changes observed longitudinally.

As discussed elsewhere in the thesis, the training literature has very mixed results with regards to the direction of structural brain changes. For instance, several studies report both increases (e.g., Bueti et al., 2012; Sampaio-Baptista et al., 2014) and decreases (e.g. Takeuchi et al., 2011) in cortical metrics as a result of training programmes lasting in the order of days-to-weeks. For these reasons, there were no particular predictions in the current study in terms of effect direction of brain changes.

4.3. Methods

4.3.1 Subjects

Seventy-two right-handed human volunteers were recruited by means of adverts placed around the University of Oxford. The sample included 36 males and 36 females, with ages 21.92±2.93 years.

4.3.2 MRI acquisition

All subjects received structural and diffusion MRI scans 1-to-3 days before the start of the training (the "pre" scan) and on the last day of training, subsequent to the end of the final session (the "post" scan). A 3T system (Siemens Magnetom Verio Syngo MR B17, Siemens Medical Solutions, Erlangen, Germany) was used to acquire T1-weighted anatomical images for each subject, in the axial plane, using a 3D MPRAX sequence (TR = 2040 ms, TE = 4.7 ms, flip angle = 8°, FoV = 192 mm, voxel size = 1 × 1 × 1 mm, 192 slices, slice thickness = 1 mm). Diffusion weighted spin-echo EPI (TR = 9600 ms, TE = 87 ms, FoV = 190 mm, maximum b factor = 1000 s/mm^2 , voxel size = 2 × 2 × 2 mm, 65 slices, slice thickness = 2 mm) was also acquired, by collecting eight images with no diffusion weighting (b_0) and 60 images with diffusion gradients applied in 60 noncollinear and equidistant (isotropically distributed) directions.

4.3.3 Behavioural tasks

4.3.3.1 Baseline tasks

During the first experimental session, before the actual training task started, all subjects were administered a number of tasks, in order to have a baseline measure of their mathematical and approximate number system (ANS) abilities. To assess the former, the Numerical Operations and Mathematical Reasoning components of the Wechsler Individual Achievement Test (Wechsler, 2005) were used. The Numerical Operations test requires subjects to solve written calculation problems and simple equations involving the four basic operations of addition, subtraction, multiplication and division, and also some more advanced ones such as powers and square roots. Mathematical Reasoning tests subjects orally on problems with whole and decimal numbers, interpreting graphs and identifying mathematical patterns.

In order to measure the acuity of the ANS, the dot enumeration task (also known as the number acuity task) was used. As opposed to Study 1 (Chapter 2), which for the same purpose used an in-house paradigm with dot arrays spatially separated on the left and right of the screen, the paradigm used here presented a single dot array, with intermixed blue and yellow dots, as illustrated in Figure 24. The task is similar to the one described in (Halberda, Ly, Wilmer, Naiman, & Germine, 2012). In short, the numerosity ratio of the two colours is varied across trials in accordance to a pre-set order; 300 trials are presented in total, and at each trial subjects had to decide whether there were more blue or more yellow dots displayed, by pressing the corresponding button on a PC keyboard.

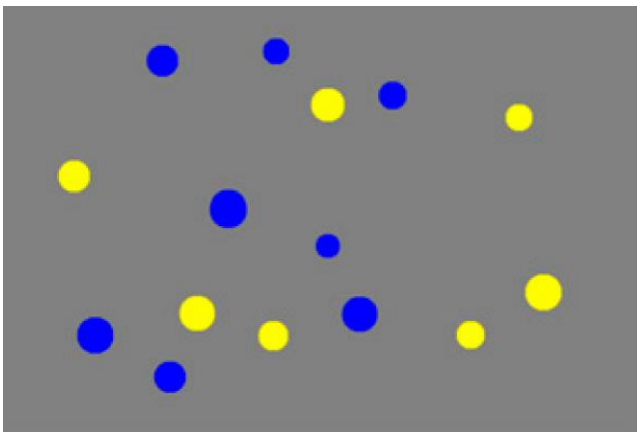


Figure 24: Example of a trial from the number acuity (dot enumeration task). Reproduced from (Halberda et al., 2012).

4.3.3.2 Control tasks

Subjects were administered a number of control tasks before the first and after the final (fifth) experimental session, in order to examine the specificity of brain stimulation. Specifically, I used control tasks that measured attention and verbal and visuo-spatial working memory, to check whether these more general abilities are also affected by brain stimulation, in addition to the targeted numerical abilities. The reason why I chose these particular general abilities is that they are the ones most intimately linked with arithmetic, and at the same time their subserving brain regions largely coincide with the frontal and parietal sites that were stimulated. Specifically, as detailed in the General Introduction (section 1.3.1, page 5), maths performance in general is heavily reliant on the working memory functions of the prefrontal and parietal cortices (Olesen et al., 2004), while these same cortices are believed to support all three attentional networks described by Posner's dominant model (Fan et al., 2005). Scores on these tasks are therefore likely to illustrate either cross-domain transfer of enhancement, or, on the contrary, perhaps a hidden cognitive cost of tES-induced enhancement (Iuculano & Cohen Kadosh, 2013).

Attention and verbal working memory were tested using the attentional networks task (ANT) and the (forward and backward) digit span task respectively; these tasks are described in detail in Study 2's Methods (section 3.3, page 62). In order to obtain a measure of visuo-spatial working memory, the Corsi block-tapping task was administered (Milner, 1971; Orsini et al., 1987). The task makes use of a flat panel on which nine identical and unmarked black cubical blocks are placed, in an irregular spatial order. During the task, the experimenter points to a number of these blocks in a particular predefined order, held constant across subjects. The subject is asked to reproduce this order by pointing to the blocks in the same way. The length of

the sequence starts from 2 and increases up to a maximum of 9 blocks, with 2 trials for each sequence length. A point is awarded for every correctly reproduced trial. As with the digit span task, the Corsi block-tapping task is ended if subjects incorrectly answer both trials from a given length, or when both trials corresponding to a length of nine has been reached. The task was first done forwards, as explained above, and then backwards, where the subject was asked to tap the blocks in reverse order to that indicated by the experimenter.

In addition to these domain-general abilities, I also used a symbolic number mapping paradigm in order to assess the accuracy of number line mapping before and after training. The task was programmed in Matlab and was based on the one by Siegler & Opfer (2003). Subjects were asked to indicate, using a PC mouse, the position of 30 integers ranging between -1000 and 1000 on a blank number line that only had its end-range values labelled. Subjects had unlimited time for each trial, and were asked to estimate the position of each number on the line visually, rather than by using any external references. The score on this task was considered to be the average absolute deviation between plotted (subjective) location and actual (objective) location, across all trials.

4.3.3.3 Training task

The arithmetic training paradigm used in this study was similar to the one used in Study 2 (and based on Delazer et al., 2005). Briefly, the task involved solving alternating blocks of "Calculation" and "Drill" problems. For Calculation, a given algorithm (such as "2×R+L-10") had to be applied to pairs of operands (such as L=3 and R=15), while Drill problems had an fixed arbitrary result that had to be memorised and later retrieved. The only difference from Study 2 had to do with manipulation of task difficulty. In Study 2, the amount of material trained (i.e. the number of distinct pairs of operands used for each type of problem throughout the training) was varied between subjects: half of the sample were trained using 6 operand pairs for Calculation and another 6 operand pairs for Drill (the "easy" condition), while for the other half 12 operand pairs were used for each type of problem. Analysis of the testing stage data from Study 2 (see section 3.3.6, page 67) revealed that a large number of subjects (11 out of 32) were not performing the Drill task according to the instructions, specifically that they were not, in fact, committing the displayed result to memory during the training stage. This prompted me to exclude those subjects from the Drill analyses in that study. Since the majority of these subjects (8 out of 11) were in the "difficult" condition (i.e. which implied a higher number of both Drill and Calculation problems), I reasoned, for the current study, that it is optimal to

only have the Drill task "easy" in the sense of only making use 6 operand pairs during training. The Calculation task, on the other hand, continued to use 12 operand pairs (as in Study 2's "difficult" condition), as that is when the effects in that study were found. All other parameters of the task were identical to those described in the Methods section of Study 2.

It is of theoretical interest to note that, in general, training paradigms can be designed with the aim of attaining constancy of either the total training duration or of total subject performance. The former model implies a fixed duration of training for all subjects, which inevitably, given inter-individual differences, leads to varying levels of performance by the end of training. On the contrary, the latter model aims to persist with training until a certain level of performance has been reached by the end of the training. For the same reason, this inevitably leads to varying training durations across all subjects. The latter model presents a clear advantage in cases (such as the current one) where behavioural training is accompanied by pre- and post-training brain imaging, since any brain-behaviour correlations that are found can rely on the presence of a uniform behavioural performance level. However, this training model also presents logistical difficulties, as each subject's training time becomes necessarily variable, and potentially endless. Another logistical difficulty of this model is the fact that, in order to truly control for the amount of behavioural gain achieved, not only the final but also the initial performance levels would have to be matched, which makes the model even more difficult to implement practically. For these reasons, the more traditional "training constancy" model was selected.

4.3.4 Brain stimulation

Subjects were assigned to one of three stimulation groups pseudo-randomly, aiming to minimise the gender and age misbalance between groups. Equal numbers of subjects were included in each group (24 in each). Two of the groups received bilateral tRNS stimulation, to the prefrontal cortex (10-20 EEG positions F3 and F4; "frontal" group) and to the posterior parietal cortex (10-20 EEG positions P3 and P4; "parietal" group) respectively. A third control group received sham stimulation as explained below. For the sham group, electrodes were placed on positions F3-F4 for half of the subjects in the group, and on P3-P4 for the other half. The main experimenter received instructions about the positioning of the electrodes for each subject without knowing which stimulation group they were in, and the double-blind was, just as in Study 2, achieved by having a second experimenter start the stimulator in the relevant mode, i.e. with either real or sham stimulation. Current in the form of high frequency noise (100-640 Hz) was delivered by a battery-driven

current stimulator (DC Stimulator Plus, from Magstim, UK). The current intensity was 1mA peak-to-peak, with each sample being drawn from a normal distribution with mean 0 μA, and with 99% of all generated amplitude values lying between -500 μA and +500 μA. The stimulation duration was set to 20 minutes for the real stimulation groups (frontal and parietal) and to 30 seconds for the sham group, both flanked by a 15 second upward and downward ramp. The stimulator was always started at the same time as the onset of the task and was turned off when the task ended. The current was delivered through two 4x4 cm electrodes, which were secured in place using an elastic strap and were placed in saline-soaked sponges to improve electrical contact with the scalp and reduce the risk of skin irritation. All subjects were asked, at the end of the experiment, whether they thought they were in the real stimulation or in the sham ("placebo") group, judging by the sensation that they felt underneath the electrodes during the stimulation.

4.3.5 Behavioural analyses

As in Study 2, a number of subjects had inordinately low accuracy scores for old drill problems in the testing phase, suggesting that, during training, they had not learnt the associations between operand pairs and their corresponding result, as instructed, instead just "transcribing" the isolated result. The "guessing range" was defined as 2.5 SDs around the mean accuracy of new drill problems, which resulted in the 0.22 mean accuracy value as a cut-off for defining outliers for old drill problems. This led to 5 subjects (i.e. roughly 7% out of the total of 72) being excluded from all analyses of drill problems, for both training and testing; 3 of these were from the Frontal group, 1 from the Parietal group and 1 from the Sham group.

Data from the training paradigm were analysed, in the first instance, in the same way as was the case in Study 2. In short, mean reaction times (RTs) for correct trials, as well as mean accuracy, were analysed using mixed-model ANOVA, with stimulation group (frontal, parietal, and sham) as a between-subjects factor, and session (1 through 5) and block (1 through 5) as within-subjects factors. Learning rate for Calculation was modelled in the same way, assuming skill acquisition evolves according to a power law (Newell & Rosenbloom, 1981). The distribution of RTs for correct Calculation and Drill problems was again modelled using an ex-Gaussian function, which produced estimates of the parameters describing this curve – μ, σ and τ – for every subject, problem type and session.

In addition, online and offline effects were computed within and between sessions respectively, for both RTs and accuracy (Reis et al., 2009). Specifically, online effects were defined as the difference between

the mean of the measure of interest (RT or accuracy) in the final block of any given session and the mean of the first block of the same session (e.g. Session1Block5–Session1Block1). Thus, a negative online effect in terms of RT would indicate an improvement in speed from the start to the end of the session, whereas an improvement in accuracy would correspond to a positive online effect for accuracy. Given the 5-session training length, 5 online effects can be defined. By contrast, offline effects for both RTs and accuracy were defined as the difference between the first block of a session and the last block of the previous session (e.g. Session2Block1–Session1Block5). Four offline effects were thus defined, as the time lapses between consecutive daily sessions (and were thus referred to as "nights"[13] in the factorial design with which offline effects were analysed). The effect direction has similar interpretation to online effects, i.e. improvement corresponds to negative offline effects for RT but positive offline effects for accuracy. Total learning in tRNS-assisted training can thus, according to this theoretical distinction, be broken down as the sum of online and offline effects.

4.3.6 MRI analyses

4.3.6.1 Grey matter analyses

4.3.6.1.1 VBM

VBM was used in order to compute and compare regional concentrations of grey matter, i.e. grey matter density (GMD). This method was implemented in FSL and details of its pre-processing steps are included in section 1.5.2.1.1, on page 17. Given the longitudinal aspect of the current data set, in a first stage analysis, an image was computed as the difference between the pre-processed image of voxel intensities for GM tissue corresponding to the post scan minus that of the pre scan. It was this image that was then used as dependent variable in a second-level analysis, where voxel-wise statistics were carried out cross-sectionally using a general linear model (GLM) implemented in FSL. The design matrices that were used aimed to identify regions where GMD changed significantly from pre-to-post within each stimulation group, where this change differed between groups and where it correlated with behavioural measures. Predictors coding for group membership (stimulation group frontal/parietal/sham), as well as for the extraneous variables age and

[13] NB: the term *night* was simply chosen as an intuitive shorthand to refer to the repeated measures factor in the offline analyses: thus, the *session* factor in online analyses runs from 1 to 5 and the *night* factor runs in-between sessions, i.e. from 1 to 4. Usage of the term does not imply that offline effects necessarily have to occur during the night (e.g. during sleep); instead these effects are simply defined as occuring between the end of one stimulation session and the start of the next one (which is typically on the next day).

gender, were included in all designs. Additionally, correlation designs also included behavioural scores, specifically ones relating to baseline maths ability (Numerical Operations) and to learning performance during the training period (e.g., offline RT effect).

4.3.6.1.2 SBM

As in the case of Study 1, SBM implemented in FreeSurfer was used to complement the VBM analyses of grey matter data. The FreeSurfer processing pipeline was similar to the one described in the Methods section of Study 1 (section 2.3.4.1.2, page 32), with the exception that the analysis of the longitudinal data of the current study required several additional steps, as per FreeSurfer's longitudinal processing stream, described in detail elsewhere (Reuter, Schmansky, Rosas, & Fischl, 2012). In short, this processing stream consists of three steps:

1. **Cross**: 3D surfaces are initially generated for all T1-weighted images, i.e. independently for both pre- and post-training scans from all subjects, using the regular FreeSurfer processing stream (recon-all).
2. **Base**: a unique template (base) is created for each subject, representing the average anatomy of each subject across the two time point scans.
3. **Long**: each subject's time points are processed longitudinally i.e. by making use of common information taken from the subject-specific template. The generated folders are subsequently used for all post-processing and analyses, as the results are more sensitive and repeatable than the independent cross runs (Reuter & Fischl, 2011).

Statistics were performed on the processed images using FreeSurfer's *longitudinal two stage model*. This procedure first reduces the temporal aspect of the data within each subject to a single statistic, which represents the percentage change from the first to the second scan, i.e. (post-pre)/pre. In a second step, this measure, referred to as "pc1", was compared across groups or correlated with a behavioural measure. The use of this percentual measure that is normalised within-subject ensures that any group differences that are found cannot be explained by pre-existing differences between groups.

4.3.6.2 White matter analyses

4.3.6.2.1 TBSS

Tract-based spatial statistics (TBSS) were used in order to compute and compare fractional anisotropy (FA) across subjects. General details of the pre-processing steps of TBSS can be found in the General Introduction (section 1.5.2.2.2, page 21), while more technical details relating to its FSL implementation can be found in Study 1's Methods (section 2.3.4.2.1, page 33). As with VBM, the statistics were performed on a version of the final pre-processed image (skeleton) that was obtained as a difference between the post and

the pre scan. The GLM designs used to check for group differences and for brain-behaviour correlations were the same as those used in the case of VBM.

4.3.6.2.2 Tractography

Given the bilateral nature of the tRNS stimulation that I used (to F3-F4 in the frontal group and to P3-P4 in the parietal group), the most suitable candidates for fibre tracts to investigate seemed to be the homotopic tracts linking these cortices (left-to-right DLPFC for F3-F4 stimulation and left-to-right posterior parietal for P3-P4 stimulation). Homotopic tracts have been suggested to display a high degree of variability between regions and between individuals, in both human and monkey studies (e.g., Putnam et al., 2009; Rouiller et al., 1994). Also, Study 1 found that the tract connecting rSFG to lSFG was stronger in mathematicians than in controls. For this reason, this tract was also investigated in the current study.

In order to isolate the rSFG to lSFG tract, the masks obtained in Study 1 were used. In order to isolate the F3-F4 and P3-P4 homotopic tracts, seed masks were defined based on the positions underlying these electrodes, as explained in section 4.3.6.3.1 (last paragraph just before Table 10). An anatomical mask of the corpus callosum (drawn from the Juelich Atlas) was used as a first waypoint, in order to ensure that the homotopic fibres aimed to be tracked are indeed transcallosal. The contralateral ROI was then used as the second waypoint, in order to find tracts connecting the two electrode-driven ROIs. In order to avoid picking up salient ipsihemispheric fibres (i.e. ones connecting cortices within a single hemisphere, such as for instance the inferior longitudinal fasciculus), exclusion masks were also defined. The exclusion masks used by the tracking software acted to guide the tracked fibres away from an area that is to be avoided. For each tract, the exclusion masks consisted of two coronal planes that delimited the anterior-posterior portion believed to contain most of the transcallosal fibres linking the two cortices of interest. Each plane was defined as having a width of 3 slices, in order to avoid rounding operations during processing that might possibly reduce their width to zero. I based my choice of a location for these planes on a previous study providing a probabilistic atlas for locations of transcallosal (homotopic) fibres (Styner, Oguz, Smith, Cascio, & Jomier, 2005). Thus, the frontal exclusion mask consisted of coronal planes y=18 to 20 and y=63 to 65; and the parietal exclusion mask consisted of coronal planes y=-9 to -7 and y=73 to 75. These exclusion masks, together with the electrode-based (F3, F4, P3, P4) and anatomically-based (corpus callosum) waypoint masks are overlaid on a standard brain template in Figure 25.

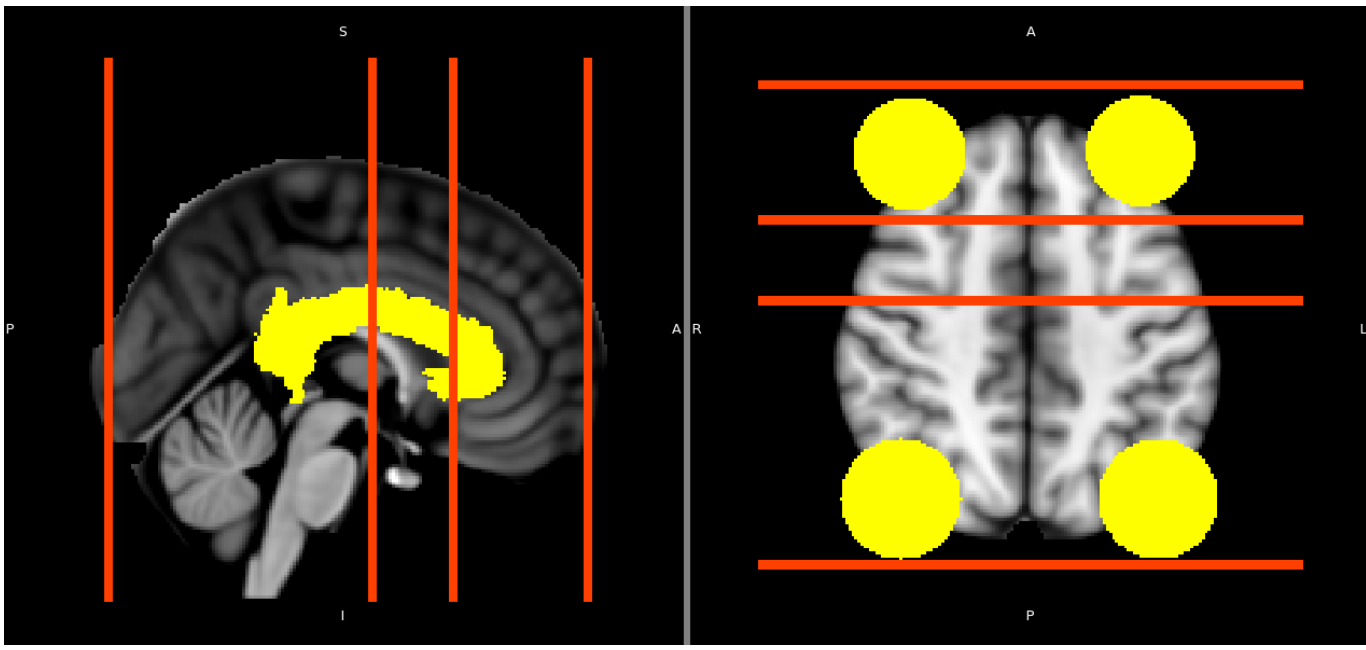


Figure 25: Masks used to delineate homotopical tracts. These tracts connect ROIs underlying the stimulating electrode positions (F3, F4; P3, P4). Waypoint masks (electrode-based and anatomically-based) are displayed in yellow, while the planes defining the exclusion masks are displayed in red. Left panel: sagittal view. Right panel: axial view.

Finally, using these parameters, tractography was performed by running the ProbTrackX program in FSL. The probability distribution in the produced tracts was then log-transformed, normalised by the maximum value in the distribution, thresholded and binarised to form a mask that was used to multiply each subject's FA image, in order to obtain the mean FA value along that specific tract, for each subject. These values were then used in a mixed-model Session (pre / post) × Tract (F3-F4 / P3-P4 / lSFG-rSFG) × Group (frontal / parietal / sham) ANOVA.

4.3.6.3 ROI analyses

4.3.6.3.1 ROI selection

Regardless of the results of a whole-brain analysis, further analyses based on predefined, hypothesis-driven regions of interest (ROI analyses) can potentially reveal new layers of interpretation to an MRI data set. In the case of the current study, I have chosen ROIs based on two lines of argument.

The first category of ROIs was defined based on the location of the stimulating electrodes. Current density modelling can predict with reasonable accuracy the flow and local field intensity of the electrical current generated in the brain by non-invasive stimulation methods such as tDCS (e.g., Datta et al., 2009). Such modelling research has shown that, while the peak of the electric field occurs at a location close to the mid-point between the two stimulating electrodes, the cortex immediately underlying the electrode location

still receives a substantial amount of electrical charge, and more than, for instance, locations in the head that are more distal to the electrodes (Marom Bikson, personal communication). On the other hand, a previous tRNS study that investigated changes in oxygenated (HbO_2) and deoxygenated (HHb) haemoglobin using near-infrared spectroscopy found the peak amplitudes of HbO_2 concentrations in the left DLPFC, immediately underlying the stimulating F3 electrode, which were smaller in the tRNS group than in sham controls (Snowball et al., 2013). Also, previous combined tDCS-fMRI studies have found signal change to be close to the stimulation site (Kwon et al., 2008; Stagg et al., 2009). Also, while non-invasive brain stimulation methods in general are believed to affect connectivity (Dayan, Censor, Buch, Sandrini, & Cohen, 2013), and thus cannot be considered truly focal, the focality in the accuracy of their application to a targeted site does vary, this accuracy being higher in, for instance, transcranial magnetic stimulation (TMS) than in tES. For these reasons, the cortical regions immediately underlying the stimulating electrodes were selected, for the current study, as nominal target ROIs.

Since the stimulation groups involved prefrontal (F3 and F4) and posterior parietal (P3 and P4) stimulation, I created ROIs that corresponded to the area of cortex underlying these electrodes that would be immediately impacted by the current flow. The coordinates, in standard MNI (Montreal Neurological Institute) space, corresponding to the four electrode positions, were calculated directly on the MNI head reconstruction, using the algorithm described in (Jurcak, Okamoto, Singh, & Dan, 2005). In order to obtain the MNI coordinates at the level of the cortex, head and brain 3D surfaces (meshes) were used to project the scalp coordinates following the perpendicular line to the tangential plane of the scalp at each electrode site (Ruth Caballero Hernández, personal communication). This led to the cortex-level coordinates described in Table 10.

The second category of ROIs are based on results from Study 1, which found structural differences in the brains of mathematicians vs non-mathematicians cross-sectionally. Defining this additional, hypothesis-based, category of ROIs was inspired by Zamarian et al.'s urge, in their review on imaging studies of arithmetic learning (2009), for training studies to *"investigate, through morphometric analysis, whether intensive, but short-term practice results in specific structural modifications of the brain areas involved in numeric processing, as was found for expert individuals with extensive mathematical experience over several years (Aydin et al., 2007)"*. Since in my own study on structural correlates of maths expertise (Study 1, Chapter 2), I found GMD group differences in three regions (left IFG, right IPS, right SPL), as well as GMD-

behaviour correlations in left SFG, these four regions (whose centre coordinates are also given in Table 10) were also defined as ROIs, in the manner explained below. The argument was that, if structural changes can be detected in these regions (presumably) as a result of years of maths training, then it might be possible that a (much shorter) training might affect brain structure in those same regions, and that hopefully the current MRI methodology would be sensitive enough to pick up on them. While in Study 1, reasonable doubt had to be placed on those clusters having been altered between expertise groups (mathematicians vs non-mathematicians) by that very expertise, the longitudinal design here would allow me more certainty in believing that any observed changes (in these clusters or elsewhere) are due to the tRNS intervention.

Table 10: MNI coordinates corresponding to the two ROI categories used. For electrode ROIs, coordinates represent estimates of the orthogonal projection from the centre of the stimulating electrodes to the level of the underlying cortex. For the ROIs based on results of Study 1 (Chapter 2), coordinates represent peak clusters from that study's VBM analyses.

ROI category	ROI	Centre coordinates (MNI, mm)		
		x	*y*	*z*
Electrode locations	**F3**	-39.97	41.55	39.46
	F4	37.8	42.32	38.35
	P3	-43	-74	46
	P4	43	-73.9	47.56
Mathematicians study	**lIFG**	-42	6	30
	rIPS	44	-38	46
	rSPL	34	-62	48
	lSFG	-42	6	30

Once the centre coordinates were ascertained, ROIs were defined as binary spherical masks of a radius proportional to the size of the stimulating electrodes (40x40 mm). For grey matter analyses, a sphere radius of 10mm was selected (cf e.g. Keeser et al., 2011). For white matter analyses, the spheres needed to be larger, in order to intersect to a sufficient degree the white matter mask and the FA skeleton mask (TBSS and tractography analyses, respectively). The masks representing the ROIs were multiplied by the masks corresponding to the tissue type involved in the analysis (GM or WM) in order to eliminate voxels not belonging in these tissue types, and finally registered to the relevant space (study-specific template for VBM, and diffusion space for TBSS). These final masks were then used to extract GMD and FA values (respectively) from the difference (post-pre) image, and those values were then fed into groupwise statistics.

An illustration of the relative positioning of these ROIs, and of the degree to which they intersected the relevant tissue type (GM/WM) is shown in Figure 26.

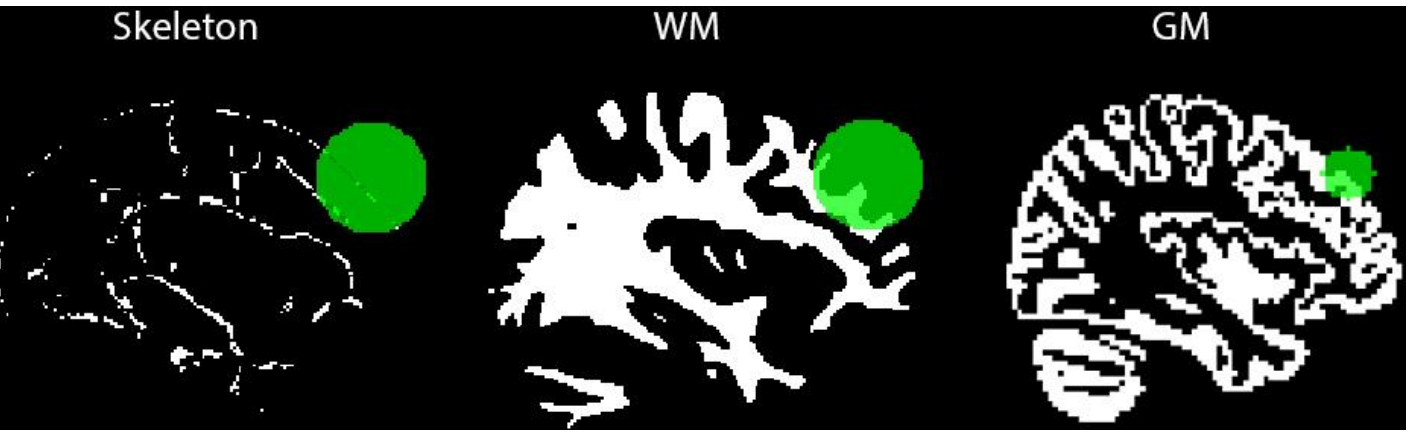


Figure 26: An example of a spherical mask used to define a region of interest (ROI) corresponding to one of the four locations of stimulating electrodes (here, the one for position F3 is shown). The masks are shown overlapped on top of the masks corresponding to the brain measure of interest, i.e. skeletonised (left panel) or non-skeletonised white matter (WM, middle panel), and grey matter (GM, right panel). The sphere was larger (radius=20mm) when WM had to be included in the ROI, and smaller (radius=10mm) when GM had to be included.

Additionally, an extra ROI was defined based on Tsang et al.'s (2009) finding of FA values in one section of the superior longitudinal fasciculus (SLF) correlating with arithmetic performance. The ROI consisted of the same coronal planes found in the said study, i.e. the middle of the anterior portion of the SLF, from Y=-21 to Y=-27.

4.3.6.3.2 Analyses

The extracted GMD and FA values were submitted to mixed-model ANOVAs. Stimulation group (frontal/parietal/sham) was always held as the between-subjects factor. For the electrode-based ROIs, the within-subject factors were site (frontal/parietal) and hemisphere (left/right). For the hypothesis-based ROIs, the within-subject factor was simply site (lIFG/rIPS/rSPL/lSFG). Additionally, for FA values extracted from the aSLF, correlations were checked with arithmetic performance as indexed by score on the WIAT components Numerical Operations and Mathematical Reasoning.

4.4. Results

4.4.1 Behavioural results

The behavioural analyses were largely similar to those performed on the data of Study 2, which used a similar behavioural paradigm as the one used here (see the Methods section of the current study for more details).

4.4.1.1 Baseline tasks

A summary of scores for the baseline tasks, categorised by stimulation group, is given in Table 11. In terms of scores on the WIAT components Numerical Operations and Mathematical Reasoning, the main effect of group was not significant, although it did approach significance ($p=.10$ and $p=.07$ respectively). Planned comparisons revealed, for Mathematical Reasoning, there was a significantly lower score for the parietal than for the frontal group ($p<.05$) and marginally significantly in comparison to the sham group ($p=.059$). Similarly, the parietal group had lower scores than the sham group on Numerical Operations ($p<.05$).

Accuracy scores from the dot enumeration task were analysed in the same way as the similar data from Study 1 (see end of section 2.3.2), and led to the computation of a Weber fraction (*w*) for each subject. For three subjects, the *w* values were greater than 1, suggesting non-standard strategies or random guessing during the task (Darko Odic, personal communication). These cases were discarded, and data for the remaining 45 subjects for which dot enumeration data was collected – of which 14 in the frontal, 13 in the parietal and 18 in the sham group – was analysed using a group ANOVA; no significant main effect of group was found ($p=.4$).

4.4.1.2 Control tasks

A summary of the pre- and post-training scores for control tasks, categorised by stimulation group, is given in Table 11. Scores from the ANT task (separately for the three attentional networks), from the number line task and from the digit span and Corsi block-tapping tasks (both forwards and backwards) were submitted to mixed-model group (frontal/parietal/sham) × timepoint (pre/post-training) ANOVAs. No significant main effects of group, nor interactions with timepoint, were found. Additionally, there was no main effect of group when analysing only the pre-training data (all $ps>.2$), suggesting that there were no initial group differences in terms of performance on these control tasks.

Table 11: Descriptive scores for baseline and control tasks. The latter are reported at both the pre-training as well as the post-training stage. Variance for Weber fraction within group F is very large due to subjects in that group who performed at chance level, thus leading to artefactually large values of *w*.

	Group F		Group P		Group S	
	Pre	**Post**	**Pre**	**Post**	**Pre**	**Post**
Numerical Operations	118.08±6.86		116.17±6.25		120.38±6.84	
Mathematical Reasoning	118.33±6.32		113.29±11.21		117.88±6.3	
Weber fraction	9.46±33.92		0.46±0.5		0.38±0.34	
ANT: alerting	27.03±22.91	27.07±17.99	17.54±11.8	24.35±16.48	16.98±23.07	30.14±21.61
ANT: orienting	29.44±32.28	41.25±22.1	53.33±23.95	53.24±24.71	44.58±26.61	49.56±24.23
ANT: executive	86.5±26.24	72.51±15.72	83.55±20.83	70.67±19.12	83.68±29.8	72.73±27.96
Digit span: forward	11.5±2.17	12±2.06	10.7±2.7	12±1.93	12±2.63	12.5±2.24
Digit span: backward	8.63±2.32	8.65±2.06	7.87±2.63	8.39±2.52	9±2.41	9.91±2.35
Number line	60.37±33.72	66.94±40.23	77.6±49.52	86.22±56.33	65.22±33.59	67.31±29.76
Corsi: forward	9.96±2.16	10.36±1.68	9.82±2.15	9.83±1.97	10.05±1.53	9.9±2.1
Corsi: backward	8.13±1.18	8.23±1.45	7.68±2.06	8.43±1.34	8.38±1.43	8.71±1.27

4.4.1.3 Training phase

4.4.1.3.1 Mean accuracy and RT

Mean accuracy and RTs for both types of problem (Calculation, Drill) were submitted to separate ANOVAs, with stimulation group (frontal/parietal/sham) as between-subjects factor and session (1 through 5) and block (1 through 5) as within-subjects factors. For sessions containing more than 5 blocks, the extra blocks were not used, for the sake of the factorial design of these analyses (Snowball et al., 2013). For Calculation analyses, the Numerical Operations score was used as a covariate, in order to control for individual differences in arithmetic abilities.

For Drill problems, there were no significant main or interaction effects when accuracy was the dependent variable. For the RT analysis, the 3-way interaction (Session × Block × Group) was significant ($F(32, 1024)=1.71$, $p<.01$, $\eta_p^2=.05$). These data are shown in the upper panel of Figure 27. When decomposing the 3-way interaction across levels of the Session factor, the 2-way interaction (Block × Group) was significant ($F(8, 256)=2.27$, $p<.05$) for session 1 only. Decomposing this interaction as well, the main effect of Group was n.s. for all levels of the Block factor.

For Calculation problems, the accuracy ANOVA produced a marginally significant Session × Group interaction ($F(8, 272)=1.73$, $p=.092$). However, when decomposing across the levels of the Session factor, the

main effect of Group was n.s. for all sessions. The RT ANOVA produced a marginally significant main effect of Group ($F(2, 68)=3.03$, $p=.055$, $\eta_p^2=.08$). Exploratory post-hoc comparisons between groups pair-wise revealed significantly lower RTs for the parietal ($F(1, 68)=4.80$, $p<.05$) and frontal groups ($F(1, 68)=4.36$, $p<.05$) as compared to sham. The Calculation RT data are shown in the lower panel of Figure 27, similarly across sessions and across blocks, for a more intuitive comparison with the homologous Drill data, presented in the upper panel of the same figure. The group means for Calculation and Drill RTs, collapsing across blocks and sessions, are shown in Figure 28.

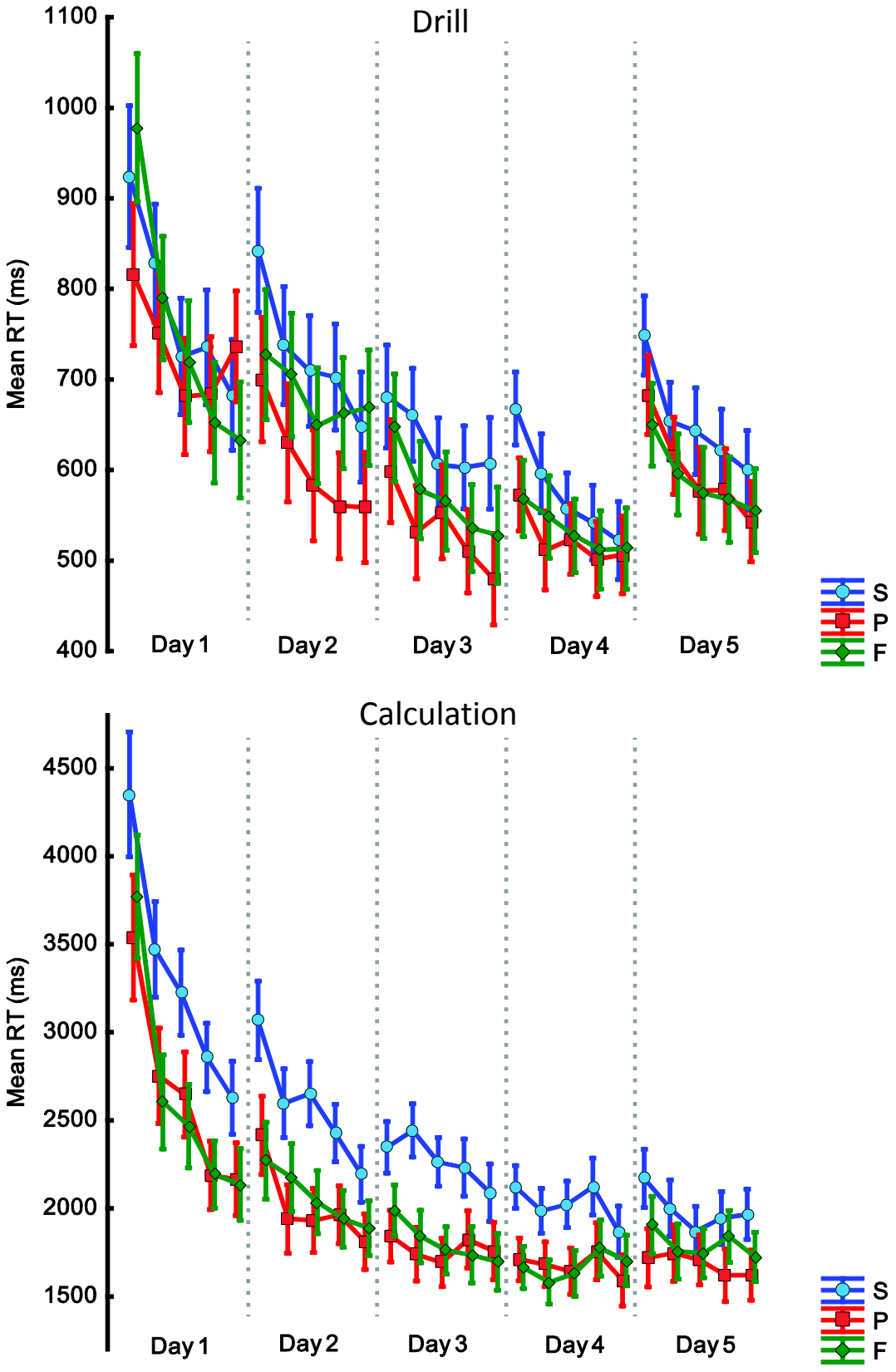


Figure 27: Drill and Calculation RTs across the duration of training. Each daily session consisted of 5 blocks, represented by individual markers colour-coded to represent the sham, parietal and frontal stimulation groups, respectively. Vertical dotted lines separate the 5 training days. Error bars represent ±1 S.E.M. (standard error of the mean).

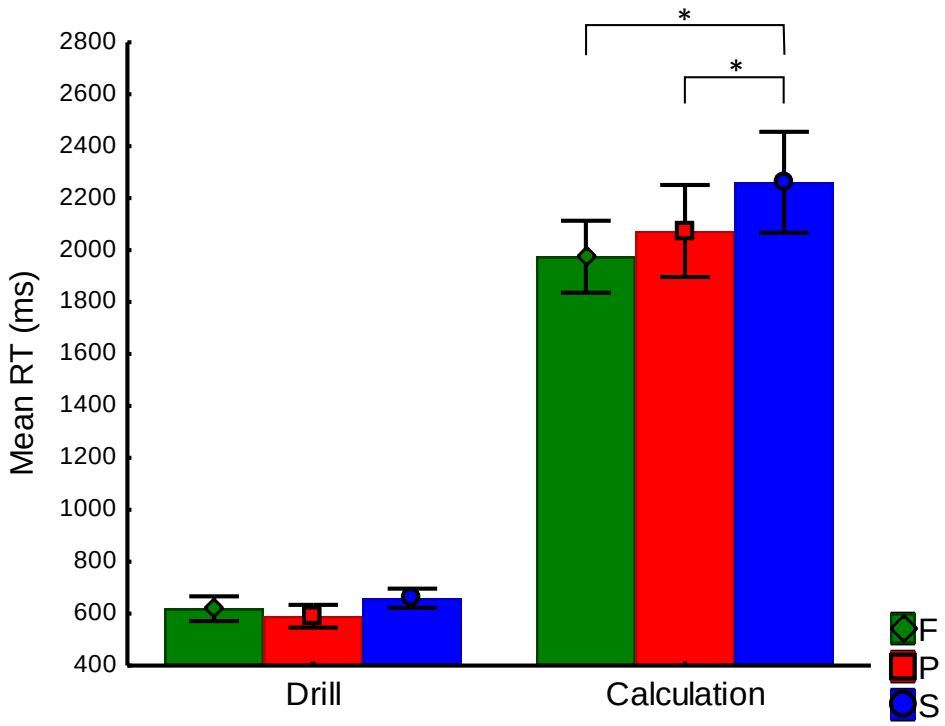


Figure 28: Means for RTs on Drill and Calculation problems, for each of the three stimulation groups. Error bars represent ±1S.E.M. (standard error of the mean).

4.4.1.3.2 Modelling learning curves

A 1-way between-subjects ANOVA for the dependent variable B (representing Calculation RTs in session 1), with stimulation group as between-subjects factor, revealed no significant main effect of group ($p>.1$), indicating no group difference in initial performance. However, in order to control for inter-subject variability in this parameter, I used B as a covariate in an analysis of covariance (ANCOVA) done for dependent variable α (Snowball et al., 2013), which indexes Calculation learning rate. Group remained as a between-subjects factor, however its main effect was non-significant ($p>.1$).

4.4.1.3.3 Online and offline effects

Separate ANOVAs were run for online and offline effects, for both RTs and accuracy levels. The between-subjects factor was always stimulation group (with 3 levels: frontal, parietal, sham), while the within-subjects factor was session (1 through 5) in the case of online effects and night (1 through 4) in the case of offline effects.

For online and offline Drill accuracy effects, as well as online and offline Calculation RT effects, there were no significant main effects of group, nor interactions with the within-subjects factor.

For offline Calculation accuracy, there was a marginally significant group × night interaction (F(6, 207)=2.10, p=.054). The main effect of group was not significant within any of the levels of the night factor, exploratory post-hoc comparisons revealed a trend for a smaller offline effect for parietal than for sham, within the fourth night (F(1, 69)=3.37, p=0.07). For online Calculation accuracy, the 2-way interaction was significant (F(8, 276)=2.02, p<.05) and was due to a main effect of group within session 5 (F(2, 69)=3.59, p<.05). In turn, this reflected higher values for the online effect for parietal vs sham (F(1, 69)=6.83, p<.05). These data are shown in Figure 29.

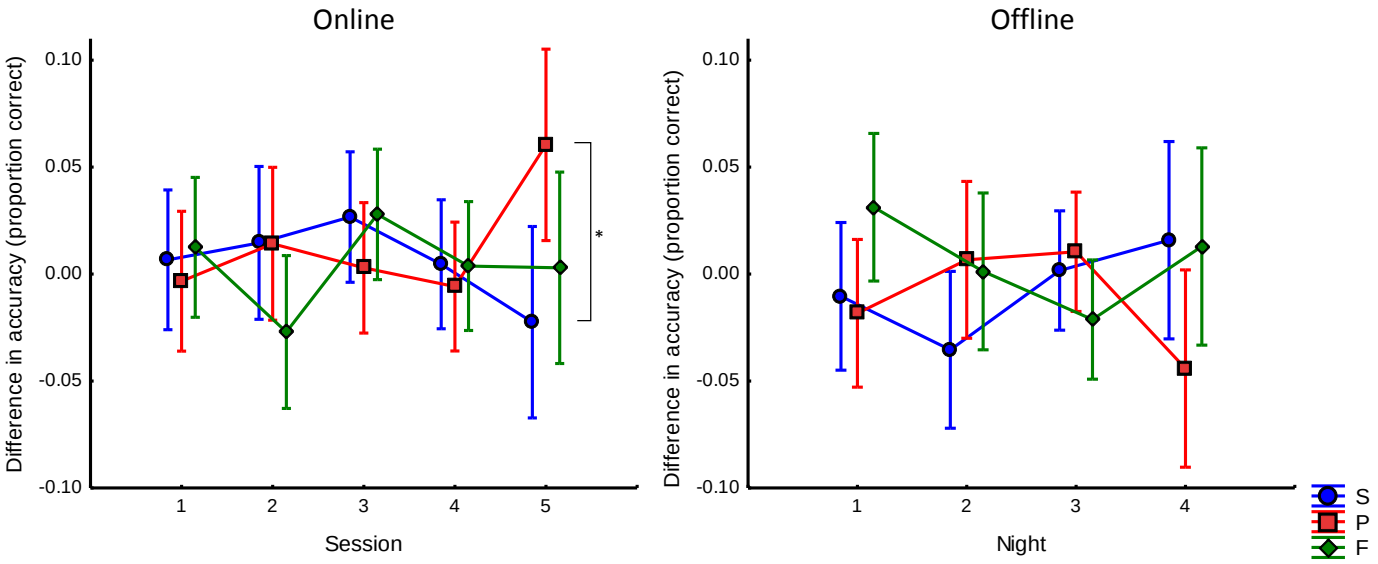


Figure 29: Online and offline effects for Calculation accuracy. Group × session interactions for online (left panel; significant) and offline (right panel; marginally significant) effect. Online effects are within-session while offline effects are between-sessions ("nights").

For offline Drill RT there was a significant main effect of group (F(2, 64)=3.39, p<.05), which was due to overall significantly greater values of the offline effect for sham than for frontal (F(1,64)=6.26, p<.05), and to marginally significantly greater values for sham than for parietal (F(1,64)=3.43, p=.069). For online Drill RT, the 2-way interaction was significant (F(8, 256)=3.3591, p<.01). This was due to a main effect of group within session 1 (F(2, 64)=4.87, p<.05), which in turn was due to significantly greater values of the online effect for parietal versus frontal (F(1, 64)=9.51, p<.005), and a trend for parietal versus sham (F(1, 64)=3.71, p=.059). These data are shown in Figure 30.

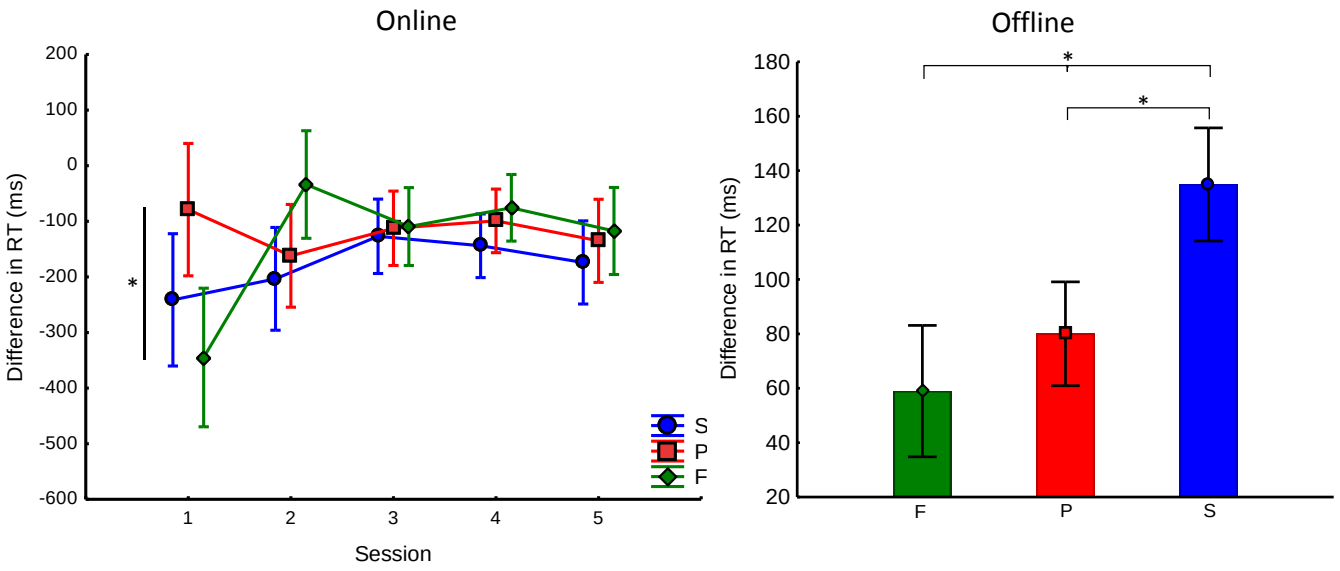


Figure 30: Online and offline effects for Drill RTs. Significant group × session interaction for the online effect (left panel) and main effect of group for the offline effect (right panel). Additionally, when online and offline effects of each type were entered into pairwise correlational analyses, a significant negative correlation was found between effects for Calculation accuracy ($r(72)=-0.87$, $p<.001$) and for Drill RT ($r(67)=-0.73$, $p<.001$), both depicted in Figure 31. These correlations remained equally significant if performed within each group as opposed to by pooling across all groups.

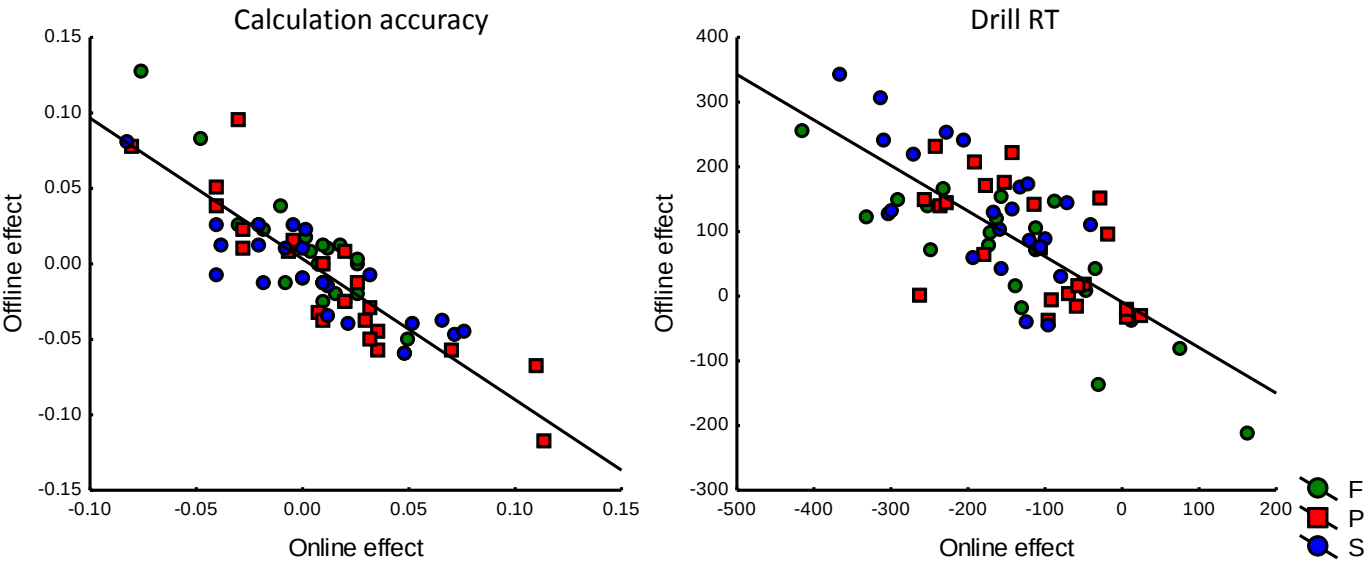


Figure 31: Significant negative correlations between online and offline effects, for Calculation accuracy (left panel) and for Drill RTs (right panel). These correlations are significant when computed both across and within the three stimulation groups.

4.4.1.3.4 Reaction time distribution analysis

For each session of each subject, the RT distribution for each type of problem was submitted to an analysis which estimated the ex-Gaussian parameters μ, σ and τ (Lacouture & Cousineau, 2008). Each of

these was then used as a dependent variable in separate mixed-model ANOVAs with stimulation group as between-subjects factor, and session as within-subjects factor. For Calculation σ, the 2-way interaction was significant ($F(8, 272)=2.36$, $\eta_p^2=.06$, $p<.05$). This was due to a significant main effect of group within session 1 ($F(2, 68)=4.40$, $p<.05$), which in turn reflected greater σ values for sham as compared to parietal ($F(1, 68)=5.90$, $p<.05$) and to frontal ($F(1, 68)=7.36$, $p<.01$). For Calculation τ, the main effect of group was significant ($F(2, 68)=3.46$, $\eta_p^2=.09$, $p<.05$). This was reflected in greater τ values for sham as compared to parietal ($F(1, 68)=3.94$, $p=.051$) and to frontal ($F(1, 68)=6.26$, $p<.05$). These effects for the ex-Gaussian parameters of the Calculation distribution are shown in Figure 32. All other effects were non-significant.

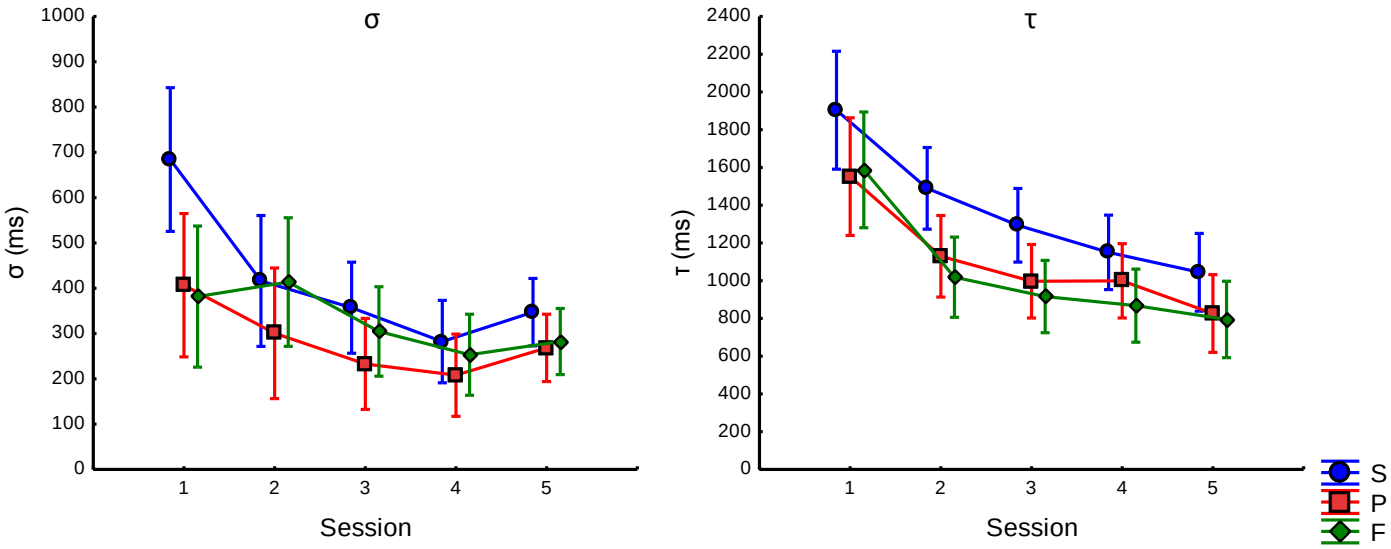


Figure 32: Ex-Gaussian parameters σ and τ, for Calculation RTs, shown across sessions and across groups.

4.4.1.4 Testing phase

Mean RTs and accuracies from the testing phase subsequent to the final training session, on both Calculation and Drill problems, were submitted to Group (frontal/parietal/sham) × Novelty (old/new problems) ANOVAs. For all of these, there were no significant main or interaction effects. For Calculation RTs (shown in Figure 33), there was a trend for a Novelty × Group interaction ($p=.066$), however exploratory post-hoc comparisons contrasting the groups pair-wise for both Novelty levels were n.s. (all $ps>.2$).

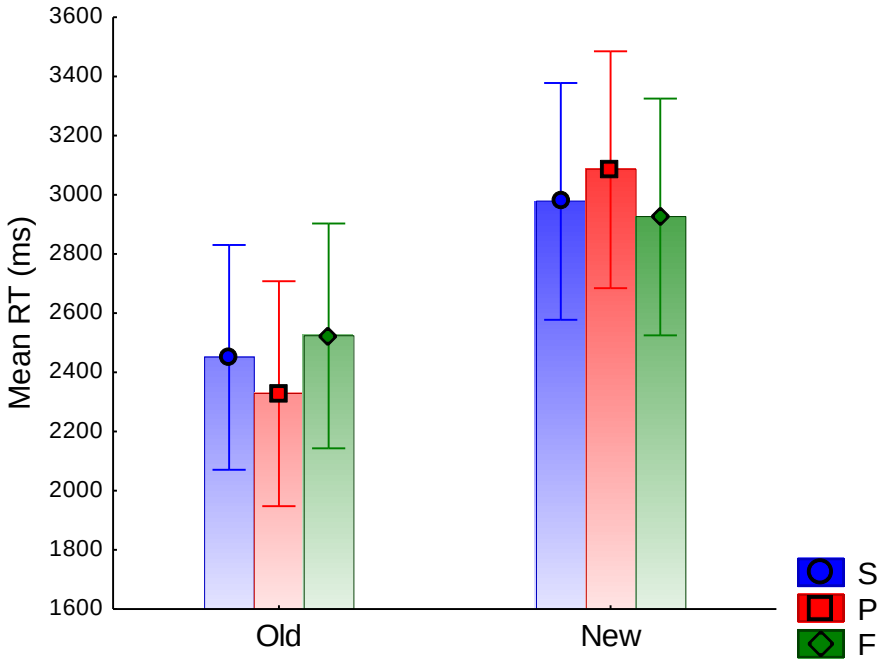


Figure 33: Mean Calculation RTs during testing phase, shown separately for old and new problems, across the three stimulation groups.

4.4.1.5 Stimulation perception

Table 9 shows the subjects' answers when asked, at the end of the experiment, whether they thought they were administered real or sham stimulation, based on their sensation during the training sessions. This 3×3 contingency table was collapsed into a 2×2 table by combining both real stimulation groups into one and considering the uncertain responses equivalent to "sham", i.e. stimulation perception not noticeable. A chi-squared test on this table revealed that subjects were unable to correctly identify their stimulation group (Yates corrected χ^2=.44, p>.1), which is in line with previous studies suggesting that tRNS offers good blinding control (Ambrus et al., 2010; Fertonani et al., 2011; Prichard et al., 2014).

Table 12: Contingency table with subjects' reflections on stimulation sensation. Cells contain number of responses of each type.

Stimulation group	Response: "tRNS"	Response: "Sham"	Response: "I do not know"	TOTAL
Frontal	5	17	2	24
Parietal	6	16	2	24
Sham	8	13	3	24
TOTAL	19	46	7	72

4.4.2 Brain results

4.4.2.1 Grey matter

4.4.2.1.1 VBM

For the whole-brain VBM analysis, the relative concentration of grey matter density (GMD) in the "post-pre" image (see Methods) was compared across groups cross-sectionally, as part of the second-level analysis (see Methods). Several contrasts were examined, with the significant clusters for each reported below. For the parietal group only, there were 4 clusters of increased GMD from pre to post stimulation, and 3 clusters of decreased GMD from pre to post. These clusters are described in Table 13. Also, clusters #3 and #4 from the Post>Pre contrast, as well as clusters #2 and #3 from the Pre>Post contrast applied within group parietal only, are displayed in Figure 33 and Figure 34 respectively, superimposed on top of a standard MNI template.

Table 13: Clusters that were significant (after correction) in the whole-brain group comparison VBM analysis.

Contrast	Cluster #	Voxels	Max t	Peak (MNI, mm)			Centre of gravity (MNI, mm)			Region (Harvard-Oxford atlas)
				X	Y	Z	X	Y	Z	
Post > Pre (parietal group only)	1	3	4.99	18	50	-16	18.7	51.3	-16	R frontal pole
	2	10	5.14	-52	-46	20	-53.1	-46.6	20.4	(L ventricle)
	3	151	5.96	6	-38	32	4.44	-39	28.2	R cingulate gyrus
	4	319	4.75	-34	18	58	-36.3	-0.5	60.3	L middle frontal gyrus
Pre > Post (parietal group only)	1	6	3.5	-10	-64	2	-11	-63	2.67	L lingual gyrus
	2	68	3.66	-18	-54	-8	-19.9	-51	-8.69	L lingual gyrus
	3	1371	6.37	4	-16	28	-2.76	-19.7	15.7	R cingulate gyrus

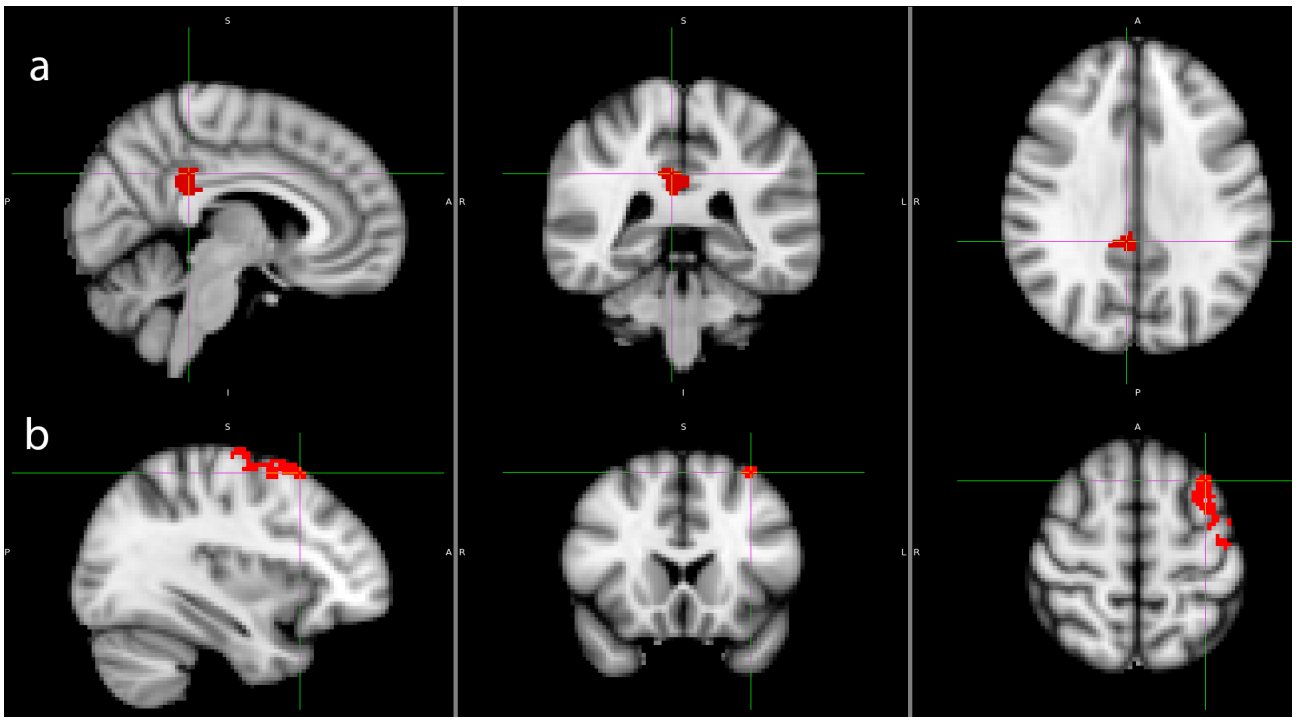


Figure 34: Clusters of significant GMD increase in the parietal group: (a) right cingulate gyrus; and (b) left middle frontal gyrus. These represent clusters #3 and #4 of the "Post>Pre (parietal group only)" contrast, detailed in Table 13.

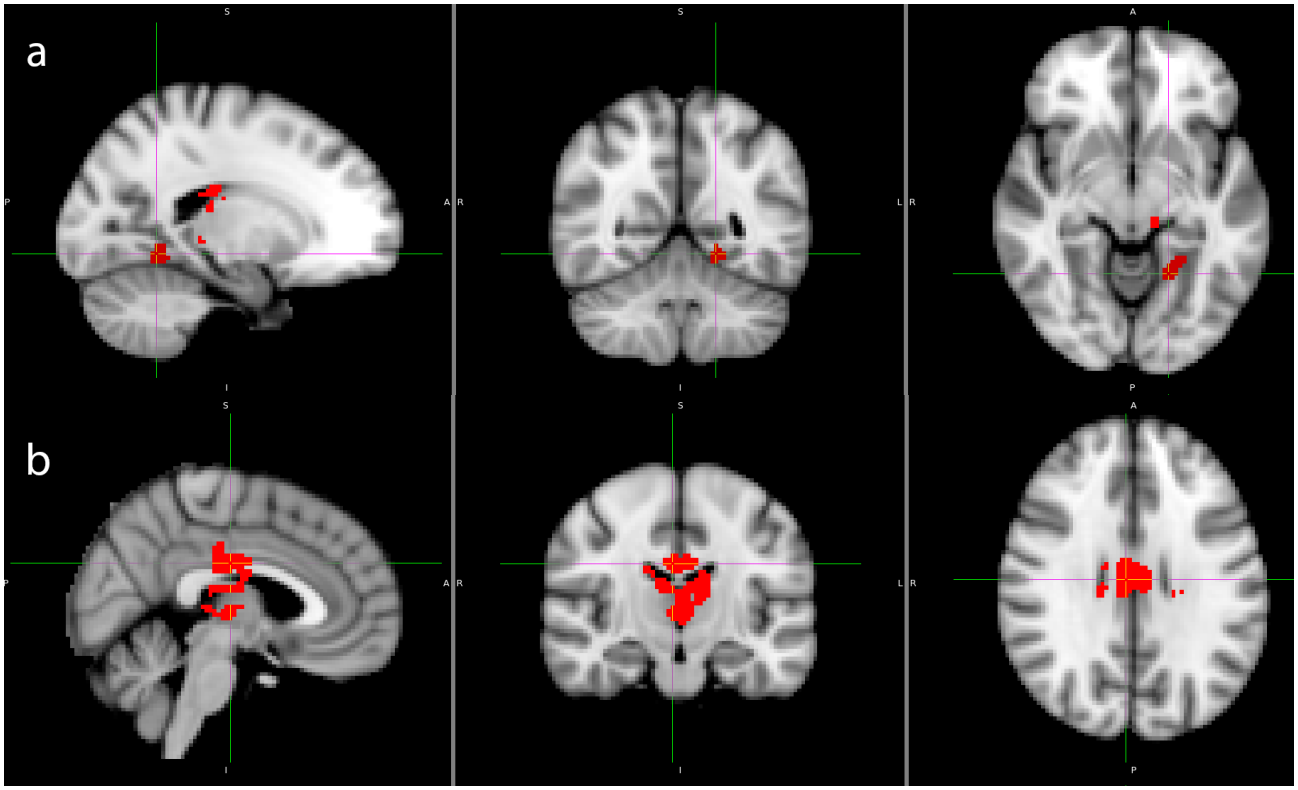


Figure 35: Clusters of significant GMD decrease in the parietal group: (a) left lingual gyrus; and (b) right cingulate gyrus. These represent clusters #2 and #3 of the "Pre>Post (parietal group only)" contrast, detailed in Table 13.

None of the contrasts where correlations were sought between GMD and behavioural measures, produced any clusters that survived correction for multiple comparisons. These results did not change even when the statistical power of the design was increased by discarding age and gender as regressors of no interest. A further analysis was done where mean GMD values were extracted across all subjects in the Parietal group, from the four clusters pictured above where effects were found for this group. Correlations were tested with behavioural measures, both relating to baseline maths ability (i.e. Numerical Operations) as well as to learning (offline effect for Drill RT, which was the effect that showed a group difference). None of these 4 ×2=8 correlations were significant at the Bonferroni-corrected significance level (α=0.05/8=0.00625), and also not even at the uncorrected level (α=0.05).

In the ROI analysis, there were no significant effects involving the stimulation group factor, neither for the electrode-based not for the hypothesis-driven ROIs.

4.4.2.1.1 SBM

Comparing the groups pair-wise, i.e. sham against each of the two real stimulation groups (frontal and parietal), produced a number of significant clusters where there were group differences in the pre-to-post change in surface-based cortical metrics. These clusters are detailed in Table 14 and shown projected on a standard inflated brain in Figure 36.

Table 14: Regions with significant group differences in term of the pre-to-post change in surface-based measures. Clusters in red font represent increases in the pre-to-post change for the real stimulation group (frontal or parietal) versus sham stimulation while clusters in blue represent decreases. Coordinates are given in standard MNI space. LGI=local gyrification index; CT=cortical thickness. LH = left hemisphere, RH = right hemisphere. Max=the maximum value ($-\log_{10}p$) in the cluster. NVtxs=number of vertices in cluster.

Groups involved	Comparison	Measure	Max	Size (mm^2)	x	y	z	N Vtxs	Hemi	Main structure	Cluster
Frontal Sham	pc1_F > pc1_S	LGI	1.5058	953.57	-48	-11	-12	2089	LH	superior temporal	a
	pc1_S > pc1_F	Volume	-1.34	533.24	9	-55	49	1403	RH	precuneus	b
Parietal Sham	pc1_P > pc1_S	CT	2.9208	895.69	8	17	52	1688	RH	superior frontal	c
		Volume	4	1015.55	7	36	19	2025	RH	rostral anterior cingulate	d
	pc1_S > pc1_P	LGI	-1.472	944.41	-16	-72	61	2066	LH	superior parietal	e

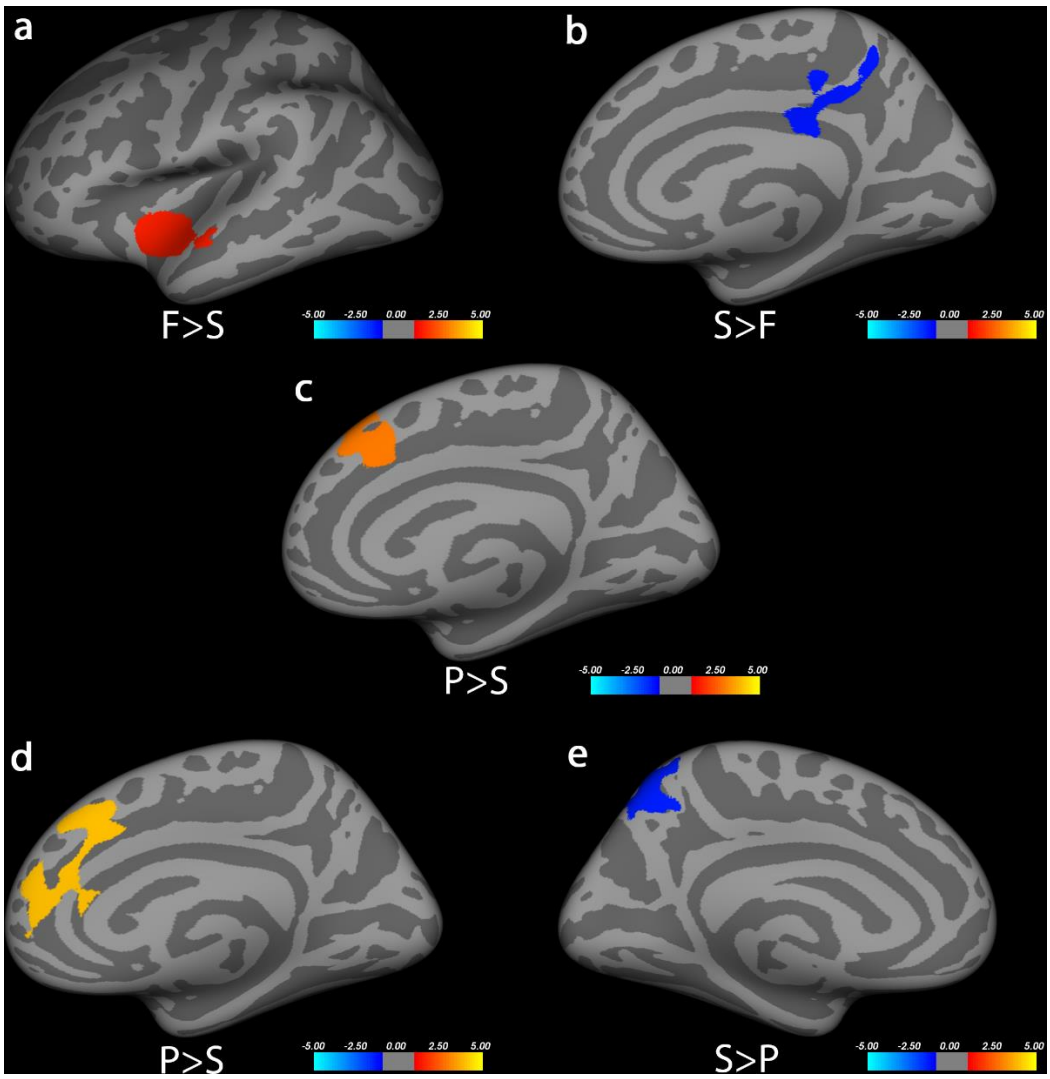


Figure 36: Clusters with significant group differences in the pre-to-post change, projected onto an inflated standard FreeSurfer brain (MNI305 space). Cluster **a** is shown in lateral view and clusters **b-e** are shown in medial view. Clusters **a** and **e** depict the left hemisphere. Clusters in shades of red represent increases in the pre-to-post change for the real stimulation group (frontal or parietal) versus sham stimulation while clusters in shades of blue represent decreases. FreeSurfer uses the colour scale to represent sign(effect)*-$\log_{10}(p)$.

4.4.2.2 White matter

4.4.2.2.1 TBSS

The whole-brain TBSS analysis that checked for the presence of pre-to-post group differences in FA, or correlations between FA and the behavioural measures, produced no significant clusters, after correction for multiple comparisons. These results did not change even when the statistical power of the design was increased by discarding age and gender as regressors of no interest.

4.4.2.2.2 Tractography

The session (pre / post) × tract (F3-F4 / P3-P4 / lSFG-rSFG) × group (frontal/parietal/sham) ANOVA did not yield any significant effects involving the stimulation factor.

4.5. Discussion

In this study, I ran a longitudinal investigation into the structural brain changes associated with previously-shown facilitating effects of tRNS on cognitive training in arithmetic. The results confirm that tRNS improves calculation performance, and it appears to do so via mechanisms (online and offline learning) that are negatively correlated across subjects and thus possibly mutually exclusive. These behavioural effects were accompanied, for the real stimulation groups as compared to sham, by both increases and decreases in the cortical volume and thickness of frontal, parietal and temporal clusters that were distal to the site of the stimulating electrodes. All these results are discussed in more detail in the following sections.

4.5.1 Behavioural results

During the training task, both frontal and parietal tRNS stimulation decreased RTs on the Calculation task, as compared to sham. This overall decrease was mainly driven by a decrease in the mean of the exponential tail (τ) of the RT distribution, as shown by an ex-Gaussian distributional analysis. Although showing no overall group differences, Drill RTs nevertheless differed between the real and the sham stimulation groups in terms of their offline (consolidation) effects, which were improved both for frontal and (marginally) for parietal stimulation.

Previous studies based on similar arithmetic training paradigms found similar facilitatory effects when using tRNS. These include the Snowball et al. study (2013) and Study 2 of this thesis (Chapter 3). In Study 2, Calculation RTs were decreased for the group receiving tRNS stimulation as opposed to sham. The tRNS group in that study received prefrontal stimulation for the first 3 days of training and parietal stimulation for the final 2 days. If, like in that study, here I would have maintained the difficulty manipulation of either few (6) or many (12) problems used as training material, it is likely that the Calculation effects would have only remained in the more difficult condition, just like they did in Study 2. This speculation is also based on the fact that tES is believed to selectively increase excitability in neural populations active at the time of stimulation (Silvanto, Muggleton, & Walsh, 2008), and that previous studies found the effectiveness of tES to increase with task difficulty (Pope & Miall, 2012).

In Snowball et al.'s study (2013), not only Calculation but also Drill RTs were decreased for frontal versus sham stimulation. Thus, while the current study replicates those results in terms of the Calculation facilitation for frontal tRNS, it does not in terms of Drill facilitation; it does, however, find the former effect

also for parietal stimulation. Unlike in Snowball et al.'s study (2013), the main effect of stimulation here was not accompanied by an interaction with the session repeated measures factor, or by an associated increase in learning rate. This might at first sight suggest that the decreased Calculation RTs observed for the parietal and frontal groups simply reflect a coincidental assignment of people with higher arithmetic performance to these groups. While still unlikely to have occurred by chance, such a scenario is further dispelled by the analysis done on the power-law parameter of initial performance (B), which showed no group difference. Thus, the decreased overall Calculation RTs (across the entire training), both in the case of the frontal and of the parietal group, have a plausible cause in the stimulation that these groups received.

The mechanisms by which tRNS achieves facilitatory behavioural effects such as the ones observed in Studies 2 and 3 of this thesis (Chapter 3 and the current Chapter 4), as well as in other tRNS studies in numerical cognition (Cappelletti et al., 2013; Snowball et al., 2013) is not completely understood. Possible mechanisms include stochastic resonance (Terney et al., 2008) and potentiation of voltage-gated channels (Chaieb et al., 2009) that may induce LTP-like cortical plasticity (Kuo & Nitsche, 2012). Regrettably, although between Snowball et al. (2013) and Studies 2&3, this Drill/Calculation learning paradigm has been used for a total of 3 times, with frontal-only, parietal-only and frontal-to-parietal stimulation between them, a direct comparison between the efficacies of these three stimulation montages is still not possible, due to slight differences in the behavioural paradigm version that was used in each case.

When fitting an ex-Gaussian function to RT distributions from the Drill and Calculation trials, group differences were found in the parameters of these curve fits, for Calculation only. The σ parameter, indexing the variance of the normal part of the distribution, was decreased – within session 1 only – in both real stimulation groups (frontal, parietal) as compared to sham, thus reflecting a more narrow distribution of RTs in those groups. Additionally, the τ parameter, indexing the exponential tail of the distribution (i.e. the longer RTs), was also significantly decreased, across all sessions, in the frontal group vs sham (and marginally significantly for parietal vs sham), just as it was for the tRNS group vs sham in Study 2. Although speculative, this suggests that one common mechanism by which tRNS contributes to speeding up calculation in these two studies (and in others discussed above) may consist of a change in the strategies underlying these calculations. In a study where subjects solved addition problems in different formats (Arabic numbers or number words) and reported their strategy (calculation or memory retrieval) for each trial, an ex-Gaussian analysis suggested that, for each format, retrieval trials contributed predominantly to μ, which indexes the

mean of the normal part of the distribution, while calculation trials contributed to τ (Campbell & Penner-Wilger, 2006). Similarly, LeFevre et al. (2006) found that τ discriminated well between the different self-reported strategies (such as retrieval, transformations or counting) used in a simple subtraction task, having small values for retrieval and large values for counting. The smaller τ caused by tRNS might therefore reflect faster and/or more efficient calculation, while the smaller σ observed in session 1 might indicate that the plateau reached in terms of consistent calculation strategies might have been reached sooner for the stimulated groups.

During the testing phase, the tRNS groups were not enhanced with regard to either RT or accuracy, for either problem type. The stimulation factor also did not interact with problem novelty. For Drill, this result is not surprising given that Drill problems were not facilitated during training either. It is, however, surprising that old Calculation problems were not facilitated, as for these problems similar performance to the final blocks of the training phase would have been expected. This might have been due to the slightly different inter-stimulus interval used in the testing phase as opposed to the training phase (in turn due to the lack of feedback in the former). The fact that the trend (see Figure 33) was not similar to what was observed in the training suggests against this being due to the lower statistical power caused by the fewer data points in this testing phase as compared to the training phase. On the other hand, the lack of facilitation for new problems suggests that the RT enhancement observed during training did not transfer from the trained material to new problems. This is in contrast to the results of Study 2 and those of Snowball et al. (2013), where transfer was found to occur for Calculation.

Stimulation did not affect performance on control tasks tapping into the auxiliary functions of attention, verbal and visuo-spatial working memory. Also unaffected was performance on the number line task. A transfer from symbolic number training to number-to-space mapping would have not been surprising. Number line training has been shown to improve improved arithmetic performance in children with dyscalculia (Kucian et al., 2011), and so the scenario of an inverse transfer – from arithmetic to number mapping – seems plausible. Also, a week-long training on symbolic magnitude comparisons using artificial digits improved performance on a number line task using those digits (R. Cohen Kadosh et al., 2010). The reason for the lack of transfer here, then, is unclear.

4.5.1.1 Online and offline (consolidation) learning

Learning achieved through tES-assisted training has been suggested to be partly due to so-called *online* effects, occurring *within* a stimulating session, and partly to *offline* (or: consolidation) effects occurring in the time *between* consecutive training sessions. In one such training programme, aimed to improve skill on a fine motor task, offline effects were found to be predominantly responsible for the observed facilitatory effects of anodal transcranial direct current stimulation (Reis et al., 2009). Intuitively, these offline effects reflect the need for knowledge or skill to "sink in" in between training sessions. On the other hand, a subsequent study from the same group found no significant improvements caused by either tDCS or tRNS stimulation over a sham group in terms of either online or offline learning, although descriptively a trend was observed for *within*-session improvements, i.e. *online* effects, for both types of stimulation (Prichard et al., 2014). Based on the theoretical suggestion made in these motor training studies, here I tried to disentangle the two types of effects in order to see whether, for cognitive training as well, one of the effects perhaps has a more dominant role in the overall learning process than the other. Although stimulation decreased overall RTs in Calculation, neither effect (online, offline) differed significantly between the groups. It is possible that the correspondence between improved RTs and improved online/offline effects might have been detected if a more sophisticated measure of skill had been used, in place of simply computing effects for RT and for accuracy. For instance, Reis et al. (2009) computed the effects for a "skill measure" that was defined using an in-house mathematical model such that a change in it reflected a change along the speed-accuracy trade-off curve for the particular motor skill task that was used. Such a skill measure would have not been easily suited to the cognitive task used in the current study's training. It is also possible that the group effect found in the analysis of RTs was simply not due to either online or offline effects, since this distinction is, of course, not the only one that can be made to disentangle different subcomponents of learning.

A group difference in offline effects was found here for Drill RTs, with significantly smaller values for frontal and a trend for parietal, versus sham. A smaller effect in this case indicates more consolidation (i.e. faster responses in the first block than the last block of the previous session), therefore this effect can be considered a form of facilitation for the real stimulation groups as compared to sham. Again, however, these Drill RT group differences in offline effects were not accompanied by a corresponding difference in overall RTs, which makes the effect more difficult to interpret. It is possible that the two effects cancelled each other out in this case, thus leading to the lack of a group difference in overall Drill RTs. In other words, while

stimulation improved learning by consolidation between sessions, it may be that it was not improving (or was even preventing) online learning within sessions, as the significant negative correlation (across all groups and within each group) indeed suggested. This suggests that, independent of stimulation condition (i.e. including sham), some people can be thought of good "online learners" but poor "offline learners", and vice-versa for others, such that no subject could ever show large or small values in both effects simultaneously. It is possible that regression to the mean is the most parsimonious explanation of this phenomenon.

4.5.2 Brain results

4.5.2.1 VBM

The structural scans acquired pre- and post-training were analysed using FSL, using both voxel-wise and regionally-averaged statistics, in order to examine how the pre-to-post changes in grey and white matter were influenced by stimulation group. Additional analyses were run by modelling the cortical surfaces in FreeSurfer, and thus obtaining an independent set of cortical parameters that complemented that obtained in FSL.

Grey matter analyses performed with VBM produced the GMD (grey matter density) metric, and the post-pre difference was analysed cross-sectionally within and between groups. There were no significant group differences in GMD anywhere in the brain. However, within the parietal group only, two clusters were found where GMD has significantly increased from pre-to-post: these were in the region of the right cingulate gyrus (posterior division) and the left middle frontal gyrus. Two other clusters showed an effect in the opposite direction, i.e. GMD within the parietal group significantly decreased pre-to-post; one cluster was in close proximity to the cingulate cluster mentioned above; the other one was in the region of the left lingual gyrus. Since these effects were only present in one stimulation group (parietal), as opposed to also in the contrast with sham, this means that they do not unambiguously tell us whether they were likely to have been caused by tRNS. In order to test the meaningfulness of these regions in the context of the tRNS-assisted arithmetic training, correlations were computed, within the parietal group, between these clusters' mean GMD and behavioural measures that characterised the groups at baseline (Numerical Operations) or that characterised the learning processed and differed between groups (offline effects for Drill RTs). No significant correlations were found, which leaves little room for meaningful interpretation of this result, which may reflect a Type I error. Such correlations between pre-to-post change in GMD and behavioural scores were

also sought by including the behavioural scores as regressors in an additional set of whole-brain VBM analyses. However, here too, no GMD-behaviour correlations were found either within or between groups. It is quite possible that the current sample size (24 subjects per group), although not small by current structural neuroimaging standards (Friston, 2012), was not enough to detect any brain-behaviour correlations based on the measures acquired and on the particular statistical tests performed in this study. If such correlations do indeed exist between those measures, it seems plausible that their effect size is quite small, such that an even larger sample size would have been needed for their detection.

In addition to these whole-brain analyses, region of interest (ROI) analyses were also performed on the pre-to-post change in the VBM-produced GMD metric. ROIs were defined based on the locations of the stimulating electrodes (frontal or parietal) and also based on the regions where group differences or GMD-behaviour correlations were found in Study 1. No group differences were found, however, which suggests that either the tRNS-assisted training did not have a differential impact in these particular regions, or that the within-ROI averaging process cancelled effects that existed in subparts of a single ROI. See the Limitations section for further discussion about the accuracy of the electrode-based ROIs' locations.

4.5.2.2 SBM

The voxel-based analyses were supplemented by surface-based morphometric analyses performed in FreeSurfer. Just as in Study 1, VBM and SBM produced divergent sets of results. The total number of significant clusters was larger with SBM, most likely due to the technique's increased sensitivity as compared to VBM (Winkler et al., 2010). Group differences in the pre-to-post change in a number of surface-based metrics were found. The local gyrification index (lGI) in the left superior temporal gyrus increased more in the frontal than in the sham group, and the lGI in the left superior parietal lobule increased more in the sham than in the parietal group[14]. Alteration of brain gyrification, as indexed by changes in lGI, has mainly been linked with conditions such as schizophrenia (Palaniyappan et al., 2011) and Williams' syndrome (Gaser et al., 2006), but very rarely with measures outside of the clinical realm. Tentative evidence for a role in cognition and experience-dependent plasticity comes, for instance, from studies reporting that differences in cortical folding predict executive processes (Fornito et al., 2004) and discriminate between meditators and non-

[14] The left superior parietal lobule cluster were group differences were found in Study 2 was located quite far from the one identified here: (30,-62,44) vs (-16,-72,61) respectively.

meditators (Luders et al., 2012). To my knowledge, no previous longitudinal results have reported lGI changes as a result of cognitive training, which makes the current results rather difficult to interpret.

Pre-to-post change in cortical volume was greater in the parietal than in the sham group in a right anterior cingulate cluster, and greater in the sham than in the frontal group in a right precuneus cluster. While the cingulate cortex holds an apparent record for level of involvement in several types of MRI study (Gage, Parikh, & Marzullo, 2007), the fact that greater increase in volume was observed in this frontal location for the group receiving parietal stimulation is surprising and will be discussed in the next paragraph.

Finally, greater pre-to-post change in cortical thickness was observed in the right superior frontal gyrus cluster, when comparing the parietal to the sham group. The mismatch between the cluster location and the stimulation location for this cluster as well as the previous cluster discussed (which showed the volume effect) raises interesting questions about the spatial distribution of cause and effect in tES. While the ability of tES to modulate connectivity in the brain is well researched (Dayan et al., 2013; Keeser et al., 2011; Polanía, Paulus, Antal, & Nitsche, 2011), the exact cause-effect spatial link between electrode location and affected area(s) is not fully understood. It is believed that areas distal to the stimulation site also receive an important amount of electrical charge (Marom Bikson, personal communication), and it is also believed that tES mainly affects neural populations that are active during the stimulation period (Silvanto et al., 2008). With these arguments in mind, it is perhaps possible that, for the frontal clusters (right cingulate and right SFG) where the above effects were found for parietal stimulation, an interaction was at play between the involvement of the frontal and parietal cortices during calculation and drill and the parietal location of the stimulation, whereby structural changes occurred within the more distal (frontal) active area, as opposed to the more proximal one.

The millimetre spatial resolution of MRI does not allow disambiguation between the different cellular mechanisms that might explain the macroscopic changes observed. For grey matter changes, such candidate mechanisms include synaptogenesis (Kleim, Lussnig, Schwarz, Comery, & Greenough, 1996), gliogenesis, changes in neuronal morphology, dendritic spine morphology (Kolb, Cioe, & Comeau, 2008), loss of neuropil and altered pruning (Voets et al., 2008). Additionally, different phases of learning might be associated with different cellular mechanisms. For instance, an initial learning phase might be associated with rapid but transient GM changes in functionally relevant brain areas, whereas the later learning phases (e.g. consolidation and long-term storage) might bring about more persistent GM changes. Indeed, evidence from

animal and human studies suggests that the time-course of grey matter change is not linear (Kleim et al., 2004; Taubert et al., 2010). MRI might therefore not able to distinguish these different phenomena, because their time-scale might be lost between the pre and the post scan of a longitudinal study such as the current one.

4.5.2.1 White matter

For white matter, none of the analyses that were carried out produced any significant results. The pre-to-post change in fractional anisotropy (FA) was not significantly different from zero within any of the stimulation groups, did not differ between those groups, and did not correlate with behaviour either at the whole-brain or at the ROI level. Also, values of this pre-to-post change in FA were extracted from homotopic tracts linking the ROIs defined in the analyses mentioned above, tracts that were probabilistically defined using tractography. Based on previous human and monkey studies suggesting that homotopic tracts display a high degree of variability between regions and between individuals (e.g., Putnam et al., 2009; Rouiller et al., 1994), I expected that the most relevant of these tracts, e.g. those linking the sites of the stimulating electrodes, would have different integrities, as would have been suggested by a group difference in their mean FAs. However, such a difference was not found.

Using diffusion MRI, a recent study found FA changes after humans had been only trained for two hours (Sagi et al., 2012). The difference between that result and the null diffusion results in this study (as well as those of Study 1) might plausibly be due to the (as of yet still) large differences between diffusion sequences employed in MRI scanners around the world (Saad Jbabdi and Mark Jenkinson, personal communication). Results from training studies using diffusion imaging will be difficult to compare across labs, and reliable conclusions drawn, before more consistency is achieved in the parameters of diffusion scans. It is also possible that, given the relatively long interval between the two scans here, many miniscule white matter changes occurred, in small time windows of perhaps a few hours (Sagi et al., 2012), in between those scans, and in parallel with the tRNS-assisted training, and were thus not picked up in the pre-to-post comparison.

4.5.3 Limitations

It is not possible, given the current data set, to identify structural brain regions that have been changed by arithmetic training alone. The comparisons necessarily involve the stimulation group factor, the two

factors (training and stimulation) being confounded. This is for two reasons. Firstly, it is likely that training alone would have affected the brain differently from tRNS-assisted training (Cappelletti et al., 2013). Secondly, even if a pre-post comparison was done at the level of the entire sample (collapsing across stimulation group), there would be no control condition which to compare any observed changes to. Thus, it would not be possible to tell whether any observed changes are indeed due to the arithmetic training, or whether they would have happened anyway, e.g. simply as a result of the passage of time. In order for this question to be answered, a control group without training or with equally long training in a different domain would have been necessary. However, the aim of this study was not to study brain plasticity due to training alone but to see how tES modulates this plasticity when administered in parallel with the training, therefore the above argument is not so much an actual limitation of the study as a theoretically-motivated observation, from the broader vantage point of training and plasticity studies.

Additionally, it is not possible to disentangle the effects of the Drill and Calculation components of training, such that the observed structural changes are necessarily an effect of both types of problems contributing to the cognitive training.

In the current study, the dot enumeration task to measure ANS acuity was only run as a baseline procedure, and not as a pre- and post-training task. In a recent study, Cappelletti et al. (2013) found improved ANS acuity for a group that received training in this domain accompanied by parietal tRNS, as opposed to training alone, stimulation alone, or non-ANS training. The training length was similar to the one used here (5 days), and it would have been desirable to investigate if any of the real stimulation conditions had an enhancement effect on ANS acuity.

4.5.3.1 MRI

Some studies (e.g. Sampaio-Baptista et al., 2014) include, in addition to a post-training scan, also an additional follow-up scan (typically in a few months' time), that enables the state of the brain to be compared to that immediately following and preceding training. Had such a long-term follow-up scan been available, this would have allowed me to see whether the structural changes evolve differently for the stimulation groups and whether there is an interaction with behavioural performance.

4.5.3.2 tRNS

The positioning of the electrode-based ROIs could be disputed. Clearly, it is of interest to examine changes in brain structure both underneath the electrodes and at the location where the impact of the electrical current, in terms of electrical field strength, is maximal. However, while the former location is relatively straightforward to pinpoint at the cortical level (and has been done in this study), the latter is not. Recent models of current flow density based on head models suggest that the maximum current density is close to the midpoint between the two stimulating electrodes (Marom Bikson, personal communication). To find this exact point, however, computations beyond the scope of this thesis would have been necessary. Also, the improved spatial resolution informing the selection of a more relevant ROI would have not come without a cost, as various assumptions need to be made which are not necessarily always met. For instance, it would have to be assumed that the dependent variable – namely, GMD – is linearly dependent to current flow at every voxel, and that all regions are equally sensitive to current flow (Marom Bikson, personal communication).

Also, the spatial precision of the stimulation itself was limited by the 10-20 procedure used to determine the F3-F4 and P3-P4 locations. In future studies, more accurate placement of the electrodes to target the specific intended regions should be used, guided by techniques such as BrainSight. However, given the low spatial resolution mentioned earlier, such a solution might not be optimal, as the spatial accuracy gained in placing the electrodes would be later lost when assessing their effects in the analyses.

One of the most relevant questions to ask of any non-invasive brain stimulation study is the degree to which any observed effects are specific to the trained function, as opposed to other cognitive processes. In the case of maths training, it can be stated that the core process involved in mental arithmetic is an interaction between long-term memory (stored facts) and the central executive. These processes cannot be completely disentangled using the current data. In particular, it is reasonable to believe tRNS applied to both sites affected both of these processes woven together.

The observed invariance in the pre-to-post performance on the control tasks suggests that, as in some tES studies of numerical cognition (R. Cohen Kadosh et al., 2010; Snowball et al., 2013) – but not in others (Cappelletti et al., 2013) – the effects here were specific to the trained cognitive process and do not extend to domain-general abilities. It would, however, be extremely speculative to try and explain the exact cause of this specificity; this is because, as mentioned earlier, the brain regions underlying the electrodes are involved

not only in numbers and arithmetic but also equally in attention and working memory. Also, it is very difficult to rule out the possibility that some hidden cost exists for a brain region other than the one stimulated, or with regard to cognitive tasks other than the ones tested (Iuculano & Cohen Kadosh, 2013). Such an assumption is supported by yet unsubstantiated – but speculatively plausible – physiological mechanisms of tES that act as a "law of conservation", or "zero sum", whereby metabolic resources are simply reallocated within the brain rather than selectively augmented (Brem, Fried, Horvath, Robertson, & Pascual-Leone, 2014) . A limitation of the current study is that it does not enable me to test whether such a cost exists and, if it does, how it is manifested.

4.5.4 Conclusions

Taken together, these results suggest that tRNS-assisted cognitive training has the ability to improve responses in an arithmetic learning task, and that it likely also has an effect on the neural circuitry underlying calculation, by modifying structural parameters of brain regions believed to be involved in it. Such findings can be important when making connections between cognitive neuroscience and education. Future research should try to more accurately ascertain how different montages of non-invasive brain stimulation (such as high-definition tDCS, allowing a much better spatial resolution) with different stimulation sites can shape the brain not just immediately after training but also in the longer term. Functional techniques such as multivoxel pattern analysis (MVPA) can be used to study how representations of numbers in the context of arithmetic change not just across groups (Iuculano et al., 2014) but also longitudinally, across stages of learning.

4.6. Acknowledgements

I thank Dr Marinella Cappelletti for providing the dot enumeration task. I am thankful to Emiliano Santarnecchi and Ruth Caballero Hernández for help with computing the MNI coordinates corresponding to the 10-20 locations of the stimulating electrodes, and to Marom Bickson for useful conversations about current density modelling and ROI definition.

5 Study 4: Number-pitch-space interactions in the lesioned brain

"Music is a hidden exercise in arithmetic, of a mind unconscious of dealing with numbers"

– Gottfried Wilhelm Leibniz

5.1. Abstract

Although the symbolical representation of numbers is a human cultural invention seemingly separate from perceptual judgements, at the neural level magnitude and numerosity are not encoded by dedicated brain modules. Instead, there is evidence that the parietal lobe holds a common metric underlying the processing of number and several other dimensions pertaining to cognition, perception and motor control. One of these dimensions is musical pitch, who shares not only neural resources with number but also the observable characteristic of interacting with space when a left/right-localised response has to be made to stimuli requesting its implicit or explicit judgement. Such phenomena are referred to as stimulus-response compatibility (SRC) effects. For **n**umber, a well-described SRC effect is the S**N**ARC, whereby responses with left hand to small numbers are faster and more accurate with the left hand, and to large numbers with the right hand. More recently, an equivalent effect in the **m**usical pitch domain, the S**M**ARC, has been demonstrated. Given the stated overlaps between the number and pitch domains, one relevant question is to what degree impairment in one domain affects the SRC effect in the other. Here I aimed to answer this by investigating the two effects (SNARC, SMARC) in lesion patients with numerical impairments, as compared to healthy controls. The results suggested a double dissociation between the two groups, with SNARC but not SMARC present in the controls, and SMARC but not SNARC present in the patients. Individual patients treated

as single cases confirmed this pattern, and also suggested that, within the number domain, not only SNARC but also performance on the number line task is outside of the normative control limits. This latter result speaks to the central role of number-to-space mapping ability within the broader realm of numerical cognition, and also to the more controversial topic concerning the importance (if not the necessity) of this concept in the existence of SNARC. Overall, these results constitute tentative evidence that SRC effects involving spatial responses to number and pitch respectively *can* dissociate in the numerically-impaired brain, which might suggest a deeper-level dissociation between the two domains also in the non-impaired brain, despite their irrefutable neural and behavioural overlaps. It seems therefore likely that, if future studies will approach the same question from the other end, patients with pitch impairments (amusia) would not necessarily show an impairment in the number domain. Further understanding the exact nature of the interaction between these different parietally-processed domains may inform rehabilitation strategies, who may want to target the affected domain either in isolation or in light of its interaction with "neighbouring" domains.

5.2. Introduction

5.2.1 Maths and music

Finding commonalities between the domains of maths and music is not a difficult task. Both have structure and periodicity (e.g., numeral systems and fractal structures in maths; phrasing and pulse repetition in music). Both have intuitive and embodied sensori-motor manifestations (e.g., finger counting and using an abacus in maths; beat-tapping and singing in music). And of course, it is ultimately complex mathematics that describes music – not just in terms of its compositional structure but also in terms of how its fundamental concepts of pitch and harmony, as well as the world's different musical tuning systems, are defined. Anecdotal reports linking maths and music abound, from the observations of the ancient Greeks and Chinese that consonant musical intervals are those for which the lengths of vibrating strings form simple numerical ratios to more recent observations by Vernon (1931, cited in Shuter-Dyson, 1968) of how a disproportionately large number of members of the Oxford University Music Club were scientists. At the same time, several studies suggest that musical (as opposed to other types of) training may benefit mathematical performance (Cheek & Smith, 1999; Gardiner, Fox, Knowles, & Jeffrey, 1996; Graziano, Peterson, & Shaw, 1999). Such studies, however, have little to say about whether this transfer takes place by

enhancing domain-specific skills used in maths or domain-general faculties such as attention. It is also possible that these maths-music concomitants are not mediated by cognitive processes at all, but simply by extraneous individual factors such as socio-economic status or level of motivation.

At the cognitive level, there is evidence of interference between simultaneous music and arithmetic processing. Hoch and Tillmann (2012) asked subjects to judge the correctness of a rule-based arithmetic series, while task-irrelevant chord sequences were presented as musical background. An interaction was found between the processing of the arithmetic and the chord series: specifically, verification of the correctness of the arithmetic series was slower when the final chord violated the harmonic expectancy created by the previously-presented chords. Also, this reaction time difference was greater for larger (more difficult) arithmetic problems. These effects prompted the authors to suggest that the processing of music and arithmetic structure share cognitive resources dedicated to structural and temporal integration, as previously suggested for language and music processing (Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009; Hoch, Poulin-Charronnat, & Tillmann, 2011; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Steinbeis & Koelsch, 2008) as well as for language and arithmetic processing (Baldo & Dronkers, 2007; Fedorenko, Gibson, & Rohde, 2007).

Furthermore, neuroimaging evidence exists that parietal structures typically involved in mathematical processing are also associated – functionally or structurally – with music processing or musical expertise. The bilateral IPL (inferior parietal lobule) is implicated in both number processing (Chochon et al., 1999; Sandrini, Rossini, & Miniussi, 2004) and pitch processing (Schmithorst & Holland, 2003). The IPS (intraparietal sulcus), one of the key "number" regions (see previous chapters), is also involved bilaterally in relative pitch judgements (Foster & Zatorre, 2010a), and while it is possible that IPS activity might have a role in the basic response selection and action perception processes common to both music and maths tasks (Göbel et al., 2004; Göbel & Rushworth, 2004), this common activation for judgements of relative number or relative pitch is still relevant to note. At the same time, individual differences in the *structure* of these cortices predict musical performance – e.g., grey matter density of the right SPL (superior parietal lobule) positively correlates with musical expertise (Gaser & Schlaug, 2003) – just like it does with maths expertise; see Study 1 of Chapter 2. On the other hand, grey matter density of the IPS – a region whose structure also differentiated between mathematicians and non-mathematicians in that study – correlates bilaterally with performance in

a relative pitch task (Foster & Zatorre, 2010b) and is increased in musicians compared with non-musicians (James et al., 2013).

In summary, there are reasons to believe a relationship exists between mathematical and musical cognitive processes, however the nature of the relationship is yet unclear, and characterising it as a whole is not straightforward. One possible approach is to focus on how the "building blocks" of maths and music interact, however while number is arguably the best candidate for the maths domain, choosing one for music involves considering a comparatively large number of dimensions, such as pitch, timbre, rhythm, metre, loudness, and so on. Pitch, however, appears, for several reasons, to be the most salient musical aspect to single out when investigating such a complex interaction. Like number, pitch can be construed as *magnitude*, either one that is analogue (the pitch of a sound can increase endlessly and in infinitely small steps, like the uncountable real number set, $\mathbb{R}$) or discrete (fixed pitches used in Western music, like the countable integer number set, $\mathbb{Z}$). Both pitch and number can be thought of as cyclical, in terms of steps of successive scales (e.g. the 12 notes of the standard chromatic scale) and the periodical digits of a numeral system (e.g. the 10 digits of base ten) respectively. Ultimately, of course, it is a number that describes a certain pitch (i.e. for a pure tone, its frequency in Hz).

Having established – for the purposes of this investigation – number and pitch as the essential dimensions of their respective domains, a way to probe their interaction is by comparing how each of them interacts with (or: is represented in terms of) a third other dimension. Among the possible choices for a third probing dimension, space seems to be among the better choices, for the reasons outlined in the next section.

5.2.2 Space as a medium for interaction

Evidence from several lines of research in neuropsychology and cognitive neuroscience strongly suggests that, rather than consisting of isolated specialised modules, the brain functions as a web of interconnected regions. While many of these regions have some degree of specialisation, this is not always exclusive for a single cognitive process, and vice-versa, a cognitive process might require more than one specialised region. Performance on tasks in which several dimensions are processed together may be indicative of shared mechanisms between those different processes. These shared mechanisms can be at different levels in the cognitive hierarchy. In turn, the existence of such shared mechanisms at a certain level may or may not reflect a physical overlap in neural populations within a single brain region. For instance,

Walsh's (2003) influential theory of magnitude brings neuropsychological, neurophysiological and behavioural evidence that magnitude, space and time are all linked by a common metric, and that judgements about any one of these dimensions will necessarily pass through a common stage of processing, performed in the inferior parietal cortex. By Walsh's account, it would be maladaptive for the developing brain *not* to use this common metric[15] for estimating time, space, and quantity, since the child can observe the covariance of these dimensions in the physical world. A later review proposed that in fact such common circuitry exists, more generally, for any sort of relative judgements made along a continuous dimension (Cantlon et al., 2009).

Common linguistic constructs suggest that many abstract concepts are cognitively represented in a spatial format. Such concepts include not just number (*low* or *high* numbers) and pitch (*low* or *high* tones), but also, for instance, time (*back* in time, looking *forward* to a future event), political orientation (*left*- or *right*-leaning) and social relationships (*close* friendships, *high* society). This is hardly surprising – as a continuous, multidimensional, linearly structured concept, space seems particularly well suited to represent other conceptual domains. In an experimental context where one of these domains acts as the nature of the *stimulus*, and the other domain as the modality through which a *response* to that stimulus is given/requested to, observing an interaction between the two - known as a stimulus-response compatibility (SRC) effect - can serve as a cue to the existence of at least one stage that is common in the processing of the two dimensions. Such pairwise interactions have been observed between all three dimensions of interest to this chapter (number, space and pitch), and each of them is discussed separately below.

5.2.2.1 Number and space

Several lines of argument can be brought to make the case for an intimate link between number and space in terms of how both are encoded and acted upon (e.g. in goal-directed movements) in the brain. Such links are the corner stone of the concept of embodiment in numerical cognition (see Fischer, 2012 for review). Anecdotally, reports exist of increased reliance on visuo-spatial strategies during mathematical reasoning by mathematicians and arithmetical prodigies (Hadamard, 1949; Scripture, 1891); this is confirmed

[15] Before the reader might start to feel frustration about the ambiguity of this term, it should be mentioned that a conceptual problem does remain with Walsh's proposal of a common magnitude system, in that the specific nature of the "common metric" remains unspecified. While it would be simplistic to interpret this commonality as a direct correspondence between how arbitrary units of measurement of the said dimensions (e.g., one second and one metre) are encoded, the "common metric" might perhaps be understood by analogy with the way in which the Weber fraction (*w*) and the numerical ratio effect (NRE) are both common metrics of the precision of the Approximate Number System.

by the questionnaire data from the mathematicians study in this thesis (Chapter 2). The simple presentation of a number, even when it is completely task-irrelevant, can cause spatial shifts in the allocation of visual attention (Fischer, Castel, Dodd, & Pratt, 2003). Deficits in both number- and space-related tasks are frequently observed simultaneously in patients with parietal lesions such as Gerstmann's syndrome (Gerstmann, 1940; Roux, Boetto, Sacko, Chollet, & Trémoulet, 2003). Neurophysiologically, the location of number-selective neurons ('numerons', Nieder et al., 2002; Nieder & Miller, 2004) coincides with that of neurons that are spatially selective (Stein, 1989; F. A. Wilson, Scalaidhe, & Goldman-Rakic, 1993). It is argued that numerosity is encoded by such numerons with increasingly higher noise and lower precision as magnitude increases (Nieder & Miller, 2003), which explains long-standing observations such as the distance effect (Moyer & Landauer, 1967), whereby the speed of a number comparison task is proportional to the numerical distance between them. The region of the parietal lobe where number and space appear to be the most tightly intertwined is the superior parietal lobule (SPL). Knops et al. (2009) used fMRI to record brain activity from this and other posterior parietal areas involved in eye movements, and used this data to train a multivariate classifier algorithm to infer whether subjects' gazes were moving left or right. Using data recorded from the posterior SPL (PSPL) alone, this training also enabled the classifier to generalise to an arithmetic task, where subtraction were distinguished from addition trials with a high degree of accuracy. The authors of the study suggested that the PSPL is involved not only in mental arithmetic problems but, more broadly, in a large variety of spatial tasks.

One of the most pervasive associations between number and space is hinted at by the intuitive model of the mental number line, which assumes numerical quantity is represented on an analogue scale that is inherently spatial in nature. Such a model provides a useful conceptualisation for explaining the distance effect as well as other effects such as the size effect (discriminability of two equally-spaced magnitudes decreases as the numbers become larger), all of which make an implicit assumption about logarithmic compression operating on the number line. While for most left-to-right writers, the orientation of the mental line follows a similar pattern (numbers increase rightwards), the shape and orientation of the mental number line leaves a lot of room for individual differences and idiosyncrasies (Fias & Fischer, 2005; Göbel, Shaki, & Fischer, 2011; Hubbard et al., 2005).

The final – and perhaps most famous – argument in favour of complex interactions occurring in the brain between number and space is the Spatial-Numerical Association of Response Codes effect (SNARC;

Dehaene et al., 1993). This effect indicates that, in either explicit (magnitude-related) or implicit (parity-related) judgements of numbers, larger magnitudes elicit faster lateralised responses with the right hand while smaller magnitudes elicit faster responses with the left hand. The SNARC effect has been widely replicated (Fias, Brysbaert, Geypens, & Ydewalle, 1996; Gevers, Caessens, & Fias, 2005; Keus, Jenks, & Schwarz, 2005; Mapelli, Rusconi, & Umiltà, 2003; Nuerk, Wood, & Willmes, 2005), although its size, i.e. the difference in reaction times between left and right responses for a given stimulus is typically only a few tens of ms. According to Walsh's "a theory of magnitude" (ATOM) model, numerical magnitude is but one of the types of magnitudes that can lead to preferentially-localised responses, such that the SNARC effect might in fact *"prove to be a SQUARC effect (spatial quantity association of response codes) in which any spatially or action-coded magnitude will yield a relationship between magnitude and space"* (Walsh, 2003, p. 487). Computational modelling of the behavioural effects of electrical brain stimulation have later added even further credibility to the ATOM model (Javadi, Brunec, Walsh, Penny, & Spiers, 2014).

However, just as Walsh's (2003) model was generalised by Cantlon et al. (2009; see above), the discovery of the SNARC was also suggested to have reflected a special case of what was later proved to be a much broader phenomenon. Thus, generalising even further, SNARC-like effects have also been observed with non-quantitative but ordered stimuli, such as letters of the alphabet, months of the year (Gevers, Reynvoet, & Fias, 2003) and days of the week (Gevers, Reynvoet, & Fias, 2004), and it has been suggested that it is an ordinal, rather than a cardinal (magnitude) code, that is critical for the spatial coding of stimuli to occur (Previtali, Hevia, & Girelli, 2010). In other words, the existence of an ordinal relationship for the class of stimuli appears to be necessary and sufficient, unlike the existence of a cardinal relationship. As a stimulus set, numbers just happen to have both relationships defined, and to have been the first stimuli to have been observed to elicit SNARC-like effects.

The SNARC effect is arguably the most investigated SRC effect in numerical cognition and has generated huge amount of interest in the field. Researchers have tried to manipulate several aspects of both the stimulus and the response, in the attempt to find, on one hand, just how much the effect can be generalised, and on the other hand what is the minimum set of stimulus and response characteristics for which the effect is still observed. For instance, the effect remains even if subjects are asked to respond with their arms crossed (Dehaene et al., 1993) and even if the numbers are presented in spoken rather than visual form (Nuerk et al., 2005). On the other hand, the effect was shown to disappear under working memory load

(van Dijck, Gevers, & Fias, 2009), and it has been suggested that the SNARC is driven by working memory (temporary position-space associations), rather than long-term memory (semantic representations) (van Dijck & Fias, 2011).

Although the SNARC is intuitively reminiscent of the concept of the horizontally-aligned mental number line, as Dehaene et al. (1993) initially proposed, the effect has been shown to exist also with vertically-mapped responses (Ito & Hatta, 2004; Schwarz & Keus, 2004), thus generalising the concept of the number line to a number map. Holmes and Lourenco (2012) proposed that the horizontal representation of numbers is the default (most salient) one, but that other orientations – such as vertical or egocentric (near to or far from oneself) – may be recruited when magnitude (and not just order) is a salient component of the representation. Taking this generalisation one step further, it was later suggested (Proctor & Cho, 2006; Santens & Gevers, 2008) that any analogue-scale (be it horizontal or vertical) is in fact not necessary to explain the effect, and that a more parsimonious model is the polarity correspondence principle. Rather than assuming a direct association between numbers spatially positioned on a line and a spatial response, the polarity correspondence principle posits that an extra step exists, in which numerical information is categorized before the activation of a response side. During this step, small (relative to the presented range) numbers are coded as "-" polarity and large numbers as "+" polarity; response location is similarly coded ("-" polarity for a left response and "+" for a right response), and faster responses are observed for corresponding polarities, thus explaining the effect. Lending support to this model is the fact that it can also account for the distance and size effects, without making use of the mental number line concept that (as mentioned earlier in this section) is usually employed to explain these effects. The authors of the model acknowledge that, "although it is possible that the use of a mental number line metaphor remains useful and that such a numerical representation may very well exist, this cannot be concluded on the basis of the SNARC effect." (Santens & Gevers, 2008, p. 269).

That the size of the SNARC effect might predict performance in higher-level numerical capacities such as arithmetic – and, consequently, that the effect might be altogether absent in dyscalculia or acalculia – is an appealing hypothesis. However, to my knowledge, so far no studies have investigated the SNARC effect in populations with purely-numerical impairment, and the only adequately-powered study using a healthy population found no difference in SNARC effect between arithmetically skilled and non-skilled subjects (Cipora & Nuerk, 2013). Beyond simply failing to reject the null hypothesis of SNARC slopes not being

dependent on measures of arithmetic performance, subsequent Bayesian analyses performed by the authors showed positive evidence of a true null effect, as opposed to just a statistical power problem. Still, the relationship between SNARC and maths abilities might be qualitatively different once the latter are impaired, and so it was one of the aims of the current study to ascertain whether the effect is present in patients with numerical impairments.

5.2.2.2 Pitch and space

It is not just number that has an intuitive spatial representation. A common metaphor for pitch is that of a spiral, the plane dimension of which represents chroma, i.e. the roles played by different notes (e.g. C_4 and D_4) in a given tonal context (e.g. the C major scale); and the vertical dimension of which represents height, i.e. the fundamental frequency that the pitch is perceptually represented by, e.g. C_4=262 Hz (Ueda, 1987); see Figure 37. Chroma is a periodic dimension, as the roles played by increasing notes (i.e. pitch chromas) start to repeat cyclically once the next scale is reached; while height is a monotonic dimension, because pitch increases continuously as fundamental frequency increases. This two-dimensional model of pitch finds support in functional magnetic resonance (fMRI) data suggesting that pitch chroma modulates activity in the anterior secondary auditory cortex while pitch height does so for the posterior part of this structure (Warren, Uppenkamp, Patterson, & Griffiths, 2003).

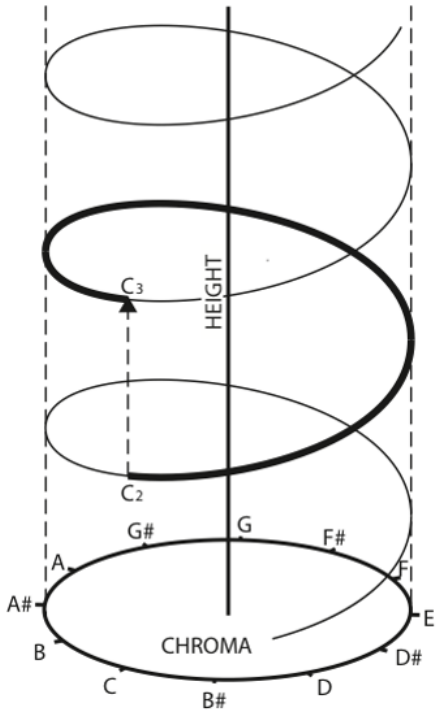


Figure 37: The helicoidal (spiral-like) visualisation of pitch. Chroma increases anticlockwise in the horizontal plane and is periodical, while height increases bottom-to-top and is monotonic (non-periodical). Figure modified from (Shepard, 1982).

At the same time, the metaphorical language used to describe stimuli along a given dimension can often reflect an underlying conceptual mapping (Lakoff & Johnson, 2008). The use of the syntagms "*low*/*high* pitch", pervasive across several languages, suggest that such an underlying mapping may exist between pitch and space, although a "chicken and the egg" problem in determining cause and effect between this association and the language metaphors is also possible.

To see if pitch, like number, interacts with space in the form of an SRC effect, Rusconi et al. (2006), in a series of SNARC-like experiments, asked musicians and non-musicians to give spatial (vertical or horizontal) responses to implicit or explicit judgements of pitch. When the pitch of a tone had to be deemed either higher or lower than that of a previous reference tone (explicit task), correct answers to low pitches were faster when responded to with the response spatially located on the left or at the bottom, whereas high pitches were associated with right/top responses. The same associations were found, in musicians only, when the task required recognition of the timbre (instrument) of an isolated sound (implicit task). By analogy with SNARC, this effect was called the Spatial-Musical Association of Response Codes (SMARC) effect, and was later replicated in several other studies (Beecham, Reeve, & Wilson, 2009; Y. S. Cho, Bae, & Proctor, 2011; Lidji, Kolinsky, Lochy, & Morais, 2007).

Another further association between the domain of pitch and that of space comes from a study suggesting that amusia, a congenital or acquired deficit in music perception, might be associated with deficits in spatial processing (Douglas & Bilkey, 2007). Evidence for this is equivocal however, as attempts to replicate this effect have been partially or completely unsuccessful (Tillmann et al., 2010; Williamson, Cocchini, & Stewart, 2011).

5.2.2.3 Number and pitch

Most studies investigating the representation of number seem to use space as the default response medium, while other response modalities have received far less attention. Arguably, however, a deeper inquiry into the role of space for the representation of number also requires paradigms where space is not forced upon the subject as a representational tool. One such paradigm was used by Marghetis et al. (2011): subjects judged the magnitude of digits 0-9 and responded with high- or low-pitched vocal responses to numbers higher or lower than 5, with mapping order counter-balanced between subjects. An interaction was observed between response pitch and number magnitude, analogous to the interaction between space and number; namely, subjects were faster to produce "high" pitches in response to large numbers and low pitches to small numbers, an effect the authors referred to as the SNAP (Seeing Number as Pitch) effect. The strength of the SNAP effect was unrelated to the strength of the SNARC effect. The fact that, when given the opportunity to respond in other modalities (i.e. non-spatially), subjects showed such an interaction between number and the pitch modality, suggests that the representation of number is not inherently or uniquely spatial. Marghetis et al. (2011) therefore argued that, despite the relative primacy of space as a representational medium, it could be that effects like the SNARC and the SMARC are but special cases of a more general cognitive capacity to think about or represent one domain in terms of another.

By analogy with SNARC, the existence of a SMARC effect suggests that the spatial response can be primed by the stimulus dimension, even when that dimension (pitch) is irrelevant to the (timbre judgement) implicit task; this suggests that the extraction of magnitude information is automatic for both number and pitch. It is also possible that the stimulus dimension (along which explicit judgement are to be made) does not need to be magnitude (i.e. number magnitude or pitch height), but can instead simply be stimulus salience, e.g. how large, long, bright or loud the stimulus is. This possible explanation of both SNARC and SMARC is suggested by a study that found that responses with a right/upper key are faster than those with a left/lower key for stimuli whose relevant feature was more salient (Y. S. Cho & Proctor, 2003).

Additional evidence for a shared representational stage for number and pitch comes from studies that have found that the distance effect, whereby reaction time in a symbolic number comparison task decreases linearly with increasing magnitude distance between the two stimuli, also exists when manipulating the pitch of auditory stimuli (R. Cohen Kadosh, Brodsky, Levin, & Henik, 2008; Elkin & Leuthold, 2011). The dependence between RTs and pitch distance was linear, despite the two-dimensional nature of pitch.

5.2.3 The present study

To my knowledge, the only other study to have examined both SNARC and SMARC in the same group of subjects did so in order to ascertain whether the way that number and pitch respectively are represented spatially stems from the same or different (domain-specific) spatial profiles. Using cluster analysis, Beecham, Reeve & Wilson (2009) identified subgroups of individuals based on their response patterns and determined that these patterns were largely independent between the two domains, suggesting it is unlikely that the *same* spatial representation processes underpin both mathematical and musical judgments. These findings challenge a spatial-mediation account, one that would assume there is a reliable and coordinated mapping between both number and space, and space and pitch.

With this study, I aimed to probe commonalities in the way that numbers and pitch are represented spatially in patients with numerical impairments versus healthy controls. Specifically, I aimed to test at what level these impairments might also affect the processing of dimensions believed to share cognitive mechanisms. I did this by investigating effects which are believed to rely on an intact representation of the relevant dimension. An improved understanding of the degree of overlap in terms of spatial (and/or mental) representation of number and pitch might have clinical implications for the conditions in which the normal processing of these two dimensions becomes impaired, i.e. acalculia and amusia, respectively. For example, whether these "specific" deficits result from the dysfunction of shared or domain-specific processes has implications for the selection of rehabilitation strategies.

One method to explore the extent to which the number and pitch dimensions interact is to observe how their explicit or implicit (via space) judgement differs in patients in which the normal processing of one of these dimensions is impaired. Long before the advent of functional neuroimaging (e.g. fMRI) and transcranial magnetic stimulation (TMS) informed scientists about the recruitment and necessity of different brain areas to different mental processes, lesion studies using neuropsychology in both animals and humans

have, for centuries, been building an increasingly complex map of the brain's functions, one piece of the puzzle at a time. Such approaches often demonstrated a brain region's necessity in a particular task, such as, famously, the necessity of Broca's area in speech production or that of the hippocampus in memory coding. Beyond informing us about a region's necessity for a given process, this approach can also inform us about the degree to which two such processes share common resources in the brain, and whether they in fact dissociate, and if so, at what cognitive stage they do so.

Here, 7 patients were selected based on the main criterion of numerical impairment as assessed in a standardised battery of neuropsychological tests. Performance on the SNARC and SMARC paradigms, as well as on several others including the mental number line, has been recorded from the patients and from a group of controls. The aim was to, first of all, evaluate whether the patients show a changed SNARC consistent with their mathematical impairment, and second, whether similar changes existed for the SMARC. The relations between the SNARC and SMARC effects were examined at a group level and also in terms of individual cases who showed particular dissociations between the procedures.

5.3. Methods

5.3.1 Subjects

5.3.1.1 Control group

The control (normative) group consisted of 12 healthy right-handed volunteers (6 males and 6 females, mean age 24.17 ± 3.29 years), all postgraduate students at the University of Oxford. In order to control for level of musical experience in the analysis of the sound-based behavioural tasks, a compound measure of musical training was computed for each subject. This was expressed as the total number of hours of musical training (instrumental or vocal), based on the subjects' self-reported description of their musical history, including average number of practice hours per week. The level of musical education within the control group varied, ranging from none to several years of instruments training. The distribution was positively skewed, as half of the controls had virtually no musical experience whatsoever. The average number of hours of musical training in the control group was 775 ± 1362, which was included as a covariate in all analyses.

5.3.1.2 Patient group

Patients were selected from the patient pool of the Cognitive Neuropsychology Centre of the Department of Experimental Psychology, University of Oxford. The main selection criterion was numerical impairment, which was mainly assessed using patients' scores on the Calculation component of the Birmingham Cognitive Screen (BCoS) battery (Humphreys, Bickerton, Samson, & Riddoch, 2012). This test consists of four easy arithmetic problems consisting of additions, subtractions and multiplications. Patients with scores of 2 or less (out of a total of 4) were selected in the first instance. Subsequently, given the need for patients to be able to follow instructions in an experimental setting and to correctly recognise numbers in the numerical tasks used in the current study, a further selection criterion was applied in order to exclude ineligible patients, defined as those having low scores on other relevant BCoS components. The cut-off scores for each component (detailed in Table 15) were chosen subjectively, in an attempt to exclude as few patients from the available database as possible, while still not compromising the anticipated quality of the data collection process.

Table 15: Exclusion criteria for patient selection. The short list of patients with numerical impairments was further trimmed by excluding patients with scores lower than an arbitrary threshold on these tests.

BCoS component	Ability tested	Max score	Patient excluded if score smaller than
Orientation: Personal Information	awareness of one's name, age etc	8	8
Picture Naming	correctly naming objects or animals depicted in pictures	14	9
Birmingham Rule Finding and Switching Test	being able to describe the relative trajectory of a dot in 2D space	18	9
Number Reading	correctly naming printed numbers	9	8
Instruction Comprehension	overall ability to follow instructions during the administration of the BCoS	3	3

After applying these criteria to the patient database, seven patients were selected for participation. Six patients were male and one female. The average age was 52.57±11.89 years. Table 16 contains a summary of the basic information concerning these patients, and the Appendix of this thesis (section 8.1, page 188) contains supplementary information, namely the patients' BCoS summary charts and, where available, MRI imagery of their lesions. None of the patients had any musical training. Five out of these 7 patients also completed the Numerical Operations component of the Wechsler Individual Achievement Test (WIAT-II UK version) battery of tests (Wechsler, 2005), which involves writing down answers to simple and relatively more

advanced arithmetic and algebra problems printed on a sheet of paper. When using a time limit based on the average completion time for control subjects, all patients had standardised scores that were at least two standard deviations below the population averages in the normative tables corresponding to their respective ages. Not all patients were able to complete the entire battery of tasks administered in this study. In particular, one patient did not complete the SNARC task, which was only available for the remaining 6 patients.

Table 16: Basic patient information. Handedness and lesion information for patient P06 was not available.

Patient code	Gender	Age	Handedness	Lesion hemisphere
P01	m	32	R	L
P02	f	54	R	L
P03	m	58	R	R
P04	m	67	R	R
P05	m	61	R	bilateral
P06	m	42	(unknown)	(unknown)
P07	m	54	R	L

5.3.2 Behavioural tasks

5.3.2.1 SNARC

The spatial-numerical association of response codes (SNARC) task was created based on specifications from the original Dehaene et al. (1993) study. The task was programmed and run in E-Prime (PST Inc.). Each trial consisted of the presentation of a digit from 0 to 9, which was kept on the screen for 5,000 ms or until a response was given. A fixation cross was displayed immediately after a response was given, until it was replaced by the stimulus for the next trial, 1,000 ms later. Subjects had to respond whether the number was odd or even, i.e. a parity judgement that only bears an implicit judgement of magnitude. For each trial, the number presented was selected at random (without replacement) from the 0-to-9 number set. The random sequence was generated uniquely for each run. Subjects responded "odd" or "even" using keys horizontally placed on a standard PC keyboard. In one mapping, the key "Z" was used for odd responses and the key "M" for even responses, while for the other mapping, the roles were reversed between these keys. Each subject performed the task once for each mapping. In each of the two runs, there were 4 blocks of 50 trials each, during which each of the 10 stimuli (0-9) was repeated 5 times, for a total of 200 trials per run (mapping). The order of the two mappings was counter-balanced between subjects, separately within each group (patients and controls). Before each of the two blocks there was a practice block containing 5 trials with stimuli

selected randomly which used the mapping of the current run. These practice trials were not included in the analysis. No feedback was given. RTs were measured from the onset of the visual stimulus to the response keypress. The need to respond both quickly and accurately was emphasised in the instructions. Subjects were seated in front of a 19" monitor, at a viewing distance of approx. 60 cm. Stimuli were displayed centrally on the screen and covered a visual angle of approximately 1.8° vertically and 1.2° horizontally (degrees of visual angle).

5.3.2.2 SMARC

As has been discussed in this chapter's Introduction, the two stimulus-response compatibility effects for number and pitch, SNARC and SMARC, can be demonstrated with paradigms using various parameters. These parameters allow the investigation of, for instance, horizontal or vertical spatial associations, various formats of the dimension in question, and implicit or explicit judgements. In the current study, the soundest choice regarding these parameters might appear to be using consistent parameters across the number and pitch tasks used to elicit evidence of their spatial associations. However, given previous researchers' difficulty in replicating the SMARC effect (Vicky Williamson and Lauren Stewart, personal communication), I chose to sacrifice the comparability of task parameters for the sake of maximising the chance of actually eliciting these effects. For that reason, the SNARC and SMARC paradigms were each run with the set of parameters that have previously led to the highest number of successful elicitations. Specifically, while the SNARC was run with the standard parameters – implicit number judgement (parity) and horizontal mapping – the SMARC was run as an explicit pitch judgement with diagonal (lower-left to upper-right) mapping. This is based on previous studies with non-musicians that have found the SMARC effect more reliably with these parameters as opposed to others, e.g. implicit rather than explicit pitch judgement, and purely horizontal or purely vertical as opposed to diagonal mapping (Beecham et al., 2009; Lidji et al., 2007; Rusconi et al., 2006).

The spatial-musical association of response codes (SMARC) task was created based on specifications from the original Rusconi et al. (2006) study. The task was programmed and run in Matlab (Mathworks Inc.), using the Psychophysics Toolbox. Each trial consisted of the presentation of two pure tones of different frequencies. Each of the tones was 1,000 ms in duration, and the inter-stimulus interval (ISI), i.e. the interval between the offset of the first tone and the onset of the second tone, was 500 ms. Tones were presented binaurally, through a pair of Sennheiser HD449 headphones. Subjects had to respond to whether the second tone was lower or higher in pitch than the first one. Since the sounds were pure tones (sinusoids), their pitch

was determined by a single frequency value. For the first tone of the interval (the reference tone), this frequency was always fixed at 262 Hz. For the second tone of the interval (the probe), a frequency was chosen pseudo-randomly among the values 165, 185, 208, 233, 294, 330, 370, 415 Hz (Rusconi et al., 2006), with the constraint that the same frequency could not be selected for two consecutive probes. The same pseudo-random sequence was used for all subjects. Subjects responded "lower" or "higher" using keys on a standard PC keyboard. Specifically, two mappings were used for each subject: a congruent mapping and an incongruent mapping. The concept of congruency refers to the hypothesis – presented in the Introduction – that low pitches tend to be associated with the left (or bottom) half of space. For this reason, the congruent mapping involved a key located in the lower-left corner of the keyboard, "Z", pressed with the left hand to answer "lower"; whereas an upper-right corner key, "0", had to be pressed with the right hand to answer "higher". The incongruent mapping had the roles of these two keys reversed. Each subject performed two blocks of trials, one using a congruent mapping and the other using an incongruent mapping. The order of the two blocks (mappings) was counter-balanced within each group (i.e., separately for patients and controls). Each block contained 40 trials, meaning that each of the 8 probe frequencies was compared to the reference for a total of 5 times per block. Before each of the two blocks there was a practice block containing 5 trials selected randomly from the list of trials of the coming block, and using the same mapping (congruent or incongruent). The practice trials were not included in the analysis. During the trials, after the instructions had been presented on the screen and orally by the experimenter, the computer monitor was set to a black background and only contained a fixation cross and correct/incorrect feedback for each trial. RTs were measured from the onset of the second tone (probe). Loudness was set a level that was sufficiently high but still comfortable, and this level was kept constant for all subjects tested, both for the SMARC and for the pitch detection threshold task. The need to respond both quickly and accurately was emphasised in the instructions, which appeared to be well comprehended by both the controls and the patients.

5.3.2.3 Pitch change detection task

Thresholds were measured for the detection of a change in pitch of a pure tone, in either direction. The paradigm involved a three-interval two-alternative forced-choice procedure in which the listener is presented with three auditory stimuli and has to decide which one of them contains a pitch glide (Liu, Patel, Fourcin, & Stewart, 2010). Subjects were informed that the correct answer was randomly placed either in the first or the last (third) interval. Stimuli in each of the three intervals were pure tones (sinusoids), 600 ms in

duration, with a 600 ms interstimulus interval (ISI). The target was a frequency glide (250 ms steady-state onset, 100 ms glide, and 250 ms steady-state offset), logarithmically centred on 500 Hz. Non-targets were steady-state tones of 500 Hz of the same length. The pitch glide was random in direction (decreasing/increasing), while its size (amount of drop/rise in pitch) was adaptively controlled using a staircase algorithm: following an incorrect response, the glide size was increased in order to make the task easier (a step known as a "reversal"). The procedure terminated following 10 reversals and thresholds (known as the just-noticeable difference, or JND) were computed as the median value of the glide size (expressed in semitones) of the last five reversals.

This task was used to ensure that both the patient and the control group have a normal ability to distinguish sounds of different pitches, a requirement for the SMARC effect to be defined. Also, it is possible that acuity in discriminating close pitches – indexed by e.g. the JND between two pitches – may be similarly related with regard to various measures of musical skill, just like acuity of the approximate number system (ANS), indexed by e.g. the Weber fraction, may underlie and predict performance in higher-level arithmetic (see previous chapters). The correlations between JND and SMARC size could, however, ultimately not be computed due to the sample size for this study eventually being smaller than that which was initially anticipated.

5.3.2.4 Number line

To measure the accuracy of mapping number onto the one-dimensional number line, the same task was used as for Study 1 (Chapter 2). In short, subjects had to approximate the position of 30 successive numbers ranging between -1000 and 1000 on an unmarked number line, with only the values at the edges of the line (-1000 and 1000) displayed. The deviation between the subject's mapping of each number and the number's actual location was computed for each number, and the mean taken across all 30 numbers, for each subject.

5.3.2.5 Mental rotation task (MRT)

To measure subjects' visuo-spatial abilities in mental rotation, the MRT-A version of the task was used (Peters et al., 1995), as used in Study 1 (Chapter 2). In short, subjects were shown an image of an abstract 3-dimensional object and had to indicate which two out of four attached images represented spatial rotations

of the initial target image. A point was awarded when both figures were correctly identified. There were a total of 24 trials in the task. This task could only be administered to N=3 patients.

5.3.3 SNARC/SMARC analyses

Since both SNARC and SMARC produce comparable output (accuracy and RTs for each of several stimuli varying across the relevant dimension), the data for both were analysed similarly, using regression of repeated-measures data, described initially by Lorch & Myers (1990). Initially, differential measures were computed for the median RT and mean accuracy of each stimulus, by taking the difference between the mean/median of a stimulus from trials where the correct answer would have been responded to with the right hand, and from trials where the correct answer would have been responded to with the left hand. These differential measures of RT and accuracy for each stimulus are referred to as dRT and dAcc respectively. A positive dAcc thus indicates a right hand advantage for the stimulus, as responses are on average more accurate with the right hand (Acc_R-Acc_L>0), whereas a positive dRT indicates a left hand advantage, as responses are on average faster with the left hand (RT_R-RT_L>0). Stimulus magnitude (number 0-to-9, in the case of SNARC; and log-transformed pitch in Hz, in the case of SMARC), and order of first mapping (coded as -0.5;+0.5 or vice-versa) were used as regressors. A linear regression model was then applied to obtain the coefficients of best fit for each subject of each group (patients and controls), for each task (SNARC and SMARC) and for each measure (dRT and dAcc). These regression coefficients are referred to as SNARC and SMARC slopes, respectively for the two tasks, or simply as the "effect slope".

Working with the measures computed as above, the procedure used for data analysis was adapted from the standard imposed by previous SMARC (Lidji et al., 2007; Rusconi et al., 2006) and SNARC studies (Dehaene et al., 1993; Fias et al., 1996; Holmes & Lourenco, 2012), which is similar for the two effects. Namely, an effect is considered to have occurred if at least one of the following two conditions is met:

1) Criterion based on difference between dRT averages for low and high stimuli. This involves the following steps:

- compute the dRT for each stimulus (number/pitch), as explained above
- compute an average dRT for responses to low numbers/pitches by pooling across the first half of the stimulus set, i.e. for the *low* numbers (0, 1, 2, 3 and 4) or for the *low* pitches (165, 185,

208 and 233 Hz); and the same for *high* numbers (5, 6, 7, 8 and 9) or pitches (294, 330, 370 and 415 Hz)

- the effect is said to exist if there is a significant main effect of stimulus size (low vs high), indicating greater values of dRT for low than for high stimuli

In the current case, since the aim is to not only see if the effect exists for each group but compare the effect across groups, the last step of the analysis here would involve the addition of the Group between-subjects factor, making the repeated-measures ANOVA into a mixed-model ANOVA with between-subjects factor Group and within-subjects factor Stimulus Size. However, due to the very large differences in variances across the two groups, a standard (parametric) ANOVA checking for a Group × Stimulus Size interaction would not be appropriate. Therefore, this analysis was replaced with 1-sample t-tests that compared, within each group (P/C), within each effect (SMARC/SNARC), within each measure (dRT/dAcc) and within each stimulus size (low/high), whether the mean dRT was significantly different from zero. The current literature defines the effect based on the left-hand advantage for small numbers/pitches (and right-hand advantage for high ones) expressed as decreased RT with the preferred hand (hemispace), e.g. a *positive* dRT for *low* stimuli. This is presumably because with healthy subjects, accuracy for these rather simple tasks is close to ceiling, leaving little variance to be analysed with e.g. ANOVAs. Since here I am including a patient group, who were expected to perform well below ceiling in terms of accuracy, an extension to the definition of the effects was used in that dAcc was a priori defined as an acceptable measure in terms of which to observe the effect. Namely, I was prepared to accept as evidence for an effect if e.g. the left-hand advantage for *low* numbers was visible as a *negative* dAcc, alternatively to its expression as a *positive* dRT as explained above. This conjecture is, of course, assuming that the data indicates no evidence for a speed-accuracy trade-off having occurred (Pachella, 1974), which was indeed suggested by their non-significant correlation. In summary, the analyses were done on both median dRT and mean dAcc values. Nevertheless, I deemed any effects based on dRT, as in previous literature, to be preferable.

2) Criterion based on non-zero regression slopes. In this approach, a regression slope is calculated for each subject, for each effect (SMARC/SNARC) and for each measure (dRT/dAcc), as explained at the beginning of this section. As explained above, an effect is indicated by the presence of a negative slope of the regression line fitting dRT values against number/pitch magnitude, or, alternatively, a positive slope for dAcc. This can be tested using 1-sample t-tests that compare each of these measures (e.g. regression slope for

SNARC dRT) against zero. An effect has occurred if the test is significant for dRT fits (indicating a negative slope) and/or for dAcc fits (indicating a positive slope). It is this approach that was the one more suited for a between-groups comparison, as the presence/magnitude of the effect was described, for each subject, by a single number (the regression slope, either the one for dRT or for dAcc).

A further analysis was done by comparing each of the patients to the control group as individual single-cases, as opposed to a comparison between the two groups. The Crawford-Howell t-test was used for each comparison (Crawford & Howell, 1998; available as an online calculator at http://homepages.abdn.ac.uk/j.crawford/pages/dept/SingleCaseMethodsComputerPrograms.HTM)

5.4. Results

Given the low sample size that prevented a correlation-based approach, the results of all tasks except for the SNARC and SMARC were scrutinised with the sole aim of identifying patients who are outside of the control limits. Patients who differed significantly from controls on the number line task are presented in Table 19 and discussed later. For all patients and controls except for one (patient P04), the just-noticeable difference (JND) thresholds produced by the pitch change detection task were within the 95% confidence interval for the controls used in the original study (Liu et al., 2010). The mean SMARC accuracy of patient P04 was above chance (64%), which suggests that either he had not comprehended the instructions for the pitch threshold task and thus performed it at chance level, or that his hearing is indeed impaired in terms of pitch discrimination, but not to such an extent as to prevent him from discriminating the reference and the probe tones in the SMARC, which were well separated in pitch. Patient P04's SMARC data was, therefore, kept in the analyses described below.

For the group-level analysis of SNARC and SMARC data, two different approaches were used, as explained in the previous section. Approach 1, which compared the mean dRT and dAcc of each group and of each effect, separately for the low and high stimuli, revealed that the patients, as a group, did not show a SNARC effect but did show a SMARC effect. The opposite pattern was observed for controls, who showed a SNARC but no SMARC. Both results were obtained based on dRT measures. Additionally, the controls' SNARC effect was also marginally significant in terms of dAcc. Table 17 contains the results of these statistical tests.

Table 17: Results of 1-sample t-tests comparing the median dRT and mean dAcc of each group, task and stimulus size against zero. Significant results are in bold. Descriptives (mean ± standard deviation) are also added for each measure.

Group	Task	Measure	Stimuli	μ±σ	1-sample t-test against zero		
					t-value	df	p
P	SMARC	median dRT	low	213.56±173.92	3.24876	6	**0.02**
			high	-171.7±147.63	-3.07703	6	**0.02**
		mean dAcc	low	-0.12±0.3	-1.06667	6	0.33
			high	0±0.16	0.00000	6	1.00
	SNARC	median dRT	low	202.33±381.3	1.29979	5	0.25
			high	-147.81±176.79	-2.04796	5	0.10
		mean dAcc	low	-0.22±0.41	-1.34959	5	0.24
			high	-0.06±0.4	-0.34595	5	0.74
C	SMARC	median dRT	low	60.14±147.89	1.40873	11	0.19
			high	-25.54±131.75	-0.67157	11	0.52
		mean dAcc	low	-0.03±0.07	-1.25357	11	0.24
			high	-0.02±0.05	-1.44887	11	0.18
	SNARC	median dRT	low	19.84±21.79	3.15468	11	**0.01**
			high	-29.13±36.7	-2.74920	11	**0.02**
		mean dAcc	low	-0.03±0.05	-2.15003	11	0.055
			high	0.02±0.01	6.63325	11	**0.00**

Relevant to the understanding of approach 1, response profiles (plots of dRT and dAcc against stimulus number) for each patient, overlapped on the means of the normative control group, are presented in figures found in the Appendix (section 8.2, page 194), separately for the two tasks (SNARC and SMARC).

Approach 2, which compared the regression line slope for each effect against zero, confirmed that the SNARC occurred for controls (in terms of both dRT and dAcc), but did not confirm the SMARC for patients.

Table 18: Results of 1-sample t-tests comparing all effect slopes against zero. Significant results are in bold.

Group	Task	Slope for	1-sample t-test against zero		
			t-value	df	p
P	SMARC	RT	1.21811	6	0.268903
		Acc	-0.49028	6	0.641345
	SNARC	RT	-1.61431	5	0.167380
		Acc	1.28769	5	0.254229
C	SMARC	RT	-0.33857	11	0.741309
		Acc	0.37929	11	0.711693
	SNARC	RT	-3.18550	11	**0.008676**
		Acc	2.24251	11	**0.046496**

The Crawford-Howell t-tests comparing individual patients to the entire control group revealed, for the RT-based SNARC as well as for the number line task, significant differences between the control group and patients P04 , P05 and P07 (see Table 19). Of these, P04 and P05 were not also significantly different on the SMARC, while P07 was. However, this latter patient appeared, during testing, to be generally impaired, presumably due to his lesion having taken out much more than just the putative common metrics for number and pitch. For this reason, the individual comparison for this patient was not taken into consideration in the interpretation of the results. Table 19 also shows that the three patients for which MRT data was available (P01, P03 and P06) are within normative limits on this task.

Table 19: P-values for the Crawford-Howell t-test used to compare scores between patients – considered individually (as single cases) – and the control group. The scores being compared were the SNARC&SMARC regression slopes, mean deviation in number line accuracy, and overall MRT score. Hyphens represent missing data. Bolded figures denote a significant difference from the normative group.

Patient	SMARC		SNARC		Number line	MRT
	RT	Acc	RT	Acc		
P01	0.93	0.13	0.98	0.82	0.58	0.25
P02	0.42	0.55	-	-	0.58	-
P03	0.83	0.92	0.54	0.26	-	0.54
P04	0.63	**E-05**	**0.00**	**0.00**	**0.00**	-
P05	0.83	0.29	**0.04**	0.49	**0.00**	-
P06	0.08	0.92	0.73	0.55	0.43	0.42
P07	**2E-05**	**0.00**	**0.00**	0.38	**0.00**	-

Figure 38 shows, for both SNARC and SMARC, the regression slope for each patient, compared to the mean of the control group.

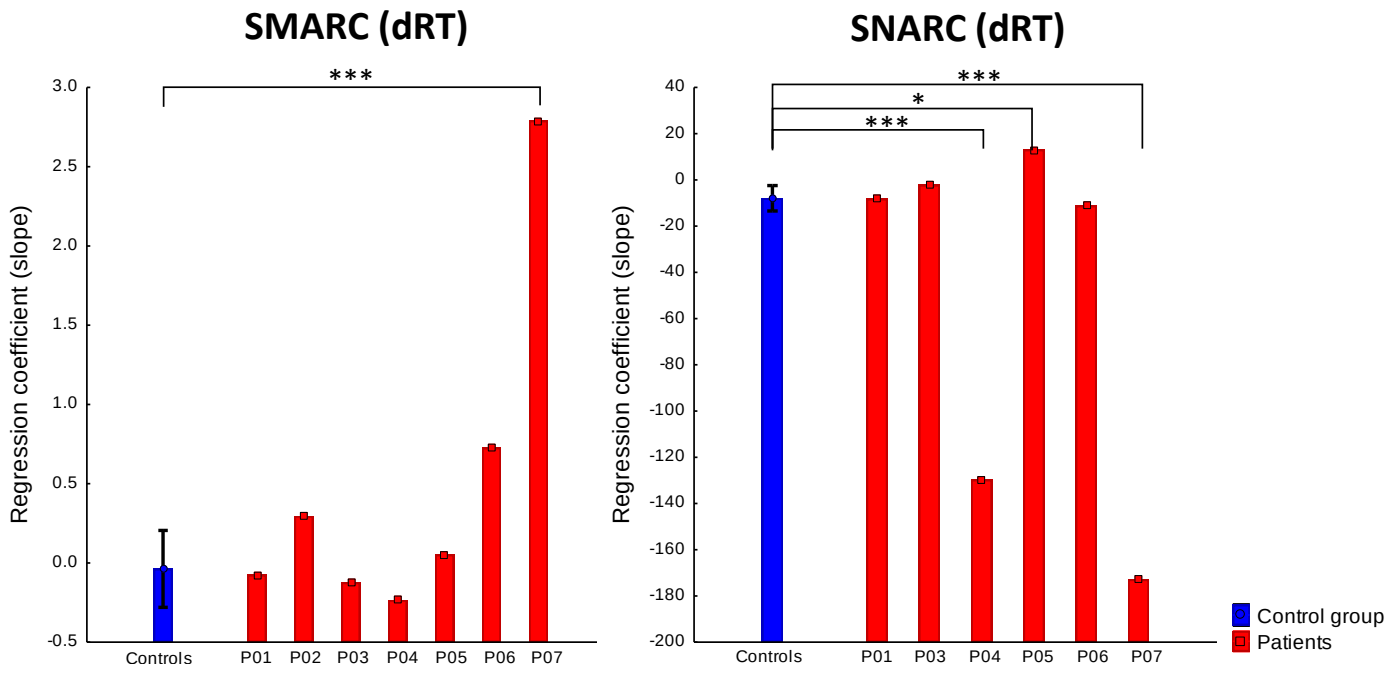


Figure 38: Regression slopes for SMARC and SNARC, computed in terms of dRT. Individual slopes are shown for patients with data on that task, while the controls are shown as a group average. Error bars indicate 95% confidence intervals and significance asterisks are based on the Crawford-Howell t-test.

5.5. Discussion

There is converging evidence for neural mechanisms in the parietal lobes that are shared between number and other dimensions, such as time (Walsh, 2003), space (Knops et al., 2009) and the pitch of musical sounds (Foster & Zatorre, 2010a). This study aimed to investigate a specific category of effects believed to arise due to the existence of such shared mechanisms between number and musical pitch. By looking at how each of these dimensions interacts with space (as part of two well-described effects known as SNARC and SMARC) in patients with disrupted numerical processing, we can better characterise each dimension's necessity in creating these interactions at the meeting point between the three dimensions.

Here, I found that impairment on number processing (including calculation), instantiated by the lack of a SNARC effect in the patients group, dissociates from a homologous impairment in pitch processing, whose absence was instantiated by the presence of the SMARC effect within that group. The controls showed the opposite pattern, manifesting the SNARC but not the SMARC. The dissociation within the controls was confirmed by two separate approaches, while in patients it only came about from the first approach. The outcomes of the two approaches used are summarised in Table 20.

Table 20: Summary of the two approaches used in assessing the existence of SNARC and SMARC effects in patients and controls. Effects are computed on the basis of dRT.

Analysis approach	Controls		Patients		Overall outcome
	SMARC	SNARC	SMARC	SNARC	
Approach 1 (low/high stimuli)	✗	✓	✓	✗	Double dissociation
Approach 2 (regression slopes)	✗	✓	✗	✗	Single dissociation

When comparing patients to controls not as two groups but as a series of single-case comparisons against the normative group provided by the controls (Crawford & Howell, 1998), the pattern of patient results previously found at the groups level – SMARC in the absence of SNARC – was shown by two individual patients. These two patients seemed to have driven the groups-level results, which lose significance when the two patients are excluded from the analyses, in both approaches. The fact that there are more (specifically, four) patients who do *not* show the said dissociation in the single-case approach must not be taken as stronger evidence for an association (rather than a dissociation) between impairments on the two domains, but rather as an equal number (four) of null results. It might be said that, qualitatively, the two patients who show the dissociation give us just as much information about the number-pitch relation as the four patients who do not (Gallistel, 2009) – although this statement needs to be qualified by the statistical caveat regarding the SNARC&SMARC reliability discussed in the next paragraph.

The controls' lack of a SMARC effect, on the other hand, cannot reliably be taken as an indication of a pitch impairment, but rather of the less robust nature of the SMARC as opposed to the SNARC. Indeed, the SNARC has been replicated many times (e.g., Fias et al., 1996; Gevers et al., 2005; Keus et al., 2005; Mapelli et al., 2003; Nuerk et al., 2005), and a meta-analysis by Wood et al. (2008) showed that about 70% of subjects show a regular SNARC effect. On the other hand, the younger SMARC has, to date, been much less replicated across labs (Beecham et al., 2009; Lidji et al., 2007; Rusconi et al., 2006), and some researchers have even been unable to replicate the effect within the same lab (Vicky Williamson and Lauren Stewart, personal communication). Also, previous studies using both SNARC and SMARC separately have shown substantial variability in performance across individuals, evident from large variance estimates (Fias et al., 1996; Gevers et al., 2005; Keus et al., 2005; Rusconi et al., 2006). This suggests that aggregating data from SNARC and SMARC measures, as has been done here, might mask individual differences in underlying cognitive representations. For these reasons, the double dissociation between patients and controls suggested by

approach 1 is more meaningful as a demonstration of the confirmation of the single dissociation in patients suggested by approach 2.

The two patients driving the group-level effect were also outside of the normative control limits in terms of scores on the number line task. This is reminiscent of the salience of the mental number line concept in explaining the SNARC effect (Dehaene et al., 1993), despite the fact that more recent reports suggest necessity only from the part of working memory, and not also of the number-to-space mapping of the number line (van Dijck & Fias, 2011). On the other hand, the SNARC literature tells us little about whether the effect survives in subjects with specific numerical disability. A previous study has suggested its absence in a sample of children with arithmetical problems (Bachot, Gevers, Fias, & Roeyers, 2005). However, those children also had visuospatial impairments in addition to arithmetical ones, which makes the result unable to speak unambiguously about the necessity of numerical processing per se in the existence of the effect, instead just reflecting possibly an abnormal representation of number magnitude on the left-to-right-orientated mental number line, a concept that by definition arises by interaction of the number and space domains. The current results are ambiguous about the necessity of intact numerical processing in SNARC, given the existence of patients who were both within and outside of the SNARC control limits.

The limited MRT data indicates that patients who were within normative limits on this task were also not impaired in the SNARC and in the number line. This can be interpreted as a tentative link between (3-dimensional) mental rotation abilities on the one hand, and more general abilities of (2-dimensional) written symbol manipulation and (1-dimensional) internal number representation on the other hand. However, as previously discussed in the context of mathematicians' superior MRT performance (Study 1, page 48), it is difficult to argue that a close relationship exists between the spatial abilities involved in these different tasks.

These results only refer to a number-pitch dissociation in terms of the way that each of these two dimensions interacts with a third dimension – space; specifically, they refer to the way in which each of them primes spatial responses. While correlatory data between SNARC size and maths performance is limited to the null result described previously (Cipora & Nuerk, 2013), and none yet exists for the SMARC and music performance, it seems reasonable to speculate that the dissociations found in the current study speak not against a complete dissociation of the two domains but of their maintaining of the ability to prime spatial responses regardless of whether or not there is impairment in the other domain. It is likely that SNARC is not integral to number just like SMARC is not integral to pitch processing, as has been shown by the normal pitch

detection thresholds of all-but-one patients in the current data set. The fact that, previously, the SMARC effect with horizontal mapping was found only in musicians but not in non-musicians (Lidji et al., 2007) may indicate that it is related to acquired knowledge in the same way that the orientation of the SNARC is related to writing direction (Dehaene et al., 1993).

5.5.1 Limitations & conclusions

The number of 1-sample t-tests used to verify the presence of the effects was high: 2 groups (F and P) × 2 tasks (SMARC and SNARC) × 2 measures (dAcc or dRT) × 2 stimulus sizes (low and high) = 16 statistical tests. A Bonferroni correction for multiple comparisons would have been appropriate here, however was not carried out due to the fact that this approach was seen as a replacement for the mixed-model ANOVA, where no correction would have been necessary. This replacement was necessary because the variances across the two groups were high, something that should not be seen as a statistical artefact but as normal given that one of the groups was a patient group. The 1-sample t-test approach was therefore selected given the lack of a widely-accepted non-parametric ANOVA for mixed models (i.e. having both between- and within-subjects factors). It should nonetheless be noted that, with the Bonferroni correction applied, all of the reported effects lose significance.

Finally, this study would have benefited from a more homogenous sample of patients, not only from the perspective of spectrum of behavioural impairments but also from that of lesion site and size. However, such an enterprise would have been difficult even for a much larger scale neuropsychology study than the current one. For the same logistical reasons, it was also not possible to match the patient and control groups in terms of age, years of formal education and years of musical training.

Further studies should also make use of a line bisection task to further increase the inferential accuracy of any replicated dissociations (such as the ones found here) between number-space and number-pitch. Also, using not only patients with numerical impairments but also patients with musical impairments (e.g. with amusia) would significantly increase the potential of characterising these interactions.

5.6. Acknowledgements

I thank Manuela Marin for helpful discussions during the early stages. I thank Rowena Beecham for providing the auditory stimuli used in the SMARC task, and Mihaela Duta for help with Psychophysics Toolbox

script-writing. I acknowledge the financial contribution of the European Society for Cognitive Psychology (ESCoP), which have supported this study by awarding me the 2012 Early Career Stimulus grant.

6 General Discussion

As it turns out, untangling The Structure Of The Mathematical Brain is no easy task. In this thesis, I addressed a number of individual questions that relate to how the brain encodes number in relation to other dimensions, how performance in arithmetic learning can be enhanced with tES, and how increased proficiency in mathematical faculties (with or without tES) shapes the brain. These questions are complementary in that, between them, they address some of the most difficult and still open questions in numerical cognition.

Structural changes in the brain are complex and their core phenomena occur at microscopic scales, where non-invasive observation in humans is difficult. Here I had at my disposition a wide range of MRI methods, which together provided information about both the exposed surface of the brain, i.e. the cerebral cortex, where most processing occurs, as well as the tremendously complex connections that exist below that surface, which allow outputs from all these individual processes to be integrated and thus lead to observable behaviour. While my MRI recordings are only as accurate as the current state of the art in that field, this observable behaviour is, at least at a first level, more straightforward to measure through the classical means of experimental psychology. I quantified several mathematically-relevant – as well as domain-general – behaviours, in an attempt to draw correlational (if not causal) links between them and the brain – an endeavour that lies at the heart of cognitive neuroscience.

I also interfered with the normal functioning of the brain using tRNS, a more recent form of tES. As a general technique, tES, just like its older cousin TMS (transcranial magnetic stimulation), has the potential to go beyond the correlational inference provided by neuroimaging, and actually allow drawing causal links between brain and behaviour, by selectively interfering with the function of relevant brain regions. On the other hand, it has the potential – by now well confirmed in the literature – to enhance this behaviour across a

wide range of cognitive domains, despite the fact that its exact cellular and molecular mechanisms of operation are not yet fully understood.

6.1. Summary of findings

Study 1 (Chapter 2) aimed to characterise the mathematical brain by cross-sectionally comparing a group of mathematicians with a group of non-mathematician academics in terms of MRI-derived structural measures and their correlation with relevant behavioural measures. Mathematical expertise was associated with better performance in a diverse range of behavioural tasks, both related and unrelated to maths. The brains of mathematicians, relative to controls, were found to have increased GMD in the right superior parietal lobule but decreased GMD in the right intraparietal sulcus and left inferior frontal gyrus, and also decreased surface area in the left fusiform gyrus. For controls, behavioural performance correlated negatively with GMD and lGI in the left superior frontal gyrus (SFG). In terms of white matter, there were no significant differences in fractional anisotropy, however tractography revealed a stronger right SFG to left SFG tract in mathematicians than in controls. These results suggested that mathematical expertise is associated with a distributed pattern of changes throughout the brain, which included both increases and decreases in the various structural measures used. The two groups were examined at a single point in time, such that the relative roles of innate predisposition and experience-dependent plasticity in causing the observed effects cannot be untangled. It is, however, plausible that these changes represent a "snapshot" from a brain that is in a perpetual state of change due, possibly, to a feedback mechanism between the nature- and nurture-type factors. Study 1 was later complemented by Study 3, in which the mathematical training was controlled in the experiment rather than being part of the subjects' personal history, and flanked this training between two scans, thus enabling a pre-to-post comparison of the changes involved.

Study 2 (Chapter 3) investigated the effect of tRNS on two different modes of learning arithmetic: by drill (fact retrieval) and by calculation (applying an algorithm). As the transition – occurring with education and development – from calculation-based to drill-based learning is known to be accompanied by a frontal-to-parietal shift in recruited brain areas, and as tES is most beneficial to neural populations actually active at the time of stimulation, here I attempted to maximise the effect of tRNS by stimulating these two parts of the brain in a similar order. Therefore, tRNS was applied to the bilateral prefrontal cortex for the first 3 days of a 5-day training, and to the posterior parietal cortex for the last 2 days. Results indicated that tRNS decreased

overall RTs and increased the rate of learning for Calculation problems during the training phase; it also decreased Calculation RTs for both trained and untrained (new) problems during a subsequent testing phase. When analysing RTs by the properties of their distributions as opposed to their overall means, the effect of tRNS was significant in reducing the exponential tail of the stimulated groups' distribution, suggesting decreased reliance on effortful strategies, strategies that tRNS might modulate. Task difficulty (load) was manipulated between groups, and all these effects only reached significance in the more difficult (higher load) condition. By replicating the results of a previous tRNS study where improvement in a similar learning task was shown, but by using a different pattern of stimulated regions, these results confirmed the robustness of the tES (and tRNS method in particular) in displaying these effects, which here too were specific to the trained material and did not extend into general-domain tasks also compared pre-to-post training. They also confirmed previous reports of the facilitatory effects of tES being modulated by task difficulty.

Study 3 (Chapter 4) was a large-scale investigation that aimed to identify brain regions whose structural parameters change as a result of tRNS-assisted arithmetic learning of the same type as the one used in Study 2. To that end, structural and diffusion MRI scans were acquired before and after training, and interactions between the pre-to-post difference in computed structural measures and stimulation group were investigated. This time, tRNS was applied as either frontal-only, parietal-only, or sham stimulation for all 5 days of training, in order to separately assess the impact of each montage on structural changes. Behaviourally, tRNS decreased RTs on the Calculation task for both frontal and parietal stimulation as compared to sham. Like in Study 2, these enhancements were accompanied by a decrease in the RT distribution tail but not also by an increase in learning rate. On the other hand, group differences in the average within-session (online) and between-session (offline) learning magnitudes were found. The two types of learning were negatively correlated both within and across stimulation groups, suggesting that online learning takes places at the expense of offline learning and vice-versa. Anatomically, several group effects were found in terms of the pre-to-post change in cortical volume and thickness of frontal, parietal and temporal clusters. These effects consisted of both increases and decreases in the frontal and parietal groups as compared to sham. These clusters were not proximal to the stimulation site, which suggests that it is not straightforward to predict either the regions of the brain receiving the maximal tRNS effect, or the direction of change in structural parameters of these regions as a result of tRNS. The results of this study on one hand partly replicate those of Study 2 and others that indicate the ability of tRNS to enhance behavioural

performance in arithmetic. They also take our understanding of the mechanisms of action of tRNS further, by showing that such behavioural enhancements are accompanied by altered structural parameters in regions of the brain that, while relevant to numerical cognition, are not necessarily proximal to the stimulation site. This study represented, to the best of my knowledge, the first longitudinal investigation into the effects of tRNS-assisted cognitive training on not only behavioural performance but also the associated structural brain changes.

Study 4 (Chapter 5) was theoretically founded on converging evidence of an overlap between the neural representation of number and that of other dimensions. Their encoding is believed to use certain parietal populations also used to encode space. One of these dimensions is musical pitch. At the behavioural level, hints of a representational overlap between dimensions can exist in tasks where a response in one dimension is required for stimuli along the other dimension. The interaction between the two is referred to as a stimulus-response compatibility (SRC) effect. The SNARC is one such effect for the number-space interaction, and implies that responses with left hand to small numbers are faster and more accurate with the left hand, and to large numbers with the right hand. A similar effect, SMARC, exists in the pitch domain, and here I administered both paradigms (SNARC, SMARC) to numerically-impaired patients and controls in order to examine the degree to which impairment in one of these domains affects an SRC effect where stimuli along the other domain are responded to spatially. The results suggested a double dissociation between the two groups, with SNARC but not SMARC present in the controls, and SMARC but not SNARC present in the patients. Individual patients treated as single cases confirmed this pattern, and also suggested that, within the number domain, not only SNARC but also performance on the number line task is outside of the normative control limits. This latter result speaks to the central role of number-to-space mapping ability within the broader realm of numerical cognition, and also to the more controversial topic concerning the importance (if not the necessity) of this concept in the existence of SNARC. These results can be taken as preliminary evidence that SRC effects involving spatial responses to number and pitch respectively can dissociate in the numerically-impaired brain. This might suggest a deeper-level dissociation between the two domains also in the non-impaired brain, despite evidence of existing neural and behavioural overlaps.

6.2. Overall conclusions

6.2.1 Maths performance and brain structure

Knowing what parameters of what brain structures systematically predict maths performance, and what the developmental trajectories of these structures are, would be very useful indeed. We could use this information in a number of ways, for instance to establish neural markers that could foresee, in advance of observations made by educators, which children are either at risk for atypical numerical development (developmental dyscalculia), or, on the contrary, likely to overachieve in maths. Either prediction would be priceless in optimising subsequent educational strategies – although it might also raise neuroethical concerns such as the ones raised about the field of neuroenhancement using tES methods (R. Cohen Kadosh, Levy, O'Shea, Shea, & Savulescu, 2012).

Unfortunately, some of the results in this thesis confirm the current opinion in the field, which is that we are "not there yet". At least when assessed with MRI, the structural correlates of existing and acquired maths expertise are spread throughout the brain, in networks including but certainly not limited to regions already known to be "maths-relevant". Examples include the right postcentral gyrus differences in local gyrification index (lGI) between mathematicians and non-mathematicians (Study 1, Chapter 2) and the differences in pre-to-post-training in the cortical volume of the right anterior cingulate, in the parietal group versus sham (Study 3, Chapter 4). Additionally, the direction of these changes are not always in the expected direction, such as, for instance, the decreased GMD in right intraparietal sulcus seen in mathematicians versus non-mathematicians (Study 1), an effect previously observed when comparing people with dyscalculia against controls (Cappelletti & Price, 2014; Molko et al., 2003; Rotzer et al., 2008; Rykhlevskaia et al., 2009). Having mentioned these caveats in the interpretation of the results, however, a noteworthy contribution of the MRI-based Studies 1 and 3, is that they have implicated, in terms of structural correlates of maths expertise and learning (respectively), changes in a measure – namely, lGI– that was previously mainly involved in clinical rather than cognitive expertise-related investigations.

6.2.2 tES to enhance maths performance

In the early stages of this doctoral project, one study had already been published demonstrating the promising behavioural effects –improved automaticity in the comparing of artificial digits and improved accuracy of their magnitude-to-space mapping – of tDCS-assisted task performance (R. Cohen Kadosh et al.,

2010). At that time, a newer tES method – tRNS – had recently been described and its enhancing effects on motor learning demonstrated (Terney et al., 2008). No studies had yet been published, however, that showed cognitive improvements. I thus chose to use this method, given tRNS' facilitatory effect at *both* electrodes (in turn due to the absence of a circuit polarity) and its apparent superiority over tDCS in terms of effect size of behavioural improvement achieved (Fertonani et al., 2011). This particular form of tES was used in an initial study where we showed tRNS' ability to enhance arithmetic (Snowball et al., 2013). Using NIRS, that study also gave a first indication of what the neural mechanisms of tRNS might be, suggesting it increases efficiency of the neurovascular coupling in the cortex underlying the left stimulating electrode (DLPFC). By contrast, Study 3 found structural changes in cortices that were rather distal to both stimulating electrodes.

Following Snowball et al. (2013), however, questions still remained about the cross-study replicability of tRNS' ability to enhance cognition in the domain of arithmetic, about the optimal montages and about the structural changes in the brain associated with the behavioural effects. Studies 2 and 3 confirmed the first of these points and also gave us some potential candidates for behavioural phenomena that might provide a slightly deeper explanation of the immediately-observable effects; these phenomena include changes in the profile of the RT distribution and separate effects relating to within-session (online) and between-session (offline) learning. Study 3 went further and also showed what structural changes occur in the brain as a result of tRNS-assisted cognitive training in arithmetic, a first-of-its kind exploration.

Some expectations existed regarding the location of these changes in the brain. First of all, changes were expected in areas involved with either number in isolation or in the context of mental arithmetic. This expectation was only partially fulfilled: some of the areas where effects were found included temporal and cingulate areas that are not immediately number-relevant. Second, based on the NIRS results of Snowball et al. (2013), i.e. that tRNS decreased peak amplitudes of oxygenated haemoglobin concentrations in the cortex immediately underlying the left stimulating electrode (the DLPFC), here too I expected at least some changes to occur in the cortex underlying stimulation site in the two groups. This expectation was, again, not met, although a deeper search of the tES current density modelling literature makes this results less surprising. It has to be noted, also, that what makes the spatial coincidence between the stimulation site and the haemodynamic effect found in Snowball et al. (2013) less remarkable is that the NIRS technique only provided prefrontal recordings as opposed to recordings from all parts of the brain. Overall, the reliability and

more fine-grained interpretation of the changes observed in Study 3 remain, of course, to hopefully be elucidated by future investigations.

6.2.3 Number-pitch-space interactions

Although number may be the most salient dimension processed in the parietal lobes, there is converging evidence for a common metric in this part of the brain that also includes dimensions such as space and pitch. The pitch-space interaction (SMARC) is a relatively young observation, and the data from the control group in Study 4 does not entirely replicate its existence, a difficulty which other labs have recently encountered when probing the effect in healthy populations. Although showing the descriptive trend of the SMARC, the effect did not reach statistical significance in the controls. The observation that patients who were within this control window were said to show the effect as single cases is, of course, qualified by the statistical aspect mentioned above. That observation was, however, more convincing when treating patients as a group. The fact that patients did not show the (comparatively reliable) SNARC fills in a part of the numerical cognition literature that is not very well described – that of number-to-space mappings in numerical impairment – and suggests that intact magnitude processing might be a prerequisite for this dimension to be able to prime spatially-polarised responses. Pieces of the puzzle that are still conspicuously missing relate to whether musical impairment leads to impairment in its own SRC effect, the SMARC, and whether in that case as well, a dissociation would be shown across the number domain, in SNARC. If that were to be the case, it would constitute believable evidence of qualitatively different mechanisms by virtue of which number and some of its allegedly parietally-related domains, such as pitch, relate to space.

6.3. Limitations

The limitations of each empirical study presented in this thesis have been presented under the relevant section of each study's chapter (Chapters 2 through 5). In summary, these limitations have to do with logistical constraints on the data acquisition and analysis, shortcomings in the diffusion MRI methodology used, and general tES-specific limitations .

In Study 1, the small range of years of experience in the mathematicians sample meant that a further result that would have lent increased credibility to the observed changes was missing; that is, a cross-subjects correlation between the amount of structural change and the amount of time spent as a mathematician, as was done for instance in Aydin et al. (2007).

Studies 2 and 3 suffer, as any other tES investigation, from the method's low spatial focality, which means that low confidence is placed not only on the exact sulcus or gyrus that is closest to the stimulating electrodes but also on the exact region of the cortex that receives from the current a peak increase in excitability. Even if this confidence were increased, it would still be difficult to predict the exact behavioural effects that receipt of this peak current distribution would entail. Despite these limitations, and despite the limited understanding of its mechanisms, tES can and has been used – as a trial-and-error or "black box" – mode of enhancing cognition. Part of the reason why such an approach is appropriate is the technique's low risk of dangerous side effects such as seizures: tDCS has, for instance, even been used in the treatment of seizures (Yook, Park, Seo, Kim, & Ko, 2011).

Finally, in order for Study 4 to completely characterise the nature of the pairwise number-pitch-space interactions, similar dissociations as the ones observed would have had to also be observed not only in the numerically impaired population but also in a musically impaired one (i.e., people with amusia). Such a desirable extension to the study was, however, not possible during the time frame of the current doctoral project.

6.4. Future directions

In order to obtain a data set that can unambiguously show the structural changes taking place in the brain as a result of acquired experience with mathematics, possibly the best-suited design is that of the cross-over study. In such a study, a group of subjects would be assigned to maths training, a control group to matched (non-maths) training, and the roles would subsequently be reversed between groups for an additional training period of the same length. The data can then be analysed longitudinally, as has been done for instance in Study 3, in order to draw a causal conclusion between each type of training and each observed set of changes. Such an experiment makes the most sense with adults, where it can be assumed that any observed structural changes are mainly due to the training intervention rather than being confounded by changes due to development. The simple longitudinal design with children, however, as is the recent study by Qin et al. (2014), also has great potential to observe what changes are accompanied by maths education. The fMRI approach used in that study, multivoxel pattern analysis (MVPA), holds great potential for unravelling brain regions that not only are involved in a single task (as univariate fMRI does) but also regions can differentiate between subtle patterns and conditions within the same task. Using algorithms borrowed from

computational fields such as machine learning, MVPA has, for instance, been used to identify brain regions in which the local pattern of activity accurately discriminated between attending to symbolic and nonsymbolic number formats (Eger et al., 2009), rising or falling melodies (Y.-S. Lee, Janata, Frost, Hanke, & Granger, 2011) or instruments of various timbres (M. Casey, Thompson, Kang, Raizada, & Wheatley, 2012). One of MVPA's more recent applications in numerical cognition has been a study where the fusiform gyrus was identified as the region most able to differentiate between a group of autistic children and a control group in terms of varying strategies used for solving complex arithmetic (Iuculano et al., 2014). Such a paradigm could conceivably also be applied to identify regions responding differently to, for instance, drill vs calculation problem solving.

In order to further explore the potential of tES to enhance numerical cognition, future studies could, in addition to continuing investigations using the still-young tRNS technique, also make use of tES modes that have been less used in numerical cognition. Such modes include tACS (transcranial alternating current stimulation) with its ability to alter the bran's natural oscillations, or HD-tES (high definition tES) with its improved spatial resolution. A wide range of tasks should be used *within* the same study that tap into as many numerical faculties as possible, noting any simultaneous cross-task improvements or, on the contrary, any "hidden costs" even within the numerical cognition realm (Iuculano & Cohen Kadosh, 2013). Such a battery of tasks would ideally include paradigms for e.g. comparison of symbolic and non-symbolic number, magnitude-to-space mapping using normal or artificial digits, exact or approximate calculation, etc. Finally, since tES applications are equally promising for enhancing normal performance as they are for neurorehabilitation (e.g. dyscalculia), studies should also continue to seek potential clinical applications for new tES modes and new behavioural paradigms.

In order to further bin and narrow down the specificity of number processing in the brain (and in particular in the parietal lobes), studies could rely on approaches such as TMS, that allow the induction of virtual lesions (e.g., Walsh & Cowey, 2000) in focal parts of the parietal lobe, and subsequently observe dissociative effects in terms of e.g. number, pitch and space processing.

7 References

Abell, F., Krams, M., Ashburner, J., Passingham, R., Friston, K., Frackowiak, R., … Frith, U. (1999). The neuroanatomy of autism: a voxel-based whole brain analysis of structural scans. *Neuroreport*, *10*(8), 1647–1651.

Agrillo, C., Piffer, L., Bisazza, A., & Butterworth, B. (2012). Evidence for two numerical systems that are similar in humans and guppies. *PloS One*, *7*(2), e31923. doi:10.1371/journal.pone.0031923

Ambrus, G. G., Paulus, W., & Antal, A. (2010). Cutaneous perception thresholds of electrical stimulation methods: Comparison of tDCS and tRNS. *Clinical Neurophysiology*, *121*(11), 1908–1914. doi:10.1016/j.clinph.2010.04.020

Ambrus, G. G., Zimmer, M., Kincses, Z. T., Harza, I., Kovács, G., Paulus, W., & Antal, A. (2011). The enhancement of cortical excitability over the DLPFC before and during training impairs categorization in the prototype distortion task. *Neuropsychologia*, *49*(7), 1974–1980. doi:10.1016/j.neuropsychologia.2011.03.026

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1120–1136. doi:10.1037/0278-7393.25.5.1120

Andres, M., Pelgrims, B., Michaux, N., Olivier, E., & Pesenti, M. (2011). Role of distinct parietal areas in arithmetic: An fMRI-guided TMS study. *NeuroImage*, *54*(4), 3048–3056. doi:10.1016/j.neuroimage.2010.11.009

Andrews, S. (2011). Improving working memory: the effect of combining cognitive activity and anodal transcranial direct current stimulation to the left dorsolateral prefrontal cortex. *Brain Stimulation*, 84–89.

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

Ansari, D., Dhital, B., & Siong, S. C. (2006). Parametric effects of numerical distance on the intraparietal sulcus during passive viewing of rapid numerosity changes. *Brain Research*, *1067*(1), 181–188. doi:10.1016/j.brainres.2005.10.083

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

Ansari, D., Grabner, R. H., Koschutnig, K., Reishofer, G., & Ebner, F. (2011). Individual differences in mathematical competence modulate brain responses to arithmetic errors: An fMRI study. *Learning and Individual Differences*, *21*(6), 636–643. doi:10.1016/j.lindif.2011.07.013

Antal, A., Begemeier, S., Nitsche, M. A., & Paulus, W. (2008). Prior state of cortical activity influences subsequent practicing of a visuomotor coordination task. *Neuropsychologia*, *46*(13), 3157–3161. doi:10.1016/j.neuropsychologia.2008.07.007

Antal, A., Nitsche, M. A., Kincses, T. Z., Kruse, W., Hoffmann, K.-P., & Paulus, W. (2004). Facilitation of visuo-motor learning by transcranial direct current stimulation of the motor and extrastriate visual areas in humans. *European Journal of Neuroscience*, *19*(10), 2888–2892. doi:10.1111/j.1460-9568.2004.03367.x

Antal, A., Nitsche, M. A., & Paulus, W. (2001). External modulation of visual perception in humans. *Neuroreport*, *12*(16), 3553–3555.

Arsalidou, M., & Taylor, M. J. (2011). Is 2 + 2 = 4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, *54*(3), 2382–2393. doi:10.1016/j.neuroimage.2010.10.009

Aydin, K., Ucar, A., Oguz, K. K., Okur, O. O., Agayev, A., Unal, Z., … 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.

Bachot, J., Gevers, W., Fias, W., & Roeyers, H. (2005). Number sense in children with visuospatial disabilities: Orientation of the mental number line. *Psychological Science*, *47*(1), 172–183.

Bailey, J. A., Zatorre, R. J., & Penhune, V. B. (2013). Early Musical Training Is Linked to Gray Matter Structure in the Ventral Premotor Cortex and Auditory–Motor Rhythm Synchronization Performance. *Journal of Cognitive Neuroscience*, 1–13. doi:10.1162/jocn_a_00527

Baldo, J. V., & Dronkers, N. F. (2007). Neural correlates of arithmetic and language comprehension: A common substrate? *Neuropsychologia*, *45*(2), 229–235. doi:10.1016/j.neuropsychologia.2006.07.014

Balota, D. A., & Yap, M. J. (2011). Moving Beyond the Mean in Studies of Mental Chronometry The Power of Response Time Distributional Analyses. *Current Directions in Psychological Science*, *20*(3), 160–166. doi:10.1177/0963721411408885

Barnea-Goraly, N., Eliez, S., Menon, V., Bammer, R., & Reiss, A. L. (2005). Arithmetic ability and parietal alterations: A diffusion tensor imaging study in Velocardiofacial syndrome. *Cognitive Brain Research*, *25*(3), 735–740. doi:10.1016/j.cogbrainres.2005.09.013

Baron-Cohen, S., Wheelwright, S., Burtenshaw, A., & Hobson, E. (2007). Mathematical Talent is Linked to Autism. *Human Nature*, *18*(2), 125–131. doi:10.1007/s12110-007-9014-0

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Barrouillet, P., & Fayol, M. (1998). From algorithmic computing to direct retrieval: Evidence from number and alphabetic arithmetic in children and adults. *Memory & Cognition*, *26*(2), 355–368. doi:10.3758/BF03201146

Basser, P. J., Mattiello, J., & Lebihan, D. (1994). Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. *Journal of Magnetic Resonance, Series B*, *103*(3), 247–254. doi:10.1006/jmrb.1994.1037

Beddington, J., Cooper, C. L., Field, J., Goswami, U., Huppert, F. A., Jenkins, R., … Thomas, S. M. (2008). The mental wealth of nations. *Nature*, *455*(7216), 1057–1060. doi:10.1038/4551057a

Beecham, R., Reeve, R. A., & Wilson, S. J. (2009). Spatial Representations Are Specific to Different Domains of Knowledge. *PLoS ONE*, *4*(5), e5543. doi:10.1371/journal.pone.0005543

Behrens, T. E. J., Johansen-Berg, H., Woolrich, M. W., Smith, S. M., Wheeler-Kingshott, C. a. M., Boulby, P. A., … Matthews, P. M. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience*, *6*(7), 750–757. doi:10.1038/nn1075

Behrens, T. E. J., Woolrich, M. w., Jenkinson, M., Johansen-Berg, H., Nunes, R. g., Clare, S., … Smith, S. m. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*, *50*(5), 1077–1088. doi:10.1002/mrm.10609

Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., & Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience*, *8*(9), 1148–1150. doi:10.1038/nn1516

Benton, A., Van Allen, M., Hamser, K., & Levin, H. (1978). *Test of Facial Recognition Manual*.

Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical Correlates of Musicianship as Revealed by Cortical Thickness and Voxel-Based Morphometry. *Cerebral Cortex*, *19*(7), 1583–1596. doi:10.1093/cercor/bhn196

Boisgueheneuc, F. du, Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., … Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: a lesion study. *Brain*, *129*(12), 3315 –3328. doi:10.1093/brain/awl244

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*(1), 189–201.

Booth, J. L., & Siegler, R. S. (2008). Numerical Magnitude Representations Influence Arithmetic Learning. *Child Development*, *79*(4), 1016–1031. doi:10.1111/j.1467-8624.2008.01173.x

Brem, A.-K., Fried, P. J., Horvath, J. C., Robertson, E. M., & Pascual-Leone, A. (2014). Is neuroenhancement by noninvasive brain stimulation a net zero-sum proposition? *NeuroImage*, *85, Part 3*, 1058–1068. doi:10.1016/j.neuroimage.2013.07.038

Brown, T. T., Lugar, H. M., Coalson, R. S., Miezin, F. M., Petersen, S. E., & Schlaggar, B. L. (2005). Developmental Changes in Human Cerebral Functional Organization for Word Generation. *Cerebral Cortex*, *15*(3), 275–290. doi:10.1093/cercor/bhh129

Bueti, D., Lasaponara, S., Cercignani, M., & Macaluso, E. (2012). Learning about Time: Plastic Changes and Interindividual Brain Differences. *Neuron*, *75*(4), 725–737. doi:10.1016/j.neuron.2012.07.019

Butterworth, B. (2005a). Developmental dyscalculia. *Handbook of Mathematical Cognition*, 455–467.

Butterworth, B. (2005b). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, *46*(1), 3–18. doi:10.1111/j.1469-7610.2004.00374.x

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, *14*(12), 534–541. doi:10.1016/j.tics.2010.09.007

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From Brain to Education. *Science*, *332*(6033), 1049 –1053. doi:10.1126/science.1201536

Bynner, J., & Parsons, S. (1997). Does numeracy matter. *Basic Skills Agency, London*.

Campbell, J., & Penner-Wilger, M. (2006). Calculation latency: The μ of memory and the τ of transformation. *Memory & Cognition*, *34*(1), 217–226. doi:10.3758/BF03193400

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, *4*(5), 844–854.

Cantlon, J. F., Platt, M., & Brannon, E. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*(2), 83–91. doi:10.1016/j.tics.2008.11.007

Cappelletti, M., Gessaroli, E., Hithersay, R., Mitolo, M., Didino, D., Kanai, R., … Walsh, V. (2013). Transfer of Cognitive Training across Magnitude Dimensions Achieved with Concurrent Brain Stimulation of the Parietal Lobe. *The Journal of Neuroscience*, *33*(37), 14899–14907. doi:10.1523/JNEUROSCI.1692-13.2013

Cappelletti, M., & Price, C. J. (2014). Residual number processing in dyscalculia. *NeuroImage: Clinical*, *4*, 18–28. doi:10.1016/j.nicl.2013.10.004

Casey, B. J., Thomas, K. M., Davidson, M. C., Kunz, K., & Franzen, P. L. (2002). Dissociating Striatal and Hippocampal Function Developmentally with a Stimulus–Response Compatibility Task. *The Journal of Neuroscience*, *22*(19), 8647–8652.

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: what have we learned about cognitive development? *Trends in Cognitive Sciences*, *9*(3), 104–110. doi:10.1016/j.tics.2005.01.011

Casey, M., Thompson, J., Kang, O., Raizada, R., & Wheatley, T. (2012). Population Codes Representing Musical Timbre for High-Level fMRI Categorization of Music Genres. *Machine Learning and Interpretation in Neuroimaging*, 34–41.

Castelli, F., Glaser, D. E., & Butterworth, B. (2006). Discrete and analogue quantity processing in the parietal lobe: A functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(12), 4693–4698. doi:10.1073/pnas.0600444103

Chaieb, L., Kovacs, G., Cziraki, C., Greenlee, M., Paulus, W., & Antal, A. (2009). Short-duration transcranial random noise stimulation induces blood oxygenation level dependent response attenuation in the human motor cortex. *Experimental Brain Research*, *198*(4), 439–444. doi:10.1007/s00221-009-1938-7

Chaieb, L., Paulus, W., & Antal, A. (2011). Evaluating Aftereffects of Short-Duration Transcranial Random Noise Stimulation on Cortical Excitability. *Neural Plasticity*, *2011*. doi:10.1155/2011/105927

Cheek, J. M., & Smith, L. R. (1999). Music training and mathematics achievement. *Adolescence*, *34*(136), 759–761.

Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15*(1), 2–11.

Chiang, M.-C., Barysheva, M., Shattuck, D. W., Lee, A. D., Madsen, S. K., Avedissian, C., … Thompson, P. M. (2009). Genetics of Brain Fiber Architecture and Intellectual Performance. *The Journal of Neuroscience*, *29*(7), 2212 –2224. doi:10.1523/JNEUROSCI.4184-08.2009

Chochon, F., Cohen, L., Moortele, P. F. van de, & Dehaene, S. (1999). Differential Contributions of the Left and Right Inferior Parietal Lobules to Number Processing. *Journal of Cognitive Neuroscience*, *11*(6), 617–630. doi:10.1162/089892999563689

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.

Cho, Y. S., Bae, G. Y., & Proctor, R. W. (2011). Referential coding contributes to the horizontal SMARC effect. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 726–734. doi:10.1037/a0026157

Cho, Y. S., & Proctor, R. W. (2003). Stimulus and response representations underlying orthogonal stimulus-response compatibility effects. *Psychonomic Bulletin & Review*, *10*(1), 45–73. doi:10.3758/BF03196468

Chung, M. K., Robbins, S. M., Dalton, K. M., Davidson, R. J., Alexander, A. L., & Evans, A. C. (2005). Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, *25*(4), 1256–1265. doi:10.1016/j.neuroimage.2004.12.052

Cipora, K., & Nuerk, H.-C. (2013). Is the SNARC effect related to the level of mathematics? No systematic relationship observed despite more power, more repetitions, and more direct assessment of arithmetic skill. *The Quarterly Journal of Experimental Psychology*, *66*(10), 1974–1991. doi:10.1080/17470218.2013.772215

Clemens, B., Jung, S., Zvyagintsev, M., Domahs, F., & Willmes, K. (2013). Modulating arithmetic fact retrieval: A single-blind, sham-controlled tDCS study with repeated fMRI measurements. *Neuropsychologia*, *51*(7), 1279–1286. doi:10.1016/j.neuropsychologia.2013.03.023

Cohen Kadosh, K., Henson, R. N. A., Cohen Kadosh, R., Johnson, M. H., & Dick, F. (2010). Task-dependent Activation of Face-sensitive Cortex: An fMRI Adaptation Study. *Journal of Cognitive Neuroscience*, *22*(5), 903–917. doi:10.1162/jocn.2009.21224

Cohen Kadosh, R., Bahrami, B., Walsh, V., Butterworth, B., Popescu, T., & Price, C. J. (2011). Specialization in the human brain: the case of numbers. *Frontiers in Human Neuroscience*, *5*(62), 1–9.

Cohen Kadosh, R., Brodsky, W., Levin, M., & Henik, A. (2008). Mental representation: What can pitch tell us about the distance effect? *Cortex*, *44*(4), 470–477. doi:16/j.cortex.2007.08.002

Cohen Kadosh, R., Cohen Kadosh, K., Kaas, A., Henik, A., & Goebel, R. (2007). Notation-Dependent and -Independent Representations of Numbers in the Parietal Lobes. *Neuron*, *53*(2), 307–314. doi:10.1016/j.neuron.2006.12.025

Cohen Kadosh, R., Henik, A., Rubinsten, O., Mohr, H., Dori, H., van de Ven, V., … Linden, D. E. J. (2005). Are numbers special?: The comparison systems of the human brain investigated by fMRI. *Neuropsychologia*, *43*(9), 1238–1248. doi:10.1016/j.neuropsychologia.2004.12.017

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.

Cohen Kadosh, R., Levy, N., O'Shea, J., Shea, N., & Savulescu, J. (2012). The neuroethics of non-invasive brain stimulation. *Current Biology*, R108–R111.

Cohen Kadosh, R., Soskic, S., Iuculano, T., Kanai, R., & Walsh, V. (2010). Modulating Neuronal Activity Produces Specific and Long-Lasting Changes in Numerical Competence. *Current Biology*, *20*(22), 2016–2020. doi:16/j.cub.2010.10.007

Cohen Kadosh, R., & Walsh, V. (2007). Dyscalculia. *Current Biology*, *17*(22), R946–R947. doi:10.1016/j.cub.2007.08.038

Cohen Kadosh, R., & Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences*, *32*(3-4), 313–328.

Cohen, L., Dehaene, S., Chochon, F., Lehéricy, S., & Naccache, L. (2000). Language and calculation within the parietal lobe: a combined cognitive, anatomical and fMRI study. *Neuropsychologia*, *38*(10), 1426–1440. doi:10.1016/S0028-3932(00)00038-5

Cohen, M. X., Schoene-Bake, J.-C., Elger, C. E., & Weber, B. (2009). Connectivity-based segregation of the human striatum predicts personality characteristics. *Nat Neurosci*, *12*(1), 32–34. doi:10.1038/nn.2228

Crawford, J. R., & Howell, D. C. (1998). Comparing an Individual's Test Score Against Norms Derived from Small Samples. *The Clinical Neuropsychologist*, *12*(4), 482–486. doi:10.1076/clin.12.4.482.7241

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, *9*(2), 179–194. doi:10.1006/nimg.1998.0395

Datta, A., Bansal, V., Diaz, J., Patel, J., Reato, D., & Bikson, M. (2009). Gyri-precise head model of transcranial direct current stimulation: Improved spatial focality using a ring electrode versus conventional rectangular pad. *Brain Stimulation*, *2*(4), 201–207.e1. doi:10.1016/j.brs.2009.03.005

Dayan, E., Censor, N., Buch, E. R., Sandrini, M., & Cohen, L. G. (2013). Noninvasive brain stimulation: from physiology to network dynamics and back. *Nature Neuroscience*, *16*(7), 838–844. doi:10.1038/nn.3422

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. Oxford Univ Pr.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396. doi:10.1037/0096-3445.122.3.371

Dehaene, S., & Changeux, J.-P. (1993). Development of Elementary Numerical Abilities: A Neuronal Model. *Journal of Cognitive Neuroscience*, *5*(4), 390–407. doi:i: 10.1162/jocn.1993.5.4.390</p>

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures. *Science*, *320*(5880), 1217–1220. doi:10.1126/science.1156540

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

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of Mathematical Thinking: Behavioral and Brain-Imaging Evidence. *Science*, *284*(5416), 970 –974. doi:10.1126/science.284.5416.970

De Lange, F. P., Koers, A., Kalkman, J. S., Bleijenberg, G., Hagoort, P., van der Meer, J. W. ., & Toni, I. (2008). Increase in prefrontal cortical volume following cognitive behavioural therapy in patients with chronic fatigue syndrome. *Brain*, *131*(8), 2172.

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. doi:10.1016/j.cogbrainres.2003.09.005

Delazer, M., Girelli, L., Semenza, C., & Denes, G. (1999). Numerical skills and aphasia. *Journal of the International Neuropsychological Society*, *5*(03), 213–221. doi:null

Delazer, M., Ischebeck, A., Domahs, F., Zamarian, L., Koppelstaetter, F., Siedentopf, C. M., … Felber, S. (2005). Learning by strategies and learning by drill–evidence from an fMRI study. *Neuroimage*, *25*(3), 838–849.

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence*, *32*(1), 25–32.

Della-Maggiore, V., Scholz, J., Johansen-Berg, H., & Paus, T. (2009). The rate of visuomotor adaptation correlates with cerebellar white-matter microstructure. *Human Brain Mapping*, *30*(12), 4048–4053. doi:10.1002/hbm.20828

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. doi:10.1016/j.neuroimage.2006.01.021

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. doi:10.1016/j.jecp.2010.09.003

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*(2), 48–55. doi:10.1016/j.tine.2013.06.001

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology*, *103*(4), 469–479. doi:10.1016/j.jecp.2009.01.010

Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nat Neurosci*, *10*(7), 915–921. doi:10.1038/nn1925

Dowker, A. (2005). *Individual Differences in Arithmetical Abilities: Implications for Psychology, Neuroscience and Education*. Psychology Press.

Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Developmental Science*, *11*(5), 650–654. doi:10.1111/j.1467-7687.2008.00713.x

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature*, *427*(6972), 311. doi:10.1038/427311a

Draganski, B., Gaser, C., Kempermann, G., Kuhn, H. G., Winkler, J., Büchel, C., & May, A. (2006). Temporal and Spatial Dynamics of Brain Structure Changes during Extensive Learning. *The Journal of Neuroscience*, *26*(23), 6314–6317. doi:10.1523/JNEUROSCI.4628-05.2006

Driemeyer, J., Boyke, J., Gaser, C., Büchel, C., & May, A. (2008). Changes in Gray Matter Induced by Learning—Revisited. *PLoS ONE*, *3*(7), e2669. doi:10.1371/journal.pone.0002669

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering Cortical Number Coding from Human Brain Activity Patterns. *Current Biology*, *19*, 1608–1615.

Elkin, J., & Leuthold, H. (2011). The representation of pitch in auditory imagery: Evidence from S-R compatibility and distance effects. *Journal of Cognitive Psychology*, *23*(1), 76–91. doi:10.1080/20445911.2011.455251

Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex*. doi:10.1016/j.cortex.2013.03.007

Engvig, A., Fjell, A. M., Westlye, L. T., Moberget, T., Sundseth, Ø., Larsen, V. A., & Walhovd, K. B. (2010). Effects of memory training on cortical thickness in the elderly. *NeuroImage*, *52*(4), 1667–1676. doi:10.1016/j.neuroimage.2010.05.041

Estrada, C. A., Martin-Hryniewicz, M., Peek, B. T., Collins, C., & Byrd, J. C. (2004). Literacy and numeracy skills and anticoagulation control. *The American Journal of the Medical Sciences*, *328*(2), 88.

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *NeuroImage*, *26*(2), 471–479. doi:10.1016/j.neuroimage.2005.02.004

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the Efficiency and Independence of Attentional Networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347. doi:10.1162/089892902317361886

Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language*, *56*(2), 246–269. doi:10.1016/j.jml.2006.06.007

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, *37*(1), 1–9. doi:10.3758/MC.37.1.1

Fehr, T., Weber, J., Willmes, K., & Herrmann, M. (2010). Neural correlates in exceptional mental arithmetic—About the neural architecture of prodigious skills. *Neuropsychologia*, *48*(5), 1407–1416. doi:10.1016/j.neuropsychologia.2010.01.007

Feldman, D. E. (2009). Synaptic Mechanisms for Plasticity in Neocortex. *Annual Review of Neuroscience*, *32*, 33–55. doi:10.1146/annurev.neuro.051508.135516

Fertonani, A., Pirulli, C., & Miniussi, C. (2011). Random Noise Stimulation Improves Neuroplasticity in Perceptual Learning. *The Journal of Neuroscience*, *31*(43), 15416–15423.

Fias, W., Brysbaert, M., Geypens, F., & Ydewalle, G. (1996). The Importance of Magnitude Information in Numerical Processing: Evidence from the SNARC Effect. *Mathematical Cognition*, *2*(1), 95–110. doi:10.1080/135467996387552

Fias, W., & Fischer, M. H. (2005). Spatial representation of numbers. In *Handbook of mathematical cognition*.

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal Representation of Symbolic and Nonsymbolic Magnitude. *Journal of Cognitive Neuroscience*, *15*(1), 47–56. doi:10.1162/089892903321107819

Fischer, M. H. (2012). A hierarchical view of grounded, embodied, and situated numerical cognition. *Cognitive Processing*, *13*(1), 161–164. doi:10.1007/s10339-012-0477-5

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, *6*(6), 555–556. doi:10.1038/nn1066

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, *97*(20), 11050 –11055. doi:10.1073/pnas.200033797

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System. *NeuroImage*, *9*(2), 195–207. doi:10.1006/nimg.1998.0396

Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, *8*(4), 272–284. doi:10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4

Fletcher, P. D., & Witoonpanich, P. (2013). The brain basis of musicophilia: evidence from frontotemporal lobar degeneration. *Frontiers in Auditory Cognitive Neuroscience*, *4*, 347. doi:10.3389/fpsyg.2013.00347

Flöel, A., de Vries, M. H., Scholz, J., Breitenstein, C., & Johansen-Berg, H. (2009). White matter integrity in the vicinity of Broca's area predicts grammar learning success. *NeuroImage*, *47*(4), 1974–1981. doi:10.1016/j.neuroimage.2009.05.046

Foong, J., Maier, M., Clark, C. A., Barker, G. J., Miller, D. H., & Ron, M. A. (2000). Neuropathological abnormalities of the corpus callosum in schizophrenia: a diffusion tensor imaging study. *Journal of Neurology, Neurosurgery & Psychiatry*, *68*(2), 242–244. doi:10.1136/jnnp.68.2.242

Fornito, A., Yücel, M., Wood, S., Stuart, G. W., Buchanan, J.-A., Proffitt, T., … Pantelis, C. (2004). Individual Differences in Anterior Cingulate/Paracingulate Morphology Are Related to Executive Functions in Healthy Males. *Cerebral Cortex*, *14*(4), 424–431. doi:10.1093/cercor/bhh004

Foster, N. E. V., & Zatorre, R. J. (2010a). A Role for the Intraparietal Sulcus in Transforming Musical Pitch Information. *Cerebral Cortex*, *20*(6), 1350–1359. doi:10.1093/cercor/bhp199

Foster, N. E. V., & Zatorre, R. J. (2010b). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage*, *53*(1), 26–36. doi:10.1016/j.neuroimage.2010.06.042

Fregni, F., Boggio, P. S., Nitsche, M. A., Bermpohl, F., Antal, A., Feredoes, E., … Pascual-Leone, A. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Experimental Brain Research*, *166*(1), 23–30. doi:10.1007/s00221-005-2334-6

Friedman, L. (1995). The space factor in mathematics: Gender differences. *Review of Educational Research*, *65*(1), 22–50.

Friston, K. (2012). Ten ironic rules for non-statistical reviewers. *NeuroImage*, *61*(4), 1300–1310. doi:10.1016/j.neuroimage.2012.04.018

Fuentemilla, L., Càmara, E., Münte, T. F., Krämer, U. M., Cunillera, T., Marco-Pallarés, J., … Rodriguez-Fornells, A. (2009). Individual Differences in True and False Memory Retrieval Are Related to White Matter Brain Microstructure. *The Journal of Neuroscience*, *29*(27), 8698 –8703. doi:10.1523/JNEUROSCI.5270-08.2009

Gage, G. J., Parikh, H., & Marzullo, T. C. (2007). The Cingular Theory Of Unification: The Cingulate Cortex Does Everything. Retrieved from http://www-personal.umich.edu/~tmarzull/CingularManuscript.pdf

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, *116*(2), 439–453. doi:10.1037/a0015251

Gardiner, M. F., Fox, A., Knowles, F., & Jeffrey, D. (1996). Learning improved by arts training. *Nature*, *381*(6580), 284–284. doi:10.1038/381284a0

Gaser, C., Luders, E., Thompson, P. M., Lee, A. D., Dutton, R. A., Geaga, J. A., … Reiss, A. L. (2006). Increased local gyrification mapped in Williams syndrome. *NeuroImage*, *33*(1), 46–54. doi:10.1016/j.neuroimage.2006.06.018

Gaser, C., & Schlaug, G. (2003). Brain Structures Differ between Musicians and Non-Musicians. *Journal of Neuroscience*, *23*(27), 9240–9245.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*(2), 345–362. doi:10.1037/0033-2909.114.2.345

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. doi:10.1016/j.jecp.2004.03.002

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia: local diagnostic value. *Archives of Neurology & Psychiatry*, *44*(2), 398–408.

Gevers, W., Caessens, B., & Fias, W. (2005). Towards a common processing architecture underlying Simon and SNARC effects. *European Journal of Cognitive Psychology*, *17*(5), 659–673. doi:10.1080/09541440540000112

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*(3), B87–95.

Gevers, W., Reynvoet, B., & Fias, W. (2004). The Mental Representation of Ordinal Sequences is Spatially Organised: Evidence from Days of the Week. *Cortex*, *40*(1), 171–172. doi:10.1016/S0010-9452(08)70938-9

Göbel, S. M., Johansen-Berg, H., Behrens, T., & Rushworth, M. F. S. (2004). Response-Selection-Related Parietal Activation during Number Comparison. *Journal of Cognitive Neuroscience*, *16*(9), 1536–1551. doi:10.1162/0898929042568442

Göbel, S. M., & Rushworth, M. F. . (2004). Acting on Numbers. *Current Biology*, *14*(13), R517–R519. doi:10.1016/j.cub.2004.06.042

Göbel, S. M., Shaki, S., & Fischer, M. H. (2011). The Cultural Number Line: A Review of Cultural and Linguistic Influences on the Development of Number Processing. *Journal of Cross-Cultural Psychology*, *42*(4), 543–565. doi:10.1177/0022022111406251

Göbel, S. M., & Snowling, M. J. (2009). Number-processing skills in adults with dyslexia. *The Quarterly Journal of Experimental Psychology*, *63*(7), 1361–1373. doi:10.1080/17470210903359206

Good, C. D., Johnsrude, I., Ashburner, J., Henson, R. N. A., Friston, K. J., & Frackowiak, R. S. J. (2001). Cerebral Asymmetry and the Effects of Sex and Handedness on Brain Structure: A Voxel-Based Morphometric Analysis of 465 Normal Adult Human Brains. *NeuroImage*, *14*(3), 685–700. doi:06/nimg.2001.0857

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N., Fristen, K. J., & Frackowiak, R. S. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, *14*(1), 21–36.

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*(2), 604–608. doi:10.1016/j.neuropsychologia.2008.10.013

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. doi:10.1016/j.neuroimage.2007.07.041

Grabner, R. H., Ischebeck, A., Reishofer, G., Koschutnig, K., Delazer, M., Ebner, F., & Neuper, C. (2009). Fact learning in complex arithmetic and figural-spatial tasks: The role of the angular gyrus and its relation to mathematical competence. *Human Brain Mapping*, *30*(9), 2936–2952. doi:10.1002/hbm.20720

Graziano, A. B., Peterson, M., & Shaw, G. L. (1999). Enhanced learning of proportional math through music training and spatial-temporal training. *Neurological Research*, *21*(2), 139–152.

Greeno, J. G. (1991). Number Sense as Situated Knowing in a Conceptual Domain. *Journal for Research in Mathematics Education*, *22*(3), 170–218. doi:10.2307/749074

Greve, D. N. (2011). An absolute beginner's guide to surface and voxelbased morphometric analysis. In *ISMRM 19th Annual Meeting & Exhibition. Montréal, Québec, Canada* (pp. 7–13). Retrieved from http://cds.ismrm.org/protected/11MProceedings/files/ISMRM2011-8410.pdf

Hadamard, J. (1949). Psychology of Invention in the Field of Mathematics. *Princeton University Press, Princeton. OLIVER HEA VISIDE'S" DINNER*, *31*, 401.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the 'number sense': The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457–1465. doi:10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. ., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668.

Han, D. H., Lyoo, I. K., & Renshaw, P. F. (2012). Differential regional gray matter volumes in patients with on-line game addiction and professional gamers. *Journal of Psychiatric Research*, *46*(4), 507–515. doi:10.1016/j.jpsychires.2012.01.004

Hänggi, J., Koeneke, S., Bezzola, L., & Jäncke, L. (2010). Structural neuroplasticity in the sensorimotor network of professional female ballet dancers. *Human Brain Mapping*, *31*(8), 1196–1206. doi:10.1002/hbm.20928

Hauser, T. U., Grabner, R. H., & Jäncke, L. (2013). Enhancing performance in numerical magnitude processing and mental arithmetic using transcranial Direct Current Stimulation (tDCS). *Frontiers in Human Neuroscience*, *7*, 244. doi:10.3389/fnhum.2013.00244

Heathcote, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods*, *28*(3), 427–445. doi:10.3758/BF03200523

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*(2), 340–347. doi:10.1037/0033-2909.109.2.340

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395.

Hitch, G. J., & Baddeley, A. D. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology*, *28*(4), 603–621. doi:10.1080/14640747608400587

Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: syntactic and semantic structures. *Auditory Cognitive Neuroscience*, *2*, 112. doi:10.3389/fpsyg.2011.00112

Hoch, L., & Tillmann, B. (2012). Shared structural and temporal integration resources for music and arithmetic processing. *Acta Psychologica*, *140*(3), 230–235. doi:10.1016/j.actpsy.2012.03.008

Hogstrom, L. J., Westlye, L. T., Walhovd, K. B., & Fjell, A. M. (2012). The Structure of the Cerebral Cortex Across Adult Life: Age-Related Patterns of Surface Area, Thickness, and Gyrification. *Cerebral Cortex*, bhs231. doi:10.1093/cercor/bhs231

Holland, R., Leff, A. P., Josephs, O., Galea, J. M., Desikan, M., Price, C. J., … Crinion, J. (2011). Speech Facilitation by Left Inferior Frontal Cortex Stimulation. *Current Biology*, *21*(16), 1403–1407. doi:10.1016/j.cub.2011.07.021

Holloway, I. D., & Ansari, D. (2009a). Developmental Specialization in the Right Intraparietal Sulcus for the Abstract Representation of Numerical Magnitude. *Journal of Cognitive Neuroscience*, *22*(11), 2627–2637. doi:10.1162/jocn.2009.21399

Holloway, I. D., & Ansari, D. (2009b). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17–29. doi:10.1016/j.jecp.2008.04.001

Holmes, K. J., & Lourenco, S. F. (2012). Orienting numbers in mental space: Horizontal organization trumps vertical. *The Quarterly Journal of Experimental Psychology*, 1–8. doi:10.1080/17470218.2012.685079

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*(6), 435–448. doi:10.1038/nrn1684

Humphreys, G. W., Bickerton, W. L., Samson, D., & Riddoch, M. J. (2012). The Birmingham Cognitive Screen (BCoS). *Psychology Press, London*.

Hu, Y., Geng, F., Tao, L., Hu, N., Du, F., Fu, K., & Chen, F. (2011). Enhanced white matter tracts integrity in children with abacus training. *Human Brain Mapping*, *32*, 10–21.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical Training Shapes Structural Brain Development. *The Journal of Neuroscience*, *29*(10), 3019–3025. doi:10.1523/JNEUROSCI.5118-08.2009

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical Thickness in Congenital Amusia: When Less Is Better Than More. *The Journal of Neuroscience*, *27*(47), 13028 – 13032. doi:10.1523/JNEUROSCI.3039-07.2007

Imfeld, A., Oechslin, M. S., Meyer, M., Loenneker, T., & Jancke, L. (2009). White matter plasticity in the corticospinal tract of musicians: A diffusion tensor imaging study. *NeuroImage*, *46*(3), 600–607. doi:10.1016/j.neuroimage.2009.02.025

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review*, *18*(6), 1222–1229. doi:10.3758/s13423-011-0154-1

Ischebeck, A., Zamarian, L., Egger, K., Schocke, M., & Delazer, M. (2007). Imaging early practice effects in arithmetic. *NeuroImage*, *36*(3), 993–1003. doi:10.1016/j.neuroimage.2007.03.051

Ischebeck, A., Zamarian, L., Schocke, M., & Delazer, M. (2009). Flexible transfer of knowledge in mental arithmetic — An fMRI study. *NeuroImage*, *44*(3), 1103–1112. doi:10.1016/j.neuroimage.2008.10.025

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition*, *32*(4), 662–673. doi:10.3758/BF03195857

Iuculano, T., & Cohen Kadosh, R. (2013). The Mental Cost of Cognitive Enhancement. *The Journal of Neuroscience*, *33*(10), 4482–4486. doi:10.1523/JNEUROSCI.4927-12.2013

Iuculano, T., & Kadosh, R. C. (2014). Preliminary evidence for performance enhancement following parietal lobe stimulation in Developmental Dyscalculia. *Frontiers in Human Neuroscience*, *8:*, 38. doi:10.3389/fnhum.2014.00038

Iuculano, T., Rosenberg-Lee, M., Supekar, K., Lynch, C. J., Khouzam, A., Phillips, J., … Menon, V. (2014). Brain Organization Underlying Superior Mathematical Abilities in Children with Autism. *Biological Psychiatry*, *75*(3), 223–230. doi:10.1016/j.biopsych.2013.06.018

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833. doi:10.1073/pnas.0801268105

James, C. E., Oechslin, M. S., Ville, D. V. D., Hauert, C.-A., Descloux, C., & Lazeyras, F. (2013). Musical training intensity yields opposite effects on grey matter density in cognitive versus sensorimotor networks. *Brain Structure and Function*, *219*(1), 353–366. doi:10.1007/s00429-013-0504-z

Javadi, A. H., Brunec, I. K., Walsh, V., Penny, W. D., & Spiers, H. J. (2014). Transcranial electrical brain stimulation modulates neuronal tuning curves in perception of numerosity and duration. *NeuroImage*, *102, Part 2*, 451–457. doi:10.1016/j.neuroimage.2014.08.016

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

Johansen-Berg, H., & Rushworth, M. F. S. (2009). Using Diffusion Imaging to Study Human Connectional Anatomy. *Annual Review of Neuroscience*, *32*(1), 75–94. doi:10.1146/annurev.neuro.051508.135735

Jolles, D. D., & Crone, E. A. (2012). Training the developing brain: a neurocognitive perspective. *Frontiers in Human Neuroscience*, *6*(76), 1–13. doi:10.3389/fnhum.2012.00076

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*, *85*(2), 103–119. doi:10.1016/S0022-0965(03)00032-8

Jurcak, V., Okamoto, M., Singh, A., & Dan, I. (2005). Virtual 10–20 measurement on MR images for inter-modal linking of transcranial and tomographic neuroimaging methods. *NeuroImage*, *26*(4), 1184–1192. doi:10.1016/j.neuroimage.2005.03.021

Kasahara, K., Tanaka, S., Hanakawa, T., Senoo, A., & Honda, M. (2013). Lateralization of activity in the parietal cortex predicts the effectiveness of bilateral transcranial direct current stimulation on performance of a mental calculation task. *Neuroscience Letters*, *545*, 86–90. doi:10.1016/j.neulet.2013.04.022

Kaufmann, L., Koppelstaetter, F., Siedentopf, C., Haala, I., Haberlandt, E., Zimmerhackl, L.-B., … Ischebeck, A. (2006). Neural correlates of the number-size interference task in children. *Neuroreport*, *17*(6), 587–591.

Kawashima, R., Taira, M., Okita, K., Inoue, K., Tajima, N., Yoshida, H., … Fukuda, H. (2004). A functional MRI study of simple arithmetic—a comparison between children and adults. *Cognitive Brain Research*, *18*(3), 227–233. doi:10.1016/j.cogbrainres.2003.10.009

Keeser, D., Meindl, T., Bor, J., Palm, U., Pogarell, O., Mulert, C., … Padberg, F. (2011). Prefrontal Transcranial Direct Current Stimulation Changes Connectivity of Resting-State Networks during fMRI. *The Journal of Neuroscience*, *31*(43), 15284–15293. doi:10.1523/JNEUROSCI.0542-11.2011

Kelly, A. M. C., & Garavan, H. (2005). Human Functional Neuroimaging of Brain Changes Associated with Practice. *Cerebral Cortex*, *15*(8), 1089–1102. doi:10.1093/cercor/bhi005

Keus, I. M., Jenks, K. M., & Schwarz, W. (2005). Psychophysiological evidence that the SNARC effect has its functional locus in a response selection stage. *Cognitive Brain Research*, *24*(1), 48–56.

Kim, M. J., & Whalen, P. J. (2009). The Structural Integrity of an Amygdala–Prefrontal Pathway Predicts Trait Anxiety. *The Journal of Neuroscience*, *29*(37), 11614 –11618. doi:10.1523/JNEUROSCI.2335-09.2009

Kleim, J. A., Hogg, T. M., VandenBerg, P. M., Cooper, N. R., Bruneau, R., & Remple, M. (2004). Cortical synaptogenesis and motor map reorganization occur during late, but not early, phase of motor skill learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *24*(3), 628–633. doi:10.1523/JNEUROSCI.3440-03.2004

Kleim, J. A., Lussnig, E., Schwarz, E. R., Comery, T. A., & Greenough, W. T. (1996). Synaptogenesis and Fos expression in the motor cortex of the adult rat after motor skill learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *16*(14), 4529–4535.

Klein, A., Ghosh, S. S., Avants, B., Yeo, B. T. T., Fischl, B., Ardekani, B., … Parsey, R. V. (2010). Evaluation of volume-based and surface-based brain image registration methods. *NeuroImage*, *51*(1), 214–220. doi:10.1016/j.neuroimage.2010.01.091

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), e55455. doi:10.1371/journal.pone.0055455

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*(7), 317–324. doi:10.1016/j.tics.2010.05.002

Knops, A., Thirion, B., Hubbard, E. M., Michel, V., & Dehaene, S. (2009). Recruitment of an Area Involved in Eye Movements During Mental Arithmetic. *Science*, *324*(5934), 1583–1585. doi:10.1126/science.1171599

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between Syntax Processing in Language and in Music: An ERP Study. *Journal of Cognitive Neuroscience*, *17*(10), 1565–1577. doi:10.1162/089892905774597290

Kolb, B., Cioe, J., & Comeau, W. (2008). Contrasting effects of motor and visual spatial learning tasks on dendritic arborization and spine density in rats. *Neurobiology of Learning and Memory*, *90*(2), 295–300. doi:10.1016/j.nlm.2008.04.012

Kong, J., Wang, C., Kwong, K., Vangel, M., Chua, E., & Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Cognitive Brain Research*, *22*(3), 397–405. doi:10.1016/j.cogbrainres.2004.09.011

Koontz, K. L. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition*, *2*(1), 1–24.

Krause, B., & Cohen Kadosh, R. (2013). Can transcranial electrical stimulation improve learning difficulties in atypical brain development? A future possibility for cognitive training. *Developmental Cognitive Neuroscience*, *6*, 176–194. doi:10.1016/j.dcn.2013.04.001

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., … von Aster, M. (2011). Mental number line training in children with developmental dyscalculia. *NeuroImage*, *57*(3), 782–795. doi:10.1016/j.neuroimage.2011.01.070

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. doi:10.1080/87565640802101474

Kuo, M.-F., & Nitsche, M. A. (2012). Effects of Transcranial Electrical Stimulation on Cognition. *Clinical EEG and Neuroscience*, *43*(3), 192–199.

Kuo, M.-F., Unger, M., Liebetanz, D., Lang, N., Tergau, F., Paulus, W., & Nitsche, M. A. (2008). Limited impact of homeostatic plasticity on motor learning in humans. *Neuropsychologia*, *46*(8), 2122–2128. doi:10.1016/j.neuropsychologia.2008.02.023

Kwon, Y. H., Ko, M.-H., Ahn, S. H., Kim, Y.-H., Song, J. C., Lee, C.-H., … Jang, S. H. (2008). Primary motor cortex activation by transcranial direct current stimulation in the human brain. *Neuroscience Letters*, *435*(1), 56–59. doi:10.1016/j.neulet.2008.02.012

Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in Quantitative Methods for Psychology*, *4*(1), 35–45.

Lakoff, G., & Johnson, M. (2008). *Metaphors We Live By*. University of Chicago Press.

Lavidor, M., Brinksman, V., & Göbel, S. M. (2004). Hemispheric asymmetry and the mental number line: comparison of double-digit numbers. *Neuropsychologia*, *42*(14), 1927–1933. doi:10.1016/j.neuropsychologia.2004.05.003

Lee, H., Boot, W. R., Basak, C., Voss, M. W., Prakash, R. S., Neider, M., … Kramer, A. F. (2012). Performance gains from directed training do not transfer to untrained tasks. *Acta Psychologica*, *139*(1), 146–158. doi:10.1016/j.actpsy.2011.11.003

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. doi:10.1002/1531-8249(200010)48:4<657::AID-ANA13>3.0.CO;2-K

Lee, Y.-S., Janata, P., Frost, C., Hanke, M., & Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *NeuroImage*, *57*(1), 293–300. doi:10.1016/j.neuroimage.2011.02.006

LeFevre, J.-A., DeStefano, D., Penner-Wilger, M., & Daley, K. E. (2006). Selection of procedures in mental subtraction. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *60*(3), 209–220. doi:10.1037/cjep2006020

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00375

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*(1), 83–97. doi:10.1037/0096-3445.124.1.83

Levine, D. R. (1982). Strategy Use and Estimation Ability of College Students. *Journal for Research in Mathematics Education*, *13*(5), 350–359. doi:10.2307/749010

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*(6), 1292–1300. doi:10.1111/j.1467-7687.2011.01080.x

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1189–1207. doi:10.1037/0096-1523.33.5.1189

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain*, *133*(6), 1682–1693. doi:10.1093/brain/awq089

Lochner, C., Fouché, J.-P., du Plessis, S., Spottiswoode, B., Seedat, S., (Psych), Mm., … Stein, D. J. (2012). Evidence for fractional anisotropy and mean diffusivity white matter abnormalities in the internal capsule and cingulum in patients with obsessive–compulsive disorder. *Journal of Psychiatry & Neuroscience : JPN*, *37*(3), 193–199. doi:10.1503/jpn.110059

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 179–195. doi:10.1037/0278-7393.17.2.179

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 149–157. doi:10.1037/0278-7393.16.1.149

Loui, P., Hohmann, A., & Schlaug, G. (2010). Inducing Disorders in Pitch Perception and Production: a Reverse-Engineering Approach. *Proceedings of Meetings on Acoustics Acoustical Society of America*, *9*(1), 50002. doi:10.1121/1.3431713

Luce, R. D. (1986). *Response times and their role in inferring elementary mental organization*. Oxford University Press.

Luders, E., Kurth, F., Mayer, E. A., Toga, A. W., Narr, K. L., & Gaser, C. (2012). The unique brain anatomy of meditation practitioners: alterations in cortical gyrification. *Frontiers in Human Neuroscience*, *6*, 34. doi:10.3389/fnhum.2012.00034

Lyons, I. M., Ansari, D., & Beilock, S. L. (2012). Symbolic estrangement: evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology. General*, *141*(4), 635–641. doi:10.1037/a0027248

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(8), 4398 –4403.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*(1), 1–22. doi:10.1037/0096-3445.111.1.1

Mapelli, D., Rusconi, E., & Umiltà, C. (2003). The SNARC effect: an instance of the Simon effect? *Cognition*, *88*(3), B1–B10. doi:10.1016/S0010-0277(03)00042-8

Marghetis, T., Walker, E. J., Bergen, B. K., & Núñez, R. (2011). Making SNAP Judgments: Rethinking the Spatial Representation of Number. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (p. 1781-1786). Cognitive Science Society.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Impaired Acuity of the Approximate Number System Underlies Mathematical Learning Disability (Dyscalculia). *Child Development*, *82*(4), 1224–1237. doi:10.1111/j.1467-8624.2011.01608.x

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011b). Preschoolers' Precision of the Approximate Number System Predicts Later School Mathematics Performance. *PLoS ONE*, *6*(9), e23749. doi:10.1371/journal.pone.0023749

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. doi:10.1037/0033-295X.102.3.419

Menon, V., Rivera, S. M., White, C. D., Glover, G. H., & Reiss, A. L. (2000). Dissociating Prefrontal and Parietal Cortex Activation during Arithmetic Processing. *NeuroImage*, *12*(4), 357–365. doi:10.1006/nimg.2000.0613

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, *27*, 272–277.

Moliadze, V., Antal, A., & Paulus, W. (2010). Electrode-distance dependent after-effects of transcranial direct and random noise stimulation with extracephalic reference electrodes. *Clinical Neurophysiology*, *121*(12), 2165–2171. doi:10.1016/j.clinph.2010.04.033

Molko, N., Cachia, A., Rivière, 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*(4), 847–858. doi:10.1016/S0896-6273(03)00670-6

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520.

Mulquiney, P. G., Hoy, K. E., Daskalakis, Z. J., & Fitzgerald, P. B. (2011). Improving working memory: Exploring the effect of transcranial random noise stimulation and transcranial direct current stimulation on the dorsolateral prefrontal cortex. *Clinical Neurophysiology*, *122*(12), 2384–2389. doi:10.1016/j.clinph.2011.05.009

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*(4), 490–502. doi:10.1016/j.jecp.2009.02.003

Mussolin, C., Mejias, S., & Noël, M.-P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition*, *115*(1), 10–25. doi:10.1016/j.cognition.2009.10.006

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In *Cognitive skills and their acquisition*.

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1204580109

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, *297*(5587), 1708.

Nieder, A., & Miller, E. K. (2003). Coding of Cognitive Magnitude: Compressed Scaling of Numerical Information in the Primate Prefrontal Cortex. *Neuron*, *37*(1), 149–157. doi:10.1016/S0896-6273(02)01144-3

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences*, *101*(19), 7457.

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., … others. (2008). Transcranial direct current stimulation: State of the art 2008. *Brain Stimulation*, *1*(3), 206–23.

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of Physiology*, *527*(3), 633–639. doi:10.1111/j.1469-7793.2000.t01-1-00633.x

Nitsche, M. A., & Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology*, *57*(10), 1899 –1901.

Nitsche, M. A., Seeber, A., Frommann, K., Klein, C. C., Rochford, C., Nitsche, M. S., … Tergau, F. (2005). Modulating parameters of excitability during and after transcranial direct current stimulation of the human motor cortex. *The Journal of Physiology*, *568*(1), 291–303. doi:10.1113/jphysiol.2005.092429

Noël, M.-P., & Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: an explanation based on a double exact-and-approximate number representation model. *Frontiers in Human Neuroscience*, *5*, 165. doi:10.3389/fnhum.2011.00165

Nuerk, H.-C., Wood, G., & Willmes, K. (2005). The Universal SNARC Effect. *Experimental Psychology (formerly 'Zeitschrift Für Experimentelle Psychologie')*, *52*(3), 187–194. doi:10.1027/1618-3169.52.3.187

O'Boyle, M. W. (2008). Mathematically Gifted Children: Developmental Brain Characteristics and Their Prognosis for Well-Being. *Roeper Review*, *30*(3), 181–186. doi:10.1080/02783190802199594

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7*(1), 75–79.

Orsini, A., Grossi, D., Capitani, E., Laiacona, M., Papagno, C., & Vallar, D. G. (1987). Verbal and spatial immediate memory span: Normative data from 1355 adults and 1112 children. *The Italian Journal of Neurological Sciences*, *8*(6), 537–548. doi:10.1007/BF02333660

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., … Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–778. doi:10.1038/nature09042

Pachella, R. G. (1974). The Interpretation of Reaction Time in Information Processing Research. In *Human information processing: Tutorials in performance and cognition*.

Palaniyappan, L., Mallikarjun, P., Joseph, V., White, T. P., & Liddle, P. F. (2011). Folding of the Prefrontal Cortex in Schizophrenia: Regional Differences in Gyrification. *Biological Psychiatry*, *69*(10), 974–979. doi:10.1016/j.biopsych.2010.12.012

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?*. National Research and Development Centre for Adult Literacy and Numeracy.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The Plastic Human Brain Cortex. *Annual Review of Neuroscience*, *28*(1), 377–401. doi:10.1146/annurev.neuro.27.070203.144216

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, *88*(4), 348–367. doi:10.1016/j.jecp.2004.04.002

Pauli, P., Lutzenberger, W., Rau, H., Birbaumer, N., Rickard, T. C., Yaroush, R. A., & Bourne Jr., L. E. (1994). Brain potentials during mental arithmetic: effects of extensive practice and problem difficulty. *Cognitive Brain Research*, *2*(1), 21–29. doi:10.1016/0926-6410(94)90017-5

Penner-Wilger, M., Fast, L., LeFevre, J., Smith-Chant, B. L., Skwarchuk, S., Kamawar, D., & Bisanz, J. (2007). The foundations of numeracy: Subitizing, finger gnosia, and fine-motor ability. In *Proceedings of the 29th annual cognitive science society* (pp. 1385–1390). Cognitive Science Society Austin, TX. Retrieved from http://csjarchive.cogsci.rpi.edu/proceedings/2007/docs/p1385.pdf

Pesenti, M. (2005). Calculation abilities in expert calculators. *Handbook of Mathematical Cognition*, 413–430.

Pesenti, M., Zago, L., Crivello, F., Mellet, E., Samson, D., Duroux, B., … Tzourio-Mazoyer, N. (2001). Mental calculation in a prodigy is sustained by right prefrontal and medial temporal areas. *Nature Neuroscience*, *4*(1), 103–108.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A Redrawn Vandenberg and Kuse Mental Rotations Test - Different Versions and Factors That Affect Performance. *Brain and Cognition*, *28*(1), 39–58. doi:10.1006/brcg.1995.1032

Pfurtscheller, G., Bauernfeind, G., Wriessnegger, S. C., & Neuper, C. (2010). Focal frontal (de)oxyhemoglobin responses during simple arithmetic. *International Journal of Psychophysiology*, *76*(3), 186–192. doi:10.1016/j.ijpsycho.2010.03.013

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., … Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, 33–41.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex. *Neuron*, *53*(2), 293–305. doi:10.1016/j.neuron.2006.11.022

Pimperton, H., & Nation, K. (2010). Understanding words, understanding numbers: An exploration of the mathematical profiles of poor comprehenders. *British Journal of Educational Psychology*, *80*(2), 255–268. doi:10.1348/000709909X477251

Pinel, P., Dehaene, S., Rivière, D., & LeBihan, D. (2001). Modulation of Parietal Activation by Semantic Distance in a Number Comparison Task. *NeuroImage*, *14*(5), 1013–1026. doi:10.1006/nimg.2001.0913

Polanía, R., Paulus, W., Antal, A., & Nitsche, M. A. (2011). Introducing graph theory to track for neuroplastic alterations in the resting human brain: A transcranial direct current stimulation study. *NeuroImage*, *54*(3), 2287–2296. doi:10.1016/j.neuroimage.2010.09.085

Pope, P. A., & Miall, R. C. (2012). Task-specific facilitation of cognition by cathodal transcranial direct current stimulation of the cerebellum. *Brain Stimulation*, *5*(2), 84–94. doi:10.1016/j.brs.2012.03.006

Poreisz, C., Boros, K., Antal, A., & Paulus, W. (2007). Safety aspects of transcranial direct current stimulation concerning healthy subjects and patients. *Brain Research Bulletin*, *72*(4–6), 208–214. doi:10.1016/j.brainresbull.2007.01.004

Previtali, P., Hevia, M. D. de, & Girelli, L. (2010). Placing order in space: the SNARC effect in serial learning. *Experimental Brain Research*, *201*(3), 599–605. doi:10.1007/s00221-009-2063-3

Price, G. R., Holloway, I., Räsänen, P., Vesterinen, M., & Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Current Biology*, *17*(24), R1042–R1043. doi:10.1016/j.cub.2007.10.013

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*(1), 50–57. doi:10.1016/j.actpsy.2012.02.008

Prichard, G., Weiller, C., Fritsch, B., & Reis, J. (2014). Effects of Different Electrical Brain Stimulation Protocols on Subcomponents of Motor Skill Learning. *Brain Stimulation*. doi:10.1016/j.brs.2014.04.005

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416–442. doi:10.1037/0033-2909.132.3.416

Putnam, M. C., Steven, M. S., Doron, K. W., Riggall, A. C., & Gazzaniga, M. S. (2009). Cortical Projection Topography of the Human Splenium: Hemispheric Asymmetry and Individual Differences. *Journal of Cognitive Neuroscience*, *22*(8), 1662–1669. doi:10.1162/jocn.2009.21290

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*, 1263–1269. doi:10.1038/nn.3788

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*(2), 110–122. doi:10.1016/j.lindif.2009.10.005

Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., … Krakauer, J. W. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proceedings of the National Academy of Sciences*, *106*(5), 1590–1595. doi:10.1073/pnas.0805413106

Reuhkala, M. (2001). Mathematical skills in ninth-graders: Relationship with visuo-spatial abilities and working memory. *Educational Psychology*, *21*(4), 387–399.

Reuter, M., & Fischl, B. (2011). Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage*, *57*(1), 19–21.

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, *61*(4), 1402–1418. doi:10.1016/j.neuroimage.2012.02.084

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, *126*(3), 288–311. doi:10.1037/0096-3445.126.3.288

Rickard, T. C. (2004). Strategy Execution in Cognitive Skill Learning: An Item-Level Test of Candidate Models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 65–82. doi:10.1037/0278-7393.30.1.65

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. doi:10.1093/cercor/bhi055

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2007). Monotonic Coding of Numerosity in Macaque Lateral Intraparietal Area. *PLoS Biol*, *5*(8), e208. doi:10.1371/journal.pbio.0050208

Roosendaal, S. D., Geurts, J. J. G., Vrenken, H., Hulst, H. E., Cover, K. S., Castelijns, J. A., … Barkhof, F. (2009). Regional DTI differences in multiple sclerosis patients. *NeuroImage*, *44*(4), 1397–1403. doi:10.1016/j.neuroimage.2008.10.026

Rosas, H. D., Liu, A. K., Hersch, S., Glessner, M., Ferrante, R. J., Salat, D. H., … Fischl, B. (2002). Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology*, *58*(5), 695–701.

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*(1), 417–422. doi:10.1016/j.neuroimage.2007.08.045

Rouiller, E. M., Babalian, A., Kazennikov, O., Moret, V., Yu, X.-H., & Wiesendanger, M. (1994). Transcallosal connections of the distal forelimb representations of the primary and supplementary motor cortical areas in macaque monkeys. *Experimental Brain Research*, *102*(2), 227–243. doi:10.1007/BF00227511

Roux, F.-E., Boetto, S., Sacko, O., Chollet, F., & Trémoulet, M. (2003). Writing, calculating, and finger recognition in the region of the angular gyrus: a cortical stimulation study of Gerstmann syndrome. *Journal of Neurosurgery*, *99*(4), 716–727.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*(2), 113–129. doi:10.1016/j.cognition.2005.01.004

Rykhlevskaia, E., Uddin, L. Q., Kondos, L., & Menon, V. (2009). Neuroanatomical Correlates of Developmental Dyscalculia: Combined Evidence from Morphometry and Tractography. *Frontiers in Human Neuroscience*, *3*(51), 1–13. doi:10.3389/neuro.09.051.2009

Sagi, Y., Tavor, I., Hofstetter, S., Tzur-Moryosef, S., Blumenfeld-Katzir, T., & Assaf, Y. (2012). Learning in the Fast Lane: New Insights into Neuroplasticity. *Neuron*, *73*(6), 1195–1203. doi:10.1016/j.neuron.2012.01.025

Sampaio-Baptista, C., Scholz, J., Jenkinson, M., Thomas, A. G., Filippini, N., Smit, G., … Johansen-Berg, H. (2014). Gray matter volume is associated with rate of subsequent skill learning after a long term training intervention. *NeuroImage*. doi:10.1016/j.neuroimage.2014.03.056

Sandrini, M., Rossini, P. M., & Miniussi, C. (2004). The differential involvement of inferior parietal lobule in number comparison: a rTMS study. *Neuropsychologia*, *42*(14), 1902–1909. doi:10.1016/j.neuropsychologia.2004.05.005

Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition*, *108*(1), 263–270. doi:10.1016/j.cognition.2008.01.002

Santens, S., Roggeman, C., Fias, W., & Verguts, T. (2010). Number Processing Pathways in Human Parietal Cortex. *Cerebral Cortex*, *20*(1), 77 –88. doi:10.1093/cercor/bhp080

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, *30*(2), 344–357. doi:10.1111/j.2044-835X.2011.02048.x

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, *114*(3), 418–431. doi:10.1016/j.jecp.2012.10.012

Sasanguie, D., Van den Bussche, E., & Reynvoet, B. (2012). Predictors for Mathematics Achievement? Evidence From a Longitudinal Study. *Mind, Brain, and Education*, *6*(3), 119–128. doi:10.1111/j.1751-228X.2012.01147.x

Schaer, M., Cuadra, M. B., Tamarit, L., Lazeyras, F., Eliez, S., & Thiran, J. (2008). A Surface-Based Approach to Quantify Local Cortical Gyrification. *IEEE Transactions on Medical Imaging*, *27*(2), 161–170. doi:10.1109/TMI.2007.903576

Schaer, M., & Eliez, S. (2009). Contribution of Structural Brain Imaging to Our Understanding of the Cortical Development Process. *European Psychiatry Reviews*, *2*, 13–16.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*(3), 414–429. doi:10.1037/0096-3445.136.3.414

Schmithorst, V. J., & Holland, S. K. (2003). The effect of musical training on music processing: a functional magnetic resonance imaging study in humans. *Neuroscience Letters*, *348*(2), 65–68. doi:10.1016/S0304-3940(03)00714-6

Schmithorst, V. J., & Wilke, M. (2002). Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neuroscience Letters*, *321*(1-2), 57–60. doi:10.1016/S0304-3940(02)00054-X

Scholz, J., Klein, M. C., Behrens, T. E. J., & Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nature Neuroscience*, *12*(11), 1370–1371. doi:10.1038/nn.2412

Schulz, R., Zimerman, M., Timmermann, J., Wessel, M. J., Gerloff, C., & Hummel, F. C. (2013). White matter integrity of motor connections related to training gains in healthy aging. *Neurobiology of Aging*. doi:10.1016/j.neurobiolaging.2013.11.024

Schwarz, W., & Keus, I. M. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Perception & Psychophysics*, *66*(4), 651–664. doi:10.3758/BF03194909

Scripture, E. W. (1891). Arithmetical prodigies. *The American Journal of Psychology*, *4*(1), 1–59.

Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J. P., Greenstein, D., … Rapoport, J. L. (2007). Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19649–19654. doi:10.1073/pnas.0707741104

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, *89*(4), 305–333. doi:10.1037/0033-295X.89.4.305

Shuter-Dyson, R. (1968). *The psychology of musical ability*. Methuen.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*(3), 258–275. doi:10.1037/0096-3445.117.3.258

Siegler, R. S., & Booth, J. L. (2004). Development of Numerical Estimation in Young Children. *Child Development*, *75*(2), 428–444. doi:10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., & Opfer, J. E. (2003). The Development of Numerical Estimation: Evidence for Multiple Representations of Numerical Quantity. *Psychological Science*, *14*(3), 237 –250. doi:10.1111/1467-9280.02438

Silvanto, J., Muggleton, N., & Walsh, V. (2008). State-dependency in brain stimulation studies of perception and cognition. *Trends in Cognitive Sciences*, *12*(12), 447–454. doi:16/j.tics.2008.09.004

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*(3), 475–487.

Simos, P. G., Kanatsouli, K., Fletcher, J. M., Sarkari, S., Juranek, J., Cirino, P., … Papanicolaou, A. C. (2008). Aberrant spatiotemporal activation profiles associated with math difficulties in children: A magnetic source imaging study. *Neuropsychology*, *22*(5), 571–584. doi:10.1037/0894-4105.22.5.571

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-Based Morphometry Reveals Increased Gray Matter Density in Broca's Area in Male Symphony Orchestra Musicians. *NeuroImage*, *17*(3), 1613–1622. doi:10.1006/nimg.2002.1288

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., … Behrens, T. E. J. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, *31*(4), 1487–1505. doi:10.1016/j.neuroimage.2006.02.024

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., … Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23, Supplement 1*(0), S208–S219. doi:10.1016/j.neuroimage.2004.07.051

Snowball, A., Tachtsidis, I., Popescu, T., Thompson, J., Delazer, M., Zamarian, L., … Cohen Kadosh, R. (2013). Long-Term Enhancement of Brain Function and Cognition Using Cognitive Training and Brain Stimulation. *Current Biology*. doi:10.1016/j.cub.2013.04.045

Soltész, F., Szűcs, D., & Szűcs, L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: A developmental study. *Behavioral and Brain Functions*, *6*(1), 13. doi:10.1186/1744-9081-6-13

Stagg, C. J., Jayaram, G., Pastor, D., Kincses, Z. T., Matthews, P. M., & Johansen-Berg, H. (2011). Polarity and timing-dependent effects of transcranial direct current stimulation in explicit motor learning. *Neuropsychologia*, *49*(5), 800–804. doi:10.1016/j.neuropsychologia.2011.02.009

Stagg, C. J., O'Shea, J., Kincses, Z. T., Woolrich, M., Matthews, P. M., & Johansen-Berg, H. (2009). Modulation of movement-associated cortical activation by transcranial direct current stimulation. *European Journal of Neuroscience*, *30*(7), 1412–1423. doi:10.1111/j.1460-9568.2009.06937.x

Stanescu-Cosson, R., Pinel, P., Moortele, P.-F. van de, Bihan, D. L., 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*(11), 2240–2255. doi:10.1093/brain/123.11.2240

Steele, C. J., Bailey, J. A., Zatorre, R. J., & Penhune, V. B. (2013). Early Musical Training and White-Matter Plasticity in the Corpus Callosum: Evidence for a Sensitive Period. *The Journal of Neuroscience*, *33*(3), 1282–1290. doi:10.1523/JNEUROSCI.3578-12.2013

Steinbeis, N., & Koelsch, S. (2008). Shared Neural Resources between Music and Language Indicate Semantic Processing of Musical Tension-Resolution Patterns. *Cerebral Cortex*, *18*(5), 1169–1178. doi:10.1093/cercor/bhm149

Stein, J. F. (1989). Representation of egocentric space in the posterior parietal cortex. *Quarterly Journal of Experimental Physiology (Cambridge, England)*, *74*(5), 583–606.

Styner, M. A., Oguz, I., Smith, R. G., Cascio, C., & Jomier, M. (2005). Corpus Callosum Subdivision Based on a Probabilistic Model of Inter-hemispheric Connectivity. In J. S. Duncan & G. Gerig (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005* (pp. 765–772). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/11566489_94

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*. doi:10.1073/pnas.1222154110

Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Developmental Psychology*, *4*, 444. doi:10.3389/fpsyg.2013.00444

Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2001). *Using multivariate statistics*.

Takeuchi, H., Taki, Y., Sassa, Y., Hashizume, H., Sekiguchi, A., Fukushima, A., & Kawashima, R. (2011). Working Memory Training Using Mental Calculation Impacts Regional Gray Matter of the Frontal and Parietal Regions. *PLoS ONE*, *6*(8), e23175. doi:10.1371/journal.pone.0023175

Tamm, L., Menon, V., & Reiss, A. L. (2002). Maturation of Brain Function Associated With Response Inhibition. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(10), 1231–1238. doi:10.1097/00004583-200210000-00013

Tanida, M., Sakatani, K., Takano, R., & Tagai, K. (2004). Relation between asymmetry of prefrontal cortex activities and the autonomic nervous system during a mental arithmetic task: near infrared spectroscopy study. *Neuroscience Letters*, *369*(1), 69–74. doi:10.1016/j.neulet.2004.07.076

Taubert, M., Draganski, B., Anwander, A., Müller, K., Horstmann, A., Villringer, A., & Ragert, P. (2010). Dynamic Properties of Human Brain Structure: Learning-Related Changes in Cortical Areas and

Associated Fiber Connections. *The Journal of Neuroscience*, *30*(35), 11670–11677. doi:10.1523/JNEUROSCI.2567-10.2010

Terhune, D. B., Tai, S., Cowey, A., Popescu, T., & Cohen Kadosh, R. (2011). Enhanced Cortical Excitability in Grapheme-Color Synesthesia and Its Modulation. *Current Biology*, *21*(23), 2006–2009. doi:10.1016/j.cub.2011.10.032

Terney, D., Chaieb, L., Moliadze, V., Antal, A., & Paulus, W. (2008). Increasing Human Brain Excitability by Transcranial High-Frequency Random Noise Stimulation. *The Journal of Neuroscience*, *28*(52), 14147 – 14155. doi:10.1523/JNEUROSCI.4248-08.2008

Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & Worden, M. S. (2004). Evidence of Developmental Differences in Implicit Sequence Learning: An fMRI Study of Children and Adults. *Journal of Cognitive Neuroscience*, *16*(8), 1339–1351. doi:10.1162/0898929042304688

Thompson, P. M., Hayashi, K. M., Sowell, E. R., Gogtay, N., Giedd, J. N., Rapoport, J. L., … Toga, A. W. (2004). Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. *NeuroImage*, *23 Suppl 1*, S2–18. doi:10.1016/j.neuroimage.2004.07.071

Thorndike, E. L. (1922). *The psychology of arithmetic*. The Macmillan Company.

Tillmann, B., Jolicœur, P., Ishihara, M., Gosselin, N., Bertrand, O., Rossetti, Y., & Peretz, I. (2010). The Amusic Brain: Lost in Music, but Not in Space. *PLoS ONE*, *5*(4), e10173. doi:10.1371/journal.pone.0010173#pone.0010173-Mudd1

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. *Proceedings of the National Academy of Sciences*, *106*(52), 22546 –22551. doi:10.1073/pnas.0906094106

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., … Morris, R. G. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76–82.

Tucholka, A., Fritsch, V., Poline, J.-B., & Thirion, B. (2012). An empirical comparison of surface-based and volume-based group studies in neuroimaging. *NeuroImage*, *63*(3), 1443–1453. doi:10.1016/j.neuroimage.2012.06.019

Ueda, K. (1987). Perceptual components of pitch: Spatial representation using a multidimensional scaling technique. *The Journal of the Acoustical Society of America*, *82*(4), 1193. doi:10.1121/1.395255

Utz, K. S., Dimova, V., Oppenländer, K., & Kerkhoff, G. (2010). Electrified minds: Transcranial direct current stimulation (tDCS) and Galvanic Vestibular Stimulation (GVS) as methods of non-invasive brain stimulation in neuropsychology—A review of current data and future implications. *Neuropsychologia*, *48*(10), 2789–2810. doi:10.1016/j.neuropsychologia.2010.06.002

Van Dijck, J.-P., & Fias, W. (2011). A working memory account for spatial–numerical associations. *Cognition*, *119*(1), 114–119. doi:10.1016/j.cognition.2010.12.013

Van Dijck, J.-P., Gevers, W., & Fias, W. (2009). Numbers are associated with different types of spatial information depending on the task. *Cognition*, *113*(2), 248–253. doi:10.1016/j.cognition.2009.08.005

Van Eimeren, L., Niogi, S. N., McCandliss, B. D., Holloway, I. D., & Ansari, D. (2008). White matter microstructures underlying mathematical abilities in children. *NeuroReport*, *19*(11), 1117–1121. doi:10.1097/WNR.0b013e328307f5c1

Van Opstal, F., Fias, W., Peigneux, P., & Verguts, T. (2009). The neural representation of extensively trained ordered sequences. *Neuroimage*, *47*(1), 367–375.

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience*, *16*(9), 1493–1504.

Voets, N. L., Hough, M. G., Douaud, G., Matthews, P. M., James, A., Winmill, L., … Smith, S. (2008). Evidence for abnormalities of cortical development in adolescent-onset schizophrenia. *NeuroImage*, *43*(4), 665–675. doi:10.1016/j.neuroimage.2008.08.013

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. doi:10.1016/j.tics.2003.09.002

Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature Reviews Neuroscience*, *1*(1), 73–80. doi:10.1038/35036239

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences*, *100*(17), 10038 –10042. doi:10.1073/pnas.1730682100

Wason, P. C. (1966). Reasoning. *New Horizons in Psychology*, *1*, 135–151.

Wechsler, D. (1997). *WAIS-III/WMS-III technical manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999). Manual for the Wechsler abbreviated intelligence scale (WASI). *San Antonio, Tex: The Psychological Corporation*.

Wechsler, D. (2005). *Wechsler Individual Achievement Test (WIAT-II UK)*. Harcourt Assessment.

Wilke, M., Kowatch, R. A., DelBello, M. P., Mills, N. P., & Holland, S. K. (2004). Voxel-based morphometry in adolescents with bipolar disorder: first results. *Psychiatry Research*, *131*(1), 57–69. doi:10.1016/j.pscychresns.2004.01.004

Williamson, V. J., Cocchini, G., & Stewart, L. (2011). The relationship between pitch and space in congenital amusia. *Brain and Cognition*, *76*(1), 70–76. doi:10.1016/j.bandc.2011.02.016

Wilson, F. A., Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science (New York, N.Y.)*, *260*(5116), 1955–1958.

Wilson, K. M., & Swanson, H. L. (2001). Are Mathematics Disabilities Due to a Domain-General or a Domain-Specific Working Memory Deficit? *Journal of Learning Disabilities*, *34*(3), 237 –248. doi:10.1177/002221940103400304

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., … Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, *53*(3), 1135–1146. doi:10.1016/j.neuroimage.2009.12.028

Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science*, *50*(4), 489–525.

Wu, S. 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*(12), 2930 –2945. doi:10.1093/cercor/bhp063

Yook, S.-W., Park, S.-H., Seo, J.-H., Kim, S.-J., & Ko, M.-H. (2011). Suppression of Seizure by Cathodal Transcranial Direct Current Stimulation in an Epileptic Patient - A Case Report -. *Annals of Rehabilitation Medicine*, *35*(4), 579. doi:10.5535/arm.2011.35.4.579

Zamarian, L., Ischebeck, A., & Delazer, M. (2009). Neuroscience of learning arithmetic - Evidence from brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *33*(6), 909–925.

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature Neuroscience*, *15*(4), 528–536. doi:10.1038/nn.3045

8 Appendices

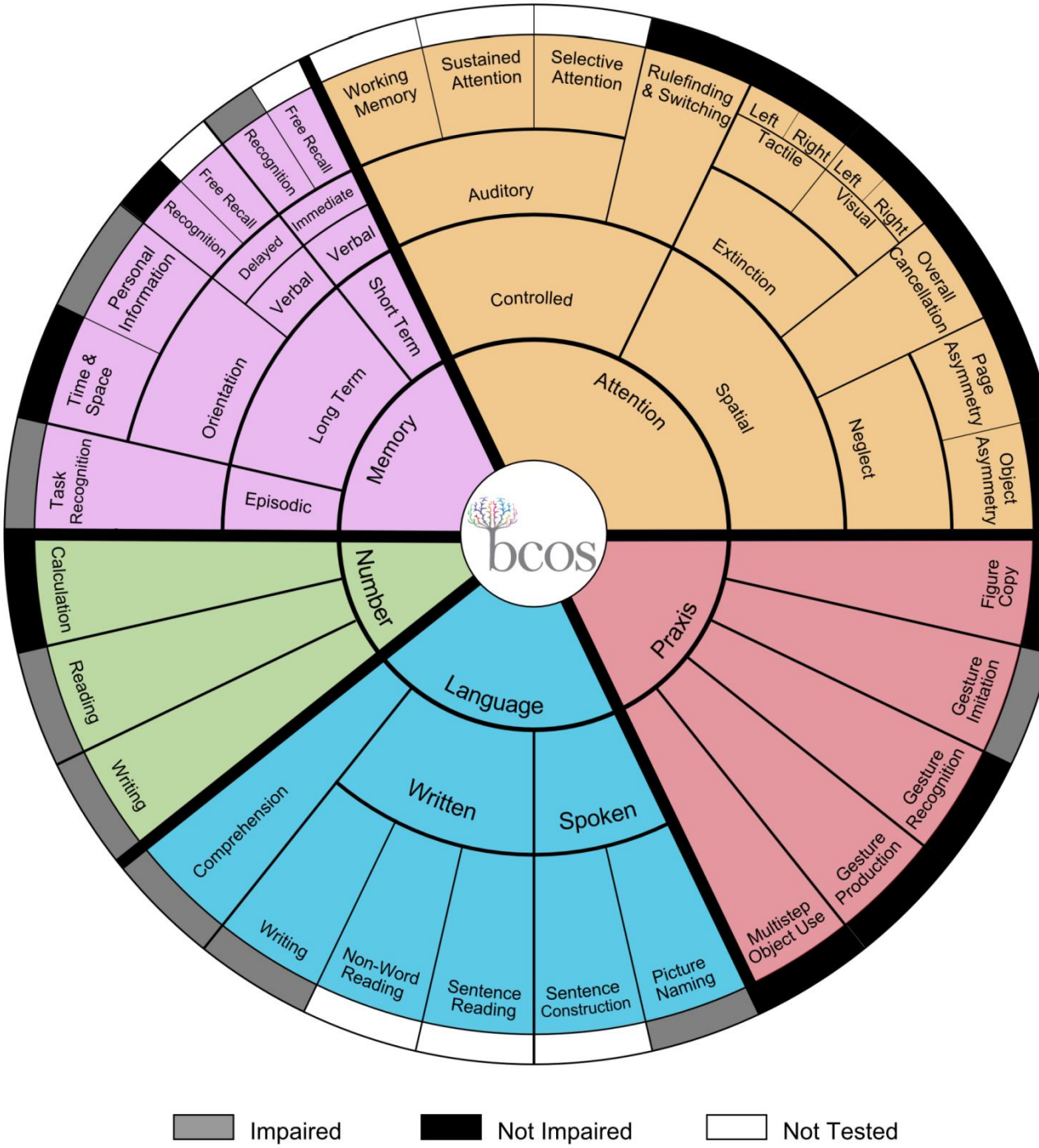
8.1. Study 4: Patient information (BCoS charts, lesion imaging)

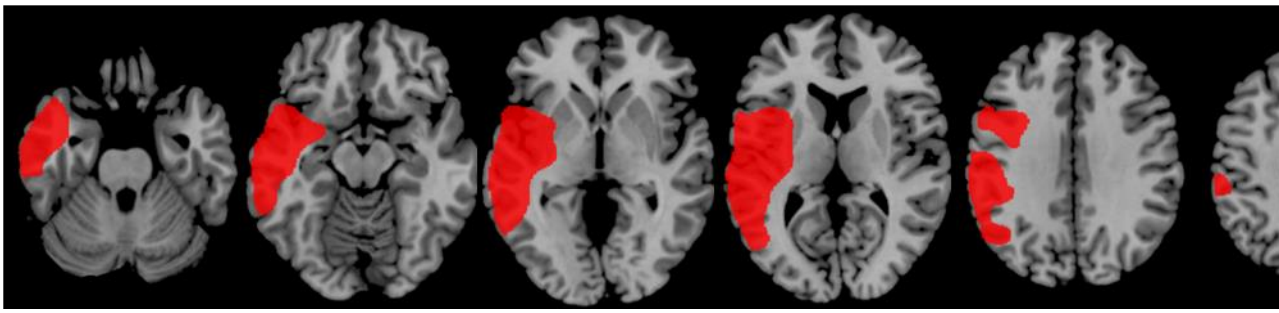
In this section, the following two pieces of visual information is given for each patient (when available):

1) The BCoS chart that summarises the patient's impairment profile, using as an input the most important components of the BCoS battery (Humphreys et al., 2012).

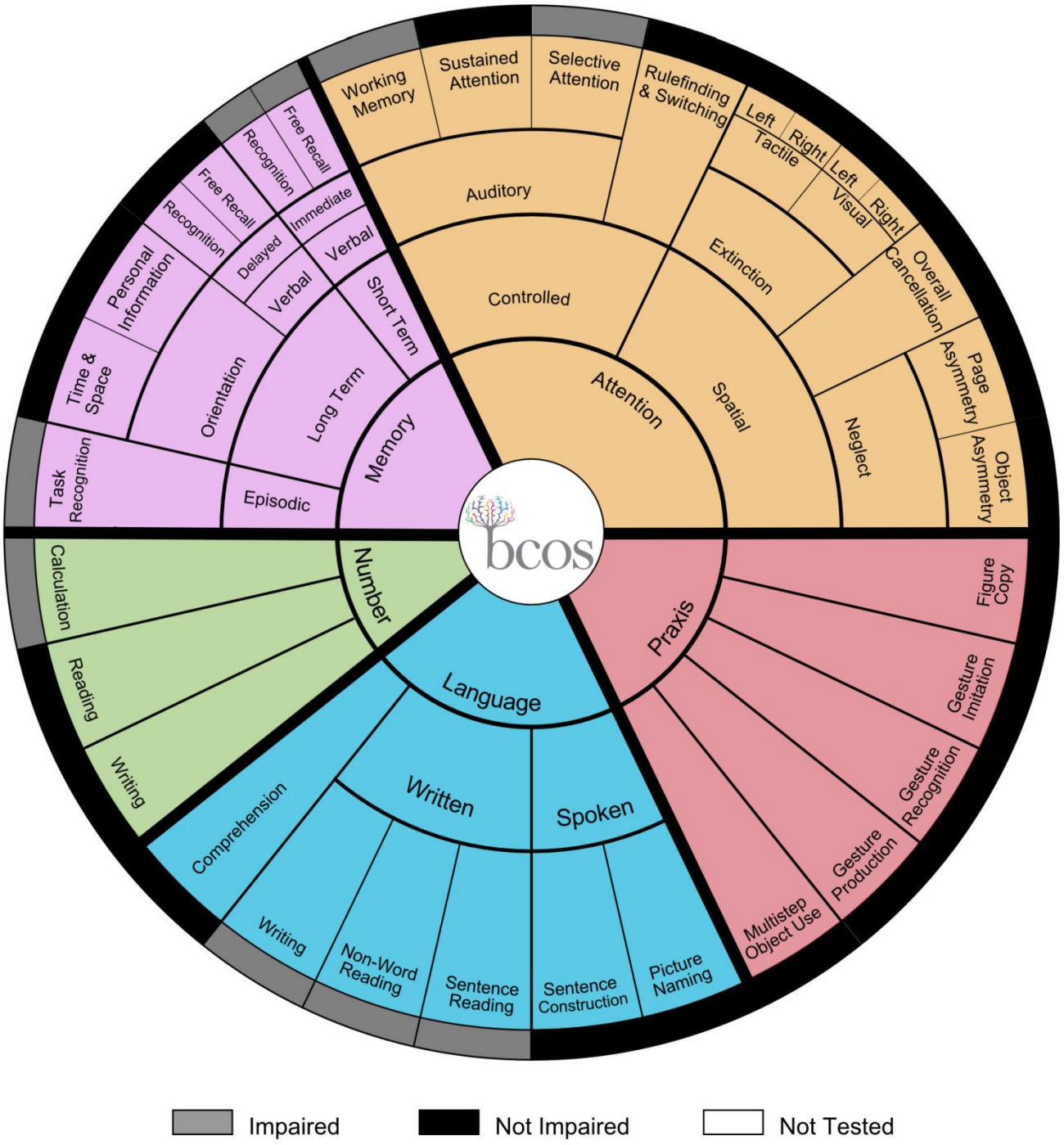
2) MRI imaging of the patient's lesion. The images are presented in MNI space, following neurological convention, i.e. the left-hand side of the image corresponds to the subject's left hemisphere. The lesions of patients P01, P02 and P04 have been manually delineated on a high quality FLAIR image and normalization to MNI space was done using a T1-weighted image. MR data were acquired at the FMRIB Centre (University of Oxford) when patients were in the chronic phase of their strokes. The lesion of P03 has been manually delineated on a CT scan, acquired in the hospital during the acute phase. No imaging is available for patients P05 and P06. The lesion of P06 involved bilateral parietal cortex and did not arise from stroke but from carbon monoxide poisoning.

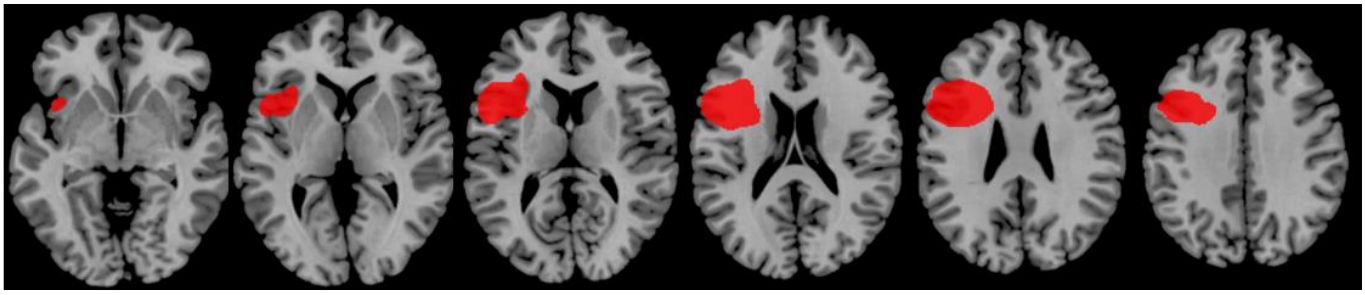
8.1.1 P01


Working Memory
Sustained Attention
Selective Attention
Rulefinding & Switching
Left
Right
Tactile
Left
Right
Visual
Auditory
Controlled
Extinction
Overall Cancellation
Page Asymmetry
Object Asymmetry
Neglect
Attention
Spatial
Free Recall
Recognition
Free Recall
Recognition
Immediate
Delayed
Verbal
Verbal
Short Term
Personal Information
Time & Space
Orientation
Long Term
Memory
Task Recognition
Episodic
bcos
Calculation
Reading
Writing
Number
Language
Written
Spoken
Comprehension
Writing
Non-Word Reading
Sentence Reading
Sentence Construction
Picture Naming
Praxis
Figure Copy
Gesture Imitation
Gesture Recognition
Gesture Production
Multistep Object Use
Impaired
Not Impaired
Not Tested

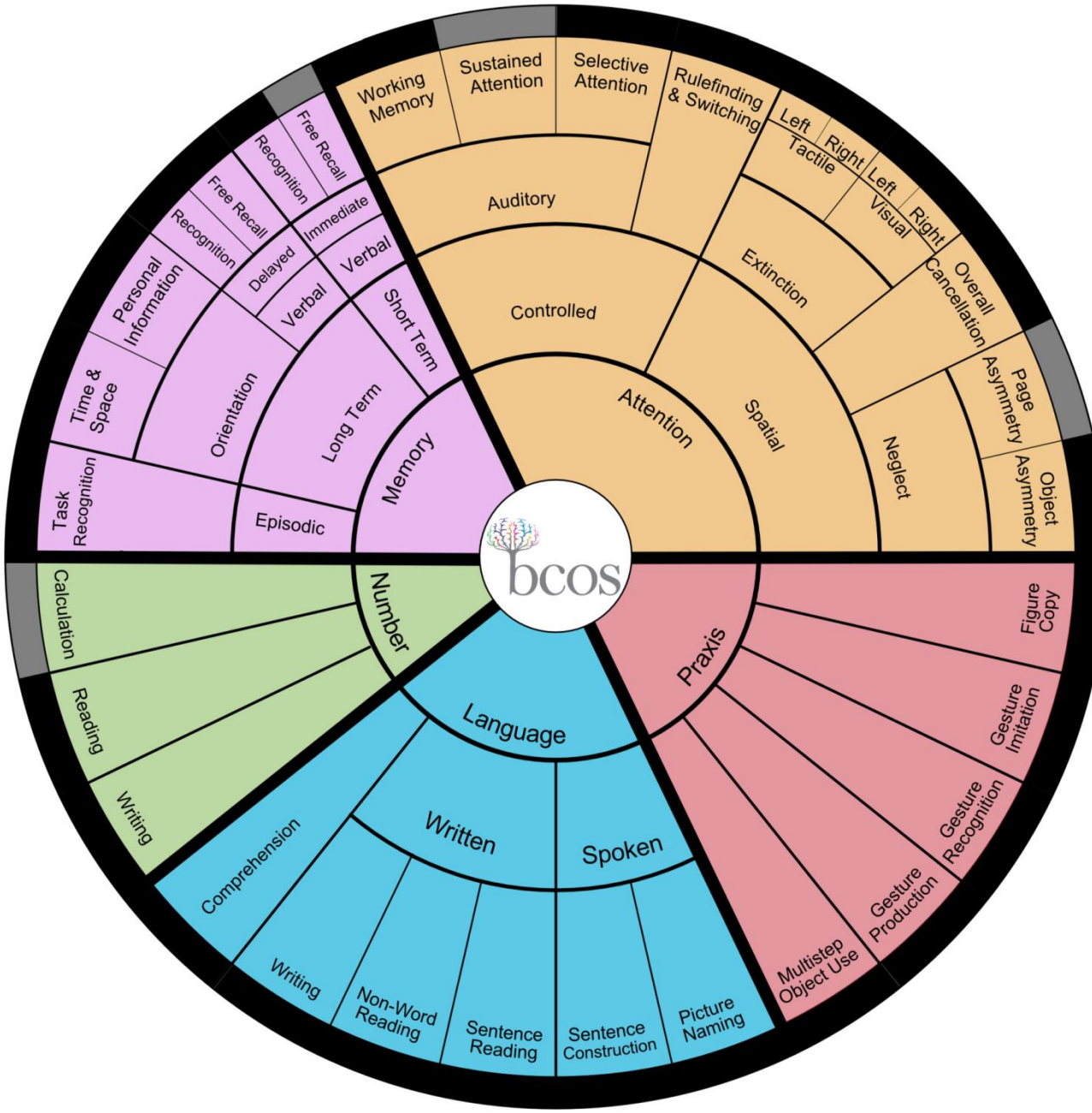


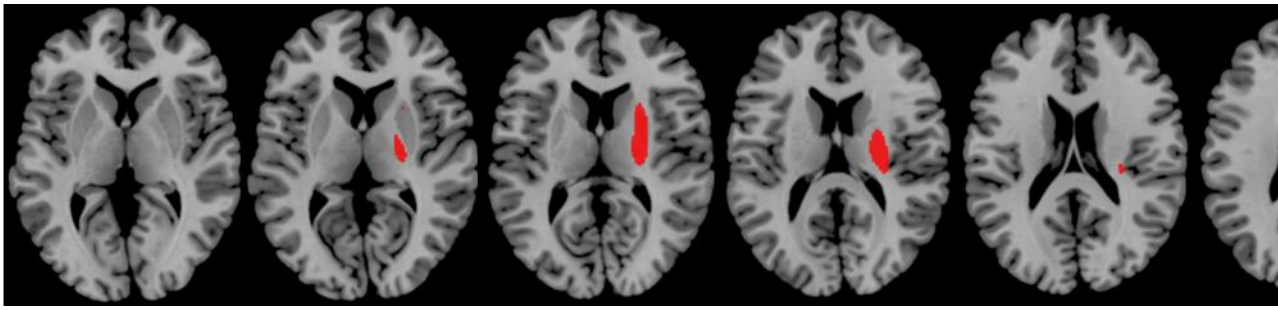
8.1.2 P02


Working Memory
Sustained Attention
Selective Attention
Rulefinding & Switching
Left
Right
Tactile
Left
Right
Visual
Auditory
Extinction
Overall Cancellation
Page Asymmetry
Object Asymmetry
Neglect
Spatial
Controlled
Attention
Free Recall
Recognition
Free Recall
Recognition
Immediate
Delayed
Verbal
Verbal
Short Term
Personal Information
Time & Space
Orientation
Long Term
Memory
Task Recognition
Episodic
bcos
Calculation
Reading
Writing
Number
Language
Comprehension
Written
Spoken
Writing
Non-Word Reading
Sentence Reading
Sentence Construction
Picture Naming
Praxis
Figure Copy
Gesture Imitation
Gesture Recognition
Gesture Production
Multistep Object Use
Impaired
Not Impaired
Not Tested

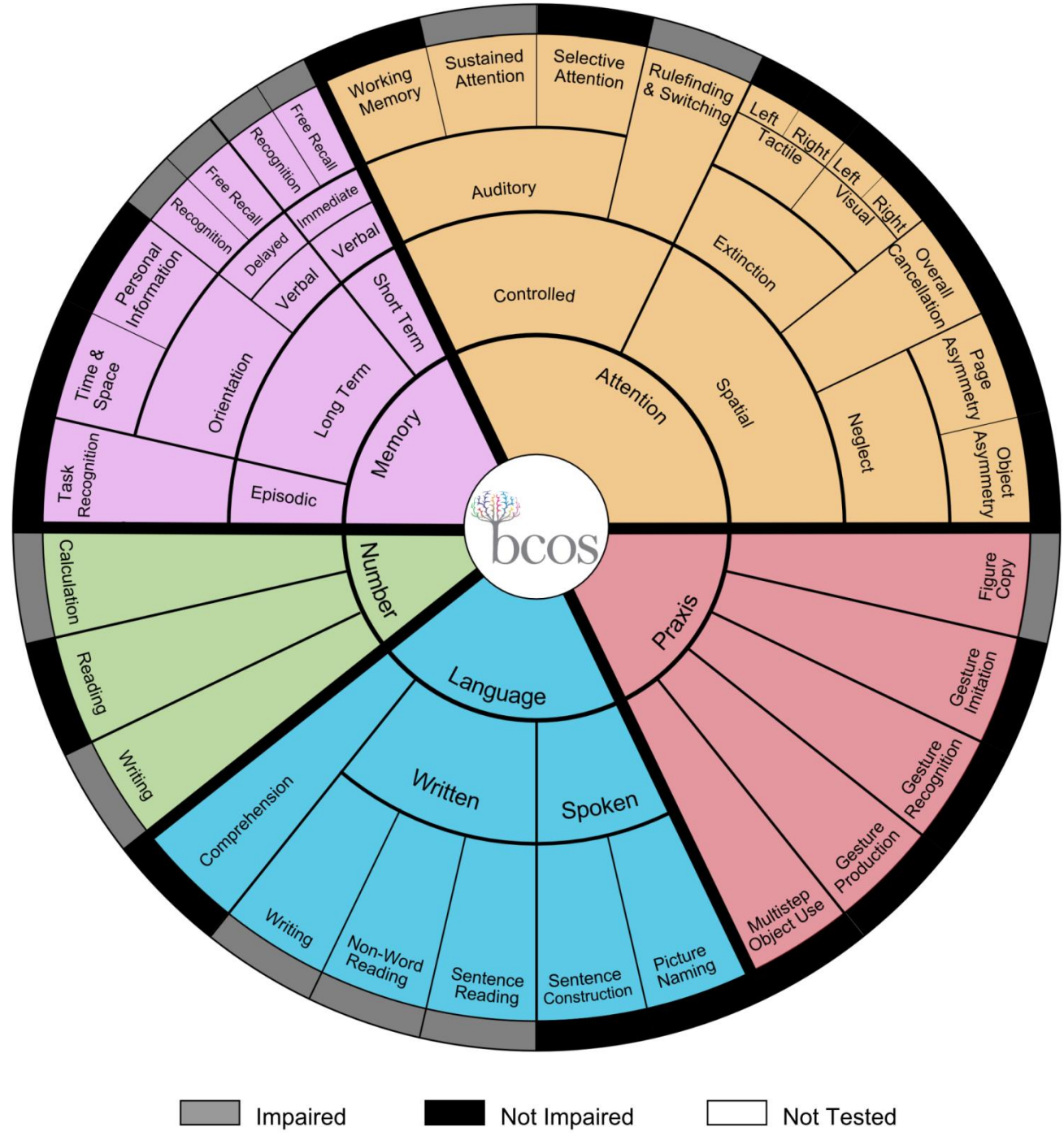


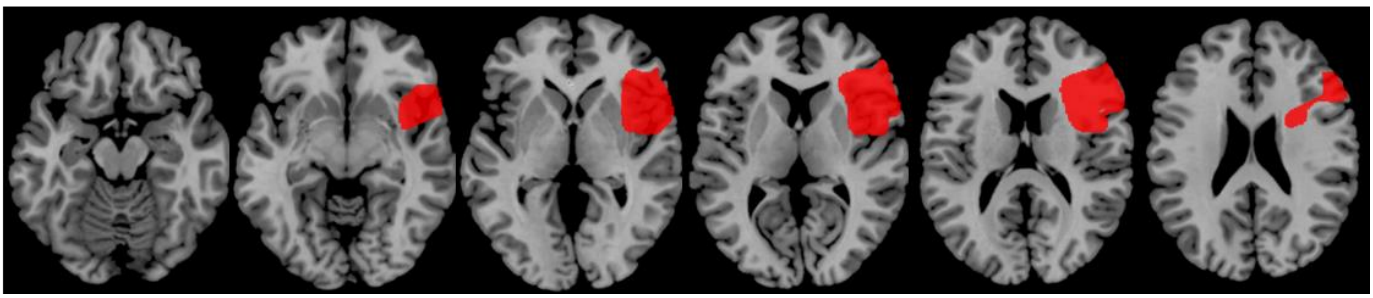
8.1.3 P03


Working Memory
Sustained Attention
Selective Attention
Rulefinding & Switching
Left
Right
Tactile
Left
Right
Visual
Auditory
Extinction
Overall Cancellation
Controlled
Attention
Spatial
Neglect
Page Asymmetry
Object Asymmetry
Free Recall
Recognition
Free Recall
Recognition
Immediate
Delayed
Verbal
Verbal
Short Term
Personal Information
Time & Space
Orientation
Long Term
Memory
Task Recognition
Episodic
bcos
Calculation
Reading
Writing
Number
Language
Written
Spoken
Comprehension
Writing
Non-Word Reading
Sentence Reading
Sentence Construction
Picture Naming
Praxis
Figure Copy
Gesture Imitation
Gesture Recognition
Gesture Production
Multistep Object Use
Impaired
Not Impaired
Not Tested

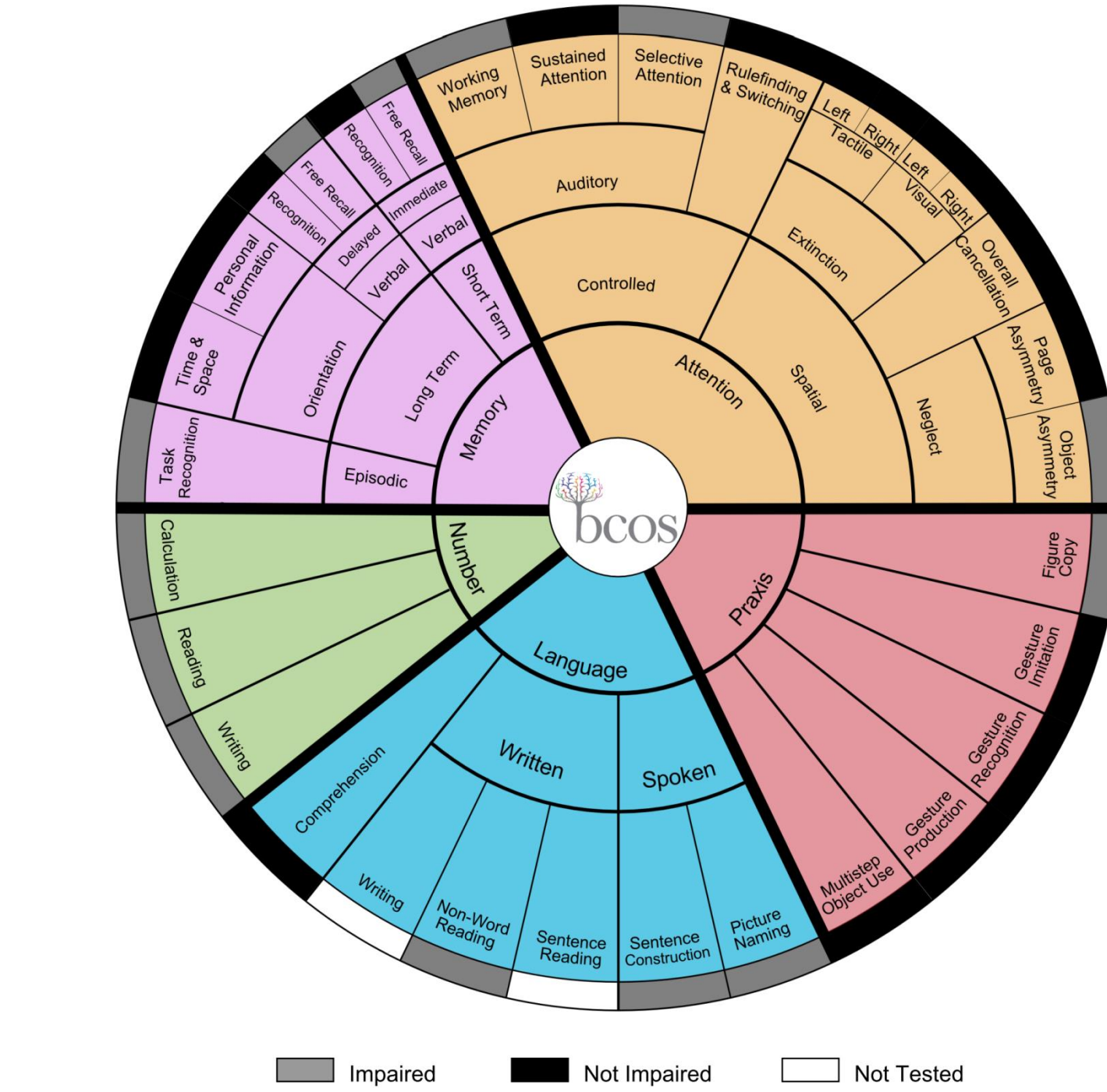


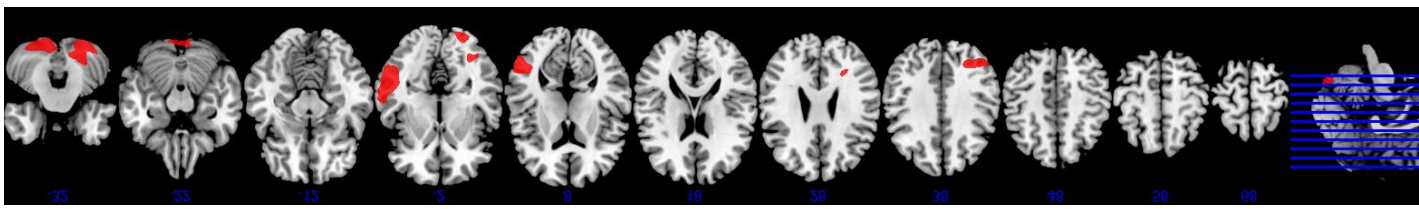
8.1.4 P04


Working Memory
Sustained Attention
Selective Attention
Rulefinding & Switching
Left
Right
Tactile
Left
Right
Visual
Overall Cancellation
Page Asymmetry
Object Asymmetry
Free Recall
Recognition
Free Recall
Recognition
Immediate
Delayed
Verbal
Verbal
Personal Information
Time & Space
Task Recognition
Short Term
Long Term
Orientation
Episodic
Memory
Auditory
Controlled
Extinction
Spatial
Neglect
Attention
bcos
Calculation
Reading
Writing
Number
Language
Praxis
Figure Copy
Gesture Imitation
Gesture Recognition
Gesture Production
Multistep Object Use
Comprehension
Writing
Written
Non-Word Reading
Sentence Reading
Spoken
Sentence Construction
Picture Naming
Impaired
Not Impaired
Not Tested



8.1.5 P07


Working Memory
Sustained Attention
Selective Attention
Rulefinding & Switching
Left Tactile
Right Tactile
Left Visual
Right Visual
Auditory
Controlled
Extinction
Overall Cancellation
Page Asymmetry
Object Asymmetry
Neglect
Attention
Spatial
Free Recall
Recognition
Free Recall
Recognition
Immediate
Delayed
Verbal
Verbal
Short Term
Personal Information
Time & Space
Orientation
Long Term
Memory
Task Recognition
Episodic
bcos
Calculation
Reading
Writing
Number
Language
Written
Spoken
Comprehension
Writing
Non-Word Reading
Sentence Reading
Sentence Construction
Picture Naming
Praxis
Figure Copy
Gesture Imitation
Gesture Recognition
Gesture Production
Multistep Object Use
Impaired
Not Impaired
Not Tested

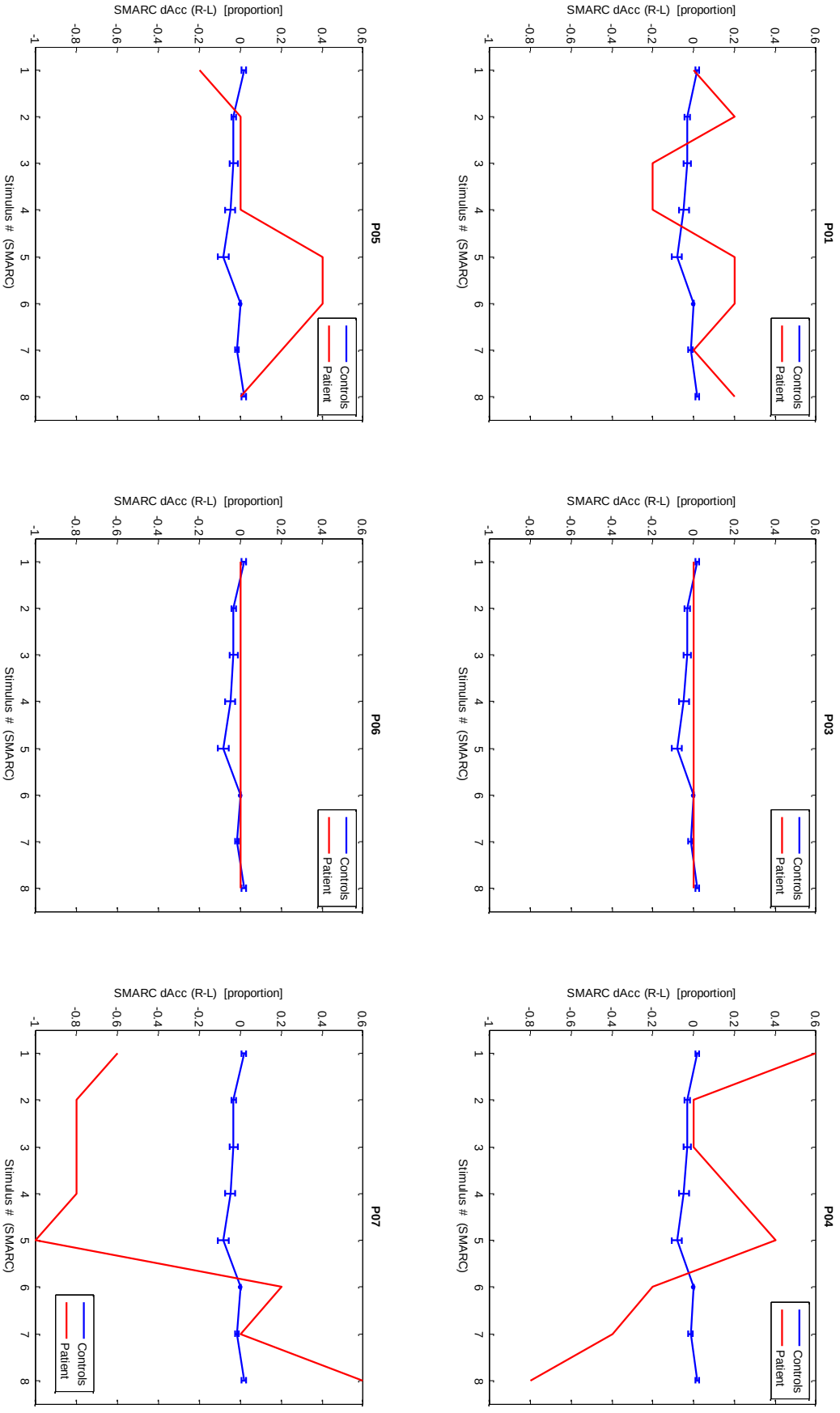


8.2. Study 4: Individual patient plots

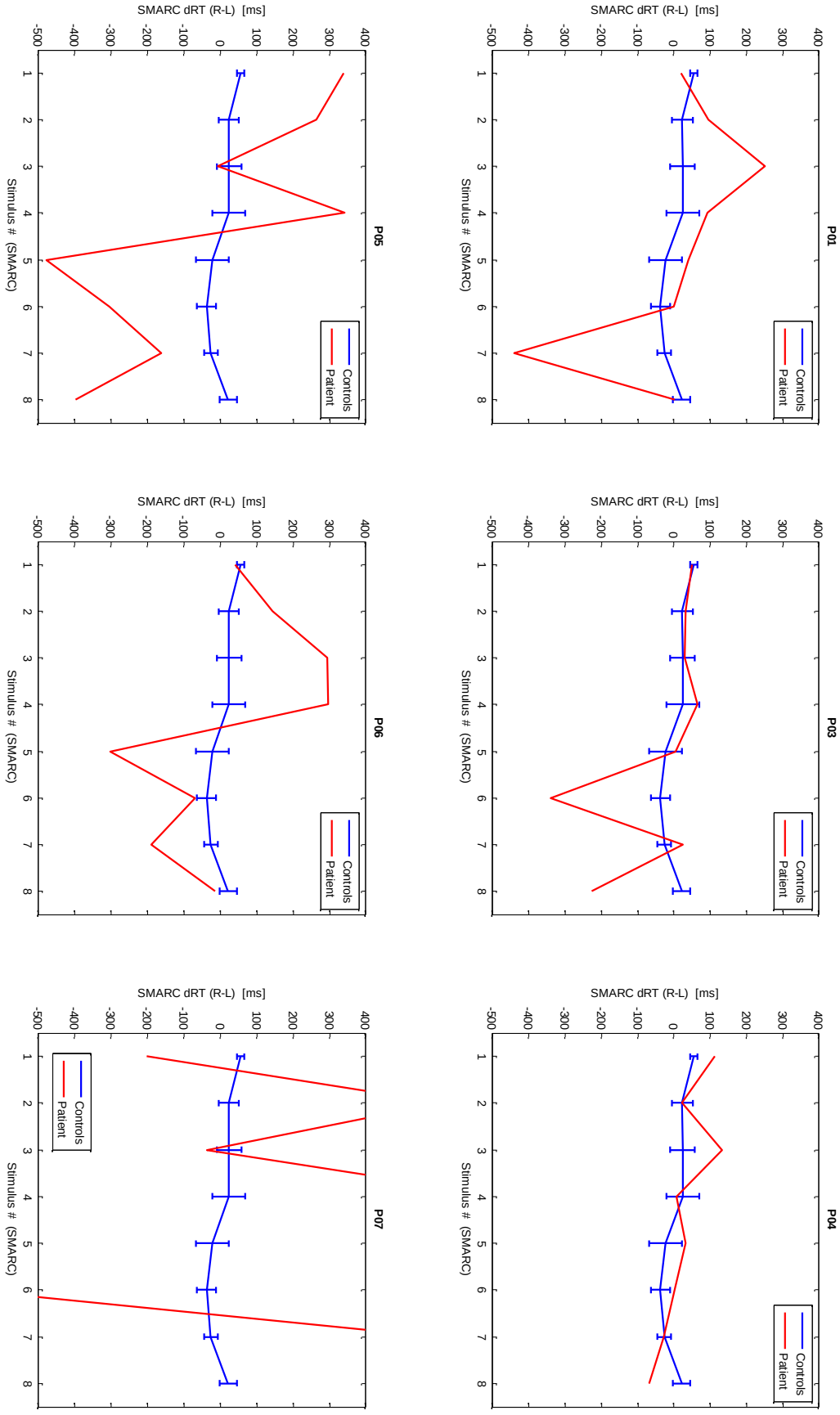
This section contains plots of dRT and dAcc against stimulus number, i.e., against number magnitude (0 to 9) for SMARC, or against sound probes sorted categorically by their pitch for SMARC (pitch #1 to pitch #8). Each plot represents each individual patient's data against the mean data from the controls. In the dRT graphs, missing datapoints indicate that at least one of the components of dRT (average for left, average for right) was blank, i.e. the subject had no correct trials for that stimulus and that response hand. In the dAcc graphs, a flat line indicates a subject whose overall accuracy was at ceiling, which meant that there was never any difference between mean accuracy for trials answered with left hand vs right hand, i.e. the dAcc was always zero.

In the SNARC dAcc figure, it can be seen that the plots of P05 and P07 both show a slopeless zig-zag pattern. This suggests an interaction with the stimulus parity factor rather than with magnitude per se. For P05, all dAcc values are <0, i.e. a left-hand advantage, while for P07 they are >0, i.e. a right-hand advantage. For P05, the left hand advantage (or: right hand *dis*advantage) is amplified, i.e. dAcc is even more negative, for even numbers. For P07, the RH advantage is amplified, i.e. dAcc is even more positive, for odd numbers. These patterns are not seen in the SMARC dAcc plots, as there we have no periodicity in stimulus parity, but simply linearly increasing pitch height.

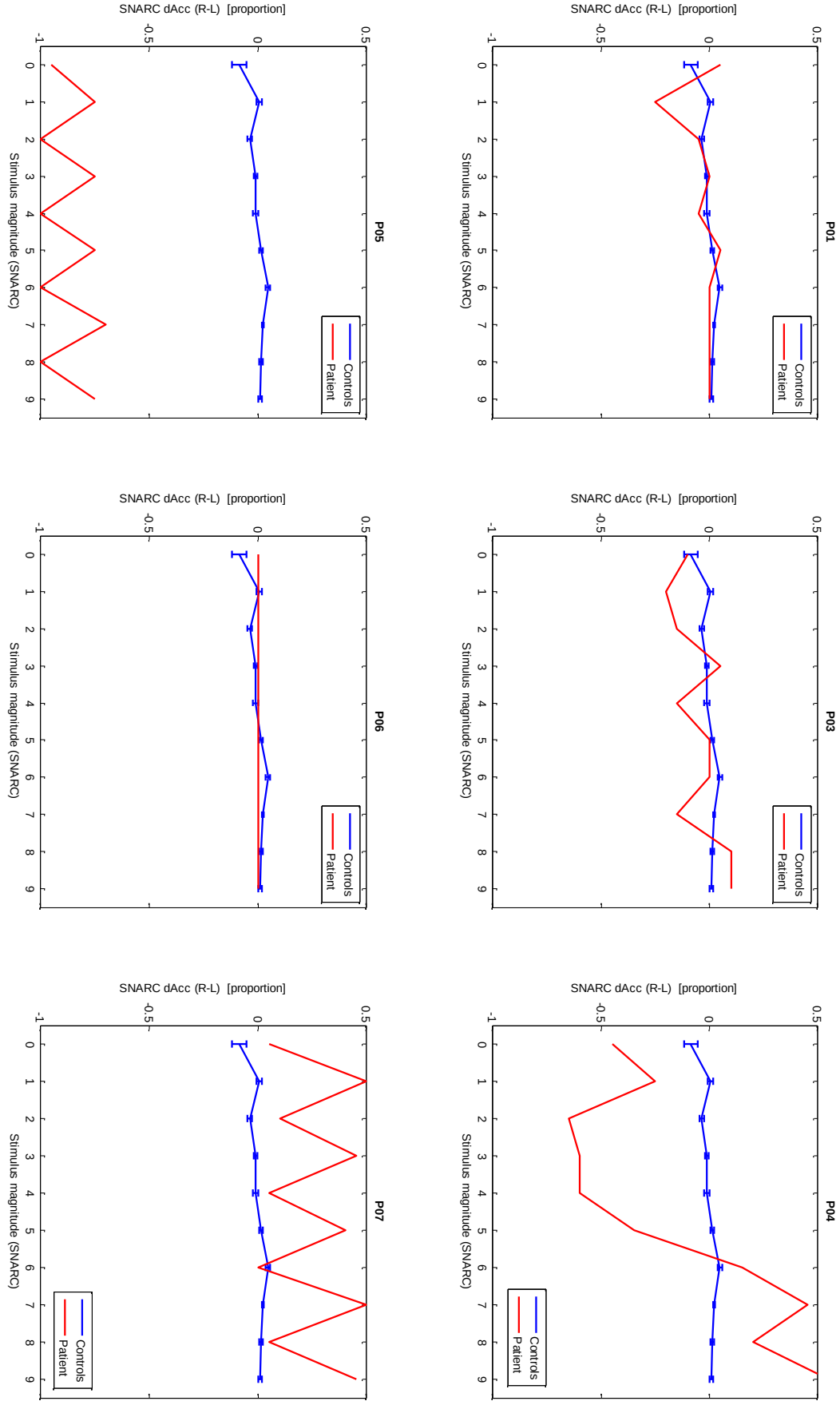
8.2.1 SMARC (dAcc)



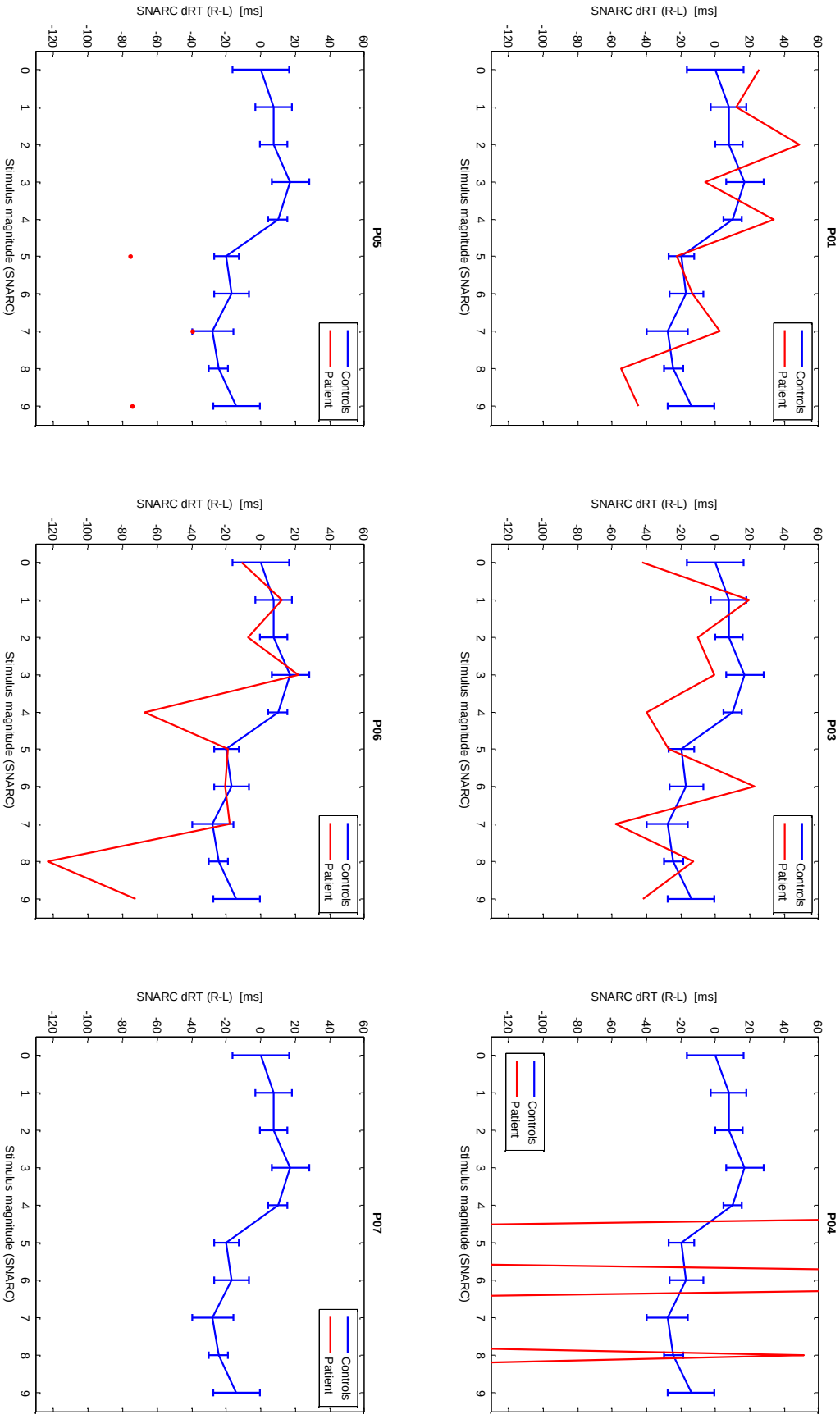
8.2.2 SMARC (dRT)



8.2.3 SNARC (dAcc)



8.2.4 SNARC (dRT)



9 List of abbreviations

Abbreviation	Meaning
Acc	accuracy
BCoS	Birmingham Cognitive Screen
CT	cortical thickness
DFA	discriminant function analysis
EEG	electroencephalography
FA	fractional anisotropy
FFG	fusiform gyrus
fMRi	functional magnetic resonance imaging
GM	grey matter
GMD	grey matter density
IFG	inferior frontal gyrus
IPS	intraparietal sulcus
lGI	local gyrification index
MD	mean diffusivity
MRI	magnetic resonance imaging
NIRS	near-infrared spectroscopy
ROI	region of interest
RT	reaction time
SBM	surface-based morphometry
SNARC	spatial-numerical association of response codes
SMARC	spatial-musical association of response codes
SFG	superior frontal gyrus
SPL	superior parietal lobule
TBSS	tract-based spatial statistics
tDCS	transcranial direct current stimulation
tES	transcranial electrical stimulation
TMS	transcranial magnetic stimulation
tRNS	transcranial random noise stimulation
VBM	voxel-based morphometry
WM	white matter