# The cognitive and neural bases of science and maths reasoning in adolescence

# The cognitive and neural bases of science and maths reasoning in adolescence

**Annie Brookman-Byrne**

Department of Psychological Sciences

Birkbeck, University of London

Thesis submitted for the degree of Doctor of Philosophy

I confirm that the work presented in this thesis is my own.

Annie Brookman-Byrne

# Acknowledgements

I would first like to thank my supervisors for their continued support: Iroise Dumontheil, for training me as an MRI operator, supporting my passion for public engagement, and always providing careful feedback; and my co-supervisors Denis Mareschal and Andy Tolmie, for regular discussions and their thoughtful contributions and enthusiasm for the work.

My thanks are also due to those who helped with data collection: Jessica Massonnié, Jack White-Foy, Minh Nguyen, Adrian Woodley-Cooper, and Sara Kapika. Thank you all for your hard work.

Thank you to all of the teachers who helped with this research, and the students, who were such willing and interested participants. In particular, thank you to Latymer Upper School for partnering with me on this project: Jackie Heywood, Zen Rogers, Ian Thompson, Zack Bassman, Caitlin Homes, and Tess Andrew; thank you for being so welcoming and easy to work with. Thank you also to Jeremy Dudman Jones from Greenford High School, and to Jack White-Foy from Dulwich College, for your ongoing enthusiasm and help.

I would also like to thank past and present members of both the iDCN lab and the CEN PhD group, for their encouragement and the invaluable opportunities to discuss the ins and outs of our research.

Thank you to Lia Commissar from the Wellcome Trust for working with me on various educational neuroscience projects over the years, helping me to work further towards our shared goal of connecting researchers and teachers, to bring scientific evidence to the classroom.

Finally, thank you to my weird and wonderful family for providing much-needed escapes from PhD life in the form of holidays, group chats, and gatherings (and Prosecco). A huge thank you to my amazing partner Max, for believing in me and being the best PhD buddy in the world. Sitting together at our shared desk every day in the final stretch gave me the motivation to get this finished. Thank you for all the coffee, memes, and your constant love and support.

# Abstract

This thesis took an educational neuroscience approach to investigate the cognitive and neural bases of science and maths reasoning in adolescence. The studies investigated the cognitive skills required to reason effectively in science and maths, and in particular about counterintuitive concepts, where misconceptions may be held. Misconceptions remain throughout schooling, likely interfering with academic success, and often persisting into adulthood. The specific roles of inhibitory control and relational reasoning were examined. Inhibitory control, the ability to suppress a prepotent response, is thought to enable the inhibition of intuitive concepts, while relational reasoning, the ability to detect patterns, is thought to allow the extension of conceptual understanding to different domains. All studies focussed on adolescence, when these skills are still developing, and when science and maths reasoning are essential for compulsory school exams. The first behavioural study showed that both response and semantic inhibition predicted variance in counterintuitive reasoning specifically, when controlling for general cognitive ability. Two classroom studies that were designed with teachers did not find that inhibitory control associated with misconception presence, before or after a lesson on a specific counterintuitive concept. The first analysis of brain data from a functional magnetic resonance imaging study showed that brain activations associated with both response and semantic inhibition overlapped with those recruited when adolescents reasoned about science and maths misconceptions. The second analysis of these brain data indicated that verbal analogical reasoning predicted unique variance in science performance and neural activation in maths, while non-verbal relational reasoning was associated with neural activation in science. Finally, the second behavioural study showed verbal analogical and non-verbal relational reasoning to relate to general science and maths performance but also specifically to counterintuitive reasoning. Overall, the results indicate that inhibitory control and relational reasoning are two skills associated with success in school-related science and maths.

# Table of contents

The work presented in this thesis was supported by an ESRC studentship, grant number ES/ J500021/1.

The work presented in Chapter 3 was published in the following paper:
Brookman-Byrne, A., Mareschal, D., Tolmie, A. K., & Dumontheil, I. (2018). Inhibitory control and counterintuitive science and maths reasoning in adolescence. *PLoS ONE, 13*(6), e0198973.

# Abbreviations

| | |
|---|---|
| AAL | Automatic Anatomical Labeling |
| ACC | anterior cingulate cortex |
| ADHD | attention deficit hyperactivity disorder |
| AG | angular gyrus |
| AI/FO | anterior insula and frontal operculum |
| ANOVA | analysis of variance |
| ANS | approximate number sense |
| BA | Brodmann area |
| BOLD | blood oxygenation level-dependent |
| CCF | Cattell Culture Fair |
| CMRR | Centre for Magnetic Resonance Research |
| DLPFC | dorsolateral prefrontal cortex |
| EEG | electroencephalography |
| EPI | echo-planar imaging |
| FD | framewise displacement |
| fMRI | functional magnetic resonance imaging |
| FWE | family-wise error |
| GLM | general linear model |
| GRAPPA | generalised autocalibrating partial parallel acquisition |
| HRF | haemodynamic response function |
| IFG | inferior frontal gyrus |
| IFS | inferior frontal sulcus |
| IPL | inferior parietal lobule |
| IPS | intraparietal sulcus |
| ISI | interstimulus interval |
| L | left hemisphere |
| LNAT | Law National Aptitude Test |
| LSAT | Law School Admission Test |
| M | mean |
| MD | multiple-demand |
| MNI | Montreal Neurological Institute |
| MPRAGE | magnetization-prepared rapid gradient-echo |
| MRI | magnetic resonance imaging |

| | |
|---|---|
| n.s. | not significant |
| PFC | prefrontal cortex |
| PTRT | primary task reaction time |
| R | right hemisphere |
| Re-Cat | Re-Categorisation task |
| RF | radiofrequency |
| RT | reaction time |
| SAT | Scholastic Assessment Test |
| SD | standard deviation |
| SE | standard error |
| SLA | Science Learning Assessment |
| SMA | supplementary motor area |
| SPL | superior parietal lobule |
| SPM | statistical parametric mapping |
| SP-Ver | Sentence-Picture Verification task |
| SSRT | stop signal reaction time |
| TARC | Test of Analogical Reasoning in Children |
| TE | echo time |
| TORR | Test of Relational Reasoning |
| TR | repetition time |
| UK | United Kingdom |
| USA | United States of America |
| VLPFC | ventrolateral prefrontal cortex |
| VS-WM | visuospatial working memory |
| VWM | verbal working memory |
| WASI | Wechsler Abbreviated Scale of Intelligence |

# List of figures

# List of tables

# Chapter 1 Introduction

Science and maths reasoning are important skills that children and adolescents must develop in order to understand the world and make informed decisions. As such, they are core components of the school curriculum (Department for Education, 2013a, 2013b), and compulsory exams are taken in these disciplines during adolescence. The material that is taught during this period of life is increasingly complex, and sometimes counterintuitive. Science and maths are also considered particularly difficult disciplines to learn (Lortie-Forgues, Tian, & Siegler, 2015; Zaitchik, Iqbal, & Carey, 2014). While the development of science and maths reasoning skills are typically studied separately, the mechanisms driving success appear to be similar. This may be especially true in the context of counterintuitive reasoning (Mareschal, 2016), which may require a specific set of cognitive skills to overcome misconceptions. Two candidate abilities which may enable successful science and maths reasoning, are inhibitory control and relational reasoning. These are domain-general skills that continue to develop through adolescence (Jablansky, Alexander, Dumas, & Compton, 2015; Leon-Carrion, García-Orza, & Pérez-Santamaría, 2004), and yet their role in reasoning about science and maths concepts during adolescence is not well understood. Establishing the cognitive and neural associations between these skills may lead to recommendations for education.

The aim of this PhD was to take an educational neuroscience approach to explore the cognitive and neural bases of science and maths reasoning in adolescence. In this introductory chapter, I first define educational neuroscience and consider how this scientific approach might inform teaching and learning. Second, I review the literature relating to the development of science and maths reasoning skills, starting with general science and maths and then considering misconceptions in these disciplines. Third, I review the measurement and development of inhibitory control, with an initial focus on this ability in isolation, and then in relation to counterintuitive science and maths reasoning. Fourth, I examine the development of relational reasoning skills, making links to executive functions and science and maths. Fifth, I outline the educational neuroscience approach taken throughout the PhD, and sixth, I summarise the aims of each following chapter.

## 1.1 Educational neuroscience

Educational neuroscience is a relatively young field that brings together psychology, neuroscience, and education (and any other discipline related to learning)

(**Figure 1.1**), with the aim of understanding more about learning, and with the ultimate goal of bringing scientific insights to the classroom to improve teaching and learning (Butterworth & Tolmie, 2014). As opposed to a new discipline, educational neuroscience can be considered a new approach to research and teaching. Researchers work with teachers to design studies of relevance for education in addition to answering important scientific questions. Newly acquired scientific knowledge is then fed back to teachers to inform their practices in the classroom.

| | | |
|---|---|---|
| Psychology | → | developing cognitive and social skills |
| Cognitive neuroscience | → | processing and storage of information |
| Neuroscience | → | biological basis of brain function |
| Behavioural genetics | → | genetic and environmental influences |
| Computer science | → | models of learning |
| Education | → | learning environment and school outcomes |
| Technology | → | tailored learning programmes |

**Figure 1.1.** Different disciplines that contribute to educational neuroscience, and how they can inform our understanding of processes involved in learning.

The emergence of educational neuroscience is linked to the rise of sophisticated technologies to image the human brain (The Royal Society, 2011). These technologies, such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), are ever-advancing, and revealing more and more about the underlying cognitive and neural mechanisms of learning and reasoning through development. The challenge lies in translating this research to the classroom (Simmonds, 2014). There are many steps between initial lab studies and educational recommendations for teachers, learners, and schools. The process is likely to be lengthy, with many different studies leading towards one or two particular recommendations for teachers (**Figure 1.2**). It may also be that new findings simply confirm what teachers already thought: that a certain technique is effective in the classroom. Some research aims to investigate *why* the practices that are known to be effective work. Even when translation of something new is ready, a language barrier between teachers and researchers remains, whereby teachers and researchers may not necessarily have a common lexicon, sometimes using the same words to mean different things. Scientists may not have the necessary knowledge of the classroom to share their findings in a manner that is useful for teachers. Similarly, teachers may lack the resources, in terms of time and sufficient background knowledge, to assess and implement new findings adequately (Simmonds, 2014). Nonetheless, it is essential that teachers are part of the translation process, since they will be best placed to fit the research into the realities of school life.

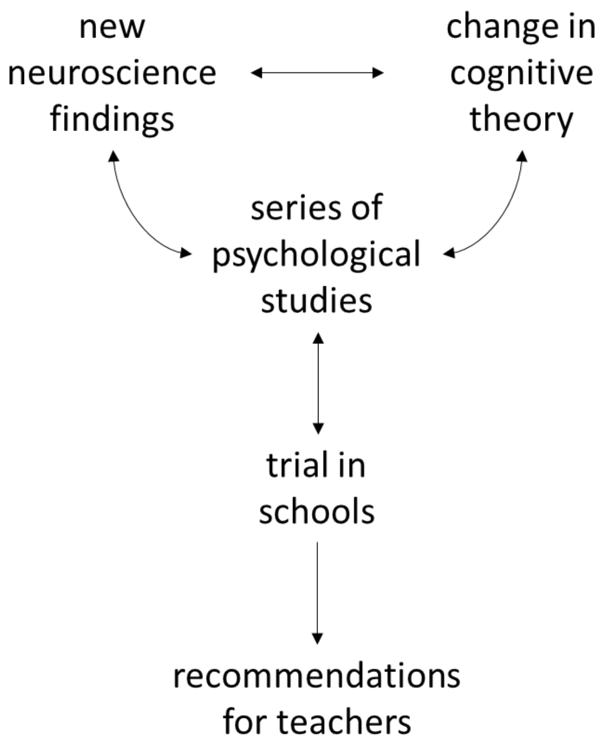



**Figure 1.2**. Example of possible steps taken between new neuroscience findings and recommendations for teachers. Note the bidirectional arrows indicating many steps and a non-linear process, with no particular starting point.

Communication and collaboration are key tenets of educational neuroscience, and conversations between researchers and teachers must be two-way, rather than scientists simply imparting their knowledge (Ansari & Coch, 2006). Researchers can also gain from listening to what teachers have to say. Discussions between educators and researchers will lead to scientific research that addresses questions of importance and interest to teachers. Where neuroscience and psychology lead to specific predictions about what might work in education, teachers will be able to provide important perspectives on the feasibility of these ideas. Being prescriptive will interfere with a teacher's professional autonomy, while providing new tools and information drawn from scientific research will enable teachers to choose the most appropriate method for a given scenario in their own classroom, as in the medical profession (Goldacre, 2013). Ideally, the involvement of teachers early on in the research process will mean that each programme of research can be shaped by the needs of educators.

There remains debate concerning whether or not educational neuroscience is a worthwhile endeavour. Criticisms of the field tend to draw on either scientific concerns (about methods, data, and theory) or pragmatic concerns (about costs, timing, and payoffs) (Varma, McCandliss, & Schwartz, 2008). For example, Bruer (1997) famously argued that the link between neuroscience and education is a bridge too far, and that cognitive psychology is instead where to look for educational guidance. While this argument is still repeated (Bowers, 2016), it does not take account of the current landscape of educational neuroscience research, which does indeed consider cognitive psychology, in addition to any other related disciplines. A common pragmatic concern is the high cost associated with some educational neuroscience research, with Bishop (2014) arguing that the payoff for such expensive neuroimaging research is not great enough. Bishop's position is that discovering neural underpinnings of learning processes cannot be of use to teachers. While there may be no direct link from neuroimaging findings to education, they can inform psychological theories and as such tell us more about learning than behavioural data on its own (Howard-Jones et al., 2016). Educational neuroscience is a long-term endeavour (Howard-Jones et al., 2016), and as such, no quick fixes are expected from carrying out neuroimaging research. While pragmatic concerns are perhaps easier to sympathise with, many of the criticisms of the field do not reflect the reality of educational neuroscience which is already making strides in linking neuroscience and education, through the convergence of many different disciplines.

The remainder of the Introduction of this thesis takes an educational neuroscience approach in considering different levels of explanation and bringing together research from psychology, neuroscience, and education. By taking this approach to the literature review, the resulting research presented in this thesis was in a position to scientifically investigate reasoning and learning during adolescence in a holistic manner: drawing from a range of relevant literatures, using psychology and neuroscience methods, and with relevance to education a key consideration throughout.

## 1.2 Science and maths

Science and maths are core components of the curriculum at both primary and secondary school and are essential if children and adolescents are to understand the world around them. Thinking scientifically and mathematically allows individuals to make informed judgments about important current issues (for instance, relating to health or the environment), and has a positive contribution to wealth in society (The Royal Society, 2014). The National Curriculum for England determines the content of taught science and maths at school, expecting pupils to improve in these core disciplines throughout the school years, solving increasingly challenging problems (e.g. Department for Education, 2013b, 2013a). The development of science and maths skills through childhood and adolescence has been studied extensively, but there remain unknowns in terms of how individual differences in science and maths relate to the development of other cognitive skills, and in terms of how different aspects of science and maths skills relate to each other.

### 1.2.1 Development of science and maths skills

The development of science and maths skills have typically been studied separately, and as such they are presented here in turn.

#### 1.2.1.1 Science

Within the discipline of science, the focus has been broad, with studies investigating the generation of hypotheses, the design of experiments, experimentation itself, the evaluation of evidence, and conceptual change. At present there is no consensus as to how best to define and categorise science skills, and consequently, different studies have taken different approaches in establishing science abilities across development.

The Science Learning Assessment (SLA) is a 24-item test that has been administered to young children to establish their understanding of the scientific inquiry process and of life sciences (Samarapungavan, Mantzicopoulos, Patrick, & French, 2009). One hundred 5-year-olds were tested on the SLA after 65 of the children had received science teaching in the SLA topics, while 35 of them received no science teaching. The group who had received science instruction performed better on items measuring both the scientific inquiry process and life sciences understanding, with a greater advantage on items relating to scientific inquiry (Samarapungavan et al., 2009). Overall the results indicate that children are able to reason effectively about science processes and scientific concepts from an early age, if they are given adequate support. They also indicate that skills relating to scientific inquiry are less likely to emerge without explicit teaching, perhaps because they are less likely to be spoken about casually at home and in school than scientific concepts.

A study of 1,581 8-, 9-, and 10-year-olds used 66 story problems to examine five components of science: goals of science, theories and interpretative frameworks, experimentation strategies, experimental designs, and data interpretation (Koerber, Mayer, Osterhaus, Schwippert, & Sodian, 2015). Performance in all five components improved with age, and analyses supported a model whereby science skills formed one single unitary construct, as opposed to being separable by the different components: there were no developmental stages according to components. Conversely, a study of 223 children aged 4 to 13 years found that three components of science reasoning followed different developmental patterns (Piekny & Maehler, 2013). The earliest component to emerge was the ability to evaluate evidence that either covaried perfectly or imperfectly, which improved steadily with age. Next to develop was an understanding of experimentation, which required children to engage in hypothesis testing. This component also saw a gradual improvement, but with a pronounced shift in performance from age 7 to 9. The final component was hypothesis generation in the face of accumulating evidence, which proved to be much harder overall but nonetheless showed improvement with age and a large shift between ages 5 and 7 (Piekny & Maehler, 2013). These findings showed that different components of science reasoning develop at different stages, in contrast to the findings from Koerber and colleagues (2015). This might be due to the larger age range, as within 8- to 10-year-olds it would be surprising to find hugely different developmental trends. It could also be down to the types of tasks administered: the study from Koerber and colleagues (2015) only used story problems, all of which might require similar types of reasoning skills. Working through 66 story problems likely relies on language ability, and the large number of problems might also require a high level of motivation. The study from Piekny and Maehler (2013) used one-to-one testing sessions that were more active in nature, requiring physical demonstration of different skills. This might have led to greater differentiation between problem types, since reading was not important and the experimenter likely helped to keep motivation high.

These studies reflect the wider literature on science skills that show differing associations with age across components of science that are tested and defined inconsistently. Tolmie, Ghazali, and Morris (2016) attempted to clarify this muddled picture and recently proposed a simplified new account of three core science skills: *prediction,* the anticipation of scientific events; *description*, the ability to reason about causal connections in observed events; and *explanation*, knowledge of the mechanisms that explain causal connections. In this analysis, explanation was linked specifically to concepts and given a central place in explicit understanding of science (Tolmie et al., 2016). The authors argued that these three skills together may form the foundation for effective science reasoning, but that detailed analysis is needed to establish the development of association between these skills. Zimmerman (2000) argued that science reasoning is a core academic skill that should be recognised alongside reading, writing, and arithmetic, as opposed to purely a content-based discipline, which is perhaps how it is often still considered despite the evidence that processes are important too.

#### 1.2.1.2 Maths

The picture is somewhat clearer in the field of maths, where there is a longstanding body of research into the development of key skills (Butterworth & Varma, 2013). Mathematical thinking is thought to rely on the integration of linguistic processes and visuospatial representations (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). Here I present a brief overview of the extensive literature into maths development, covering illustrative examples that demonstrate the developmental trajectory of core skills. There is evidence of an understanding of basic number even in infancy. An early study used a dishabituation paradigm to show that 72 22-week-olds were able to distinguish two dots from three dots, which the authors argued demonstrated the ability of infants to subitize, i.e. to recognise how many items are present without counting (Starkey & Cooper, 1980). Further research indicated that subitizing ability increases until approximately 5 years of age, where sets of five items can be subitized, and that this ability matches the capacity of adults (Starkey & Cooper, 1995). This number ability therefore emerges early in development and rapidly matures.

The ongoing development of arithmetic skills is thought to be supported by innate approximate representations of numbers (Feigenson, Dehaene, & Spelke, 2004), called an approximate number sense (ANS), which would explain how infants are able to discriminate small numbers. Feigenson and colleagues (2004) presented a large body of evidence to argue that the ANS is one of two core systems that underlie the ability to reason about numbers from infancy and throughout childhood to adulthood, with precision of the ANS improving through childhood and adolescence. The second core system they posit is the tracking and precise representation of small numbers of distinct items (Feigenson et al., 2004). According to this view, the same core systems are involved in number throughout life, forming the foundations for sophisticated maths. Spelke and Kinzler (2007) later argued that there is a single innate core number system, as evidenced by the infant data and cross-cultural data showing the same number representations independent of education. According to Spelke and Kinzler (2007), the core number system forms one part of ‘core knowledge’ – the systems that guide human cognition.

A longitudinal study emphasised the role of digit knowledge in arithmetic abilities, indicating that the role of the ANS is perhaps not as crucial as others have argued (Göbel, Watson, Lervåg, & Hulme, 2014). One hundred and sixty-five 6-year-olds were tested at baseline and ANS was not a unique predictor of arithmetic after 11 months, indicating that individual differences in the ANS are less of a constraint than previously imagined. Instead, knowledge of Arabic digits was associated with arithmetic growth. The authors argued that digit knowledge is therefore a key foundation of maths skills, just as letter knowledge is a key foundation of reading skills (Göbel et al., 2014). Nonetheless, there was a strong correlation between ANS and arithmetic, and it is possible that the ANS variance was shared with other control variables, such as earlier arithmetic skills.

Another longitudinal study, this time of 182 children, investigated the development of different maths skills in 4- to 7-year-olds (LeFevre et al., 2010). Modelling revealed three different pathways of skills that led to maths competence. Linguistic skills were specifically related to Arabic digit naming, supporting Göbel and colleagues' (2014) suggestion that Arabic digit knowledge in maths is analogous to letter knowledge in reading. Subitizing latency was related to nonlinguistic arithmetic ability. Finally, spatial attention was related to digit knowledge and numerical magnitude processing. Overall, the findings indicate that maturing maths ability relies on the development of multiple separable components (LeFevre et al., 2010). Cragg and Gilmore (2014) argued that this pathway approach is a good start but does not go far enough; that more work needs to be done to explore skills in factual, procedural, and conceptual knowledge in maths separately, as they likely require different demands that vary with age. Despite the relatively advanced literature on the emergence of maths skills, evidence relating to changing associations between components of maths and their underlying demands across development is still in its infancy, in part because standardised tests group these components together (Cragg & Gilmore, 2014). Factual maths (knowledge of number facts), procedural maths (applying a strategy) and conceptual maths (understanding mathematical rules) are dissociable (Cragg & Gilmore, 2014), and thus require in depth exploration.

Academic difficulties are more readily recognised in maths than science, with dyscalculia being diagnosed as a result of poor maths ability in the absence of a known cause such as low intelligence (Butterworth & Varma, 2013). Since the core skills in science are still under investigation there is no corresponding diagnosis in science, which is demonstrative of the difference in understanding between the two disciplines. While Tolmie and colleagues (2016) posited three core science skills, these have not been examined across development. Age-specific norms would be required in order to establish what might constitute a diagnosable scientific difficulty.

Despite the varied literature, it is clear that abilities in both science and maths begin to emerge from a young age (from infancy in maths), and continue to develop throughout the school years. In both disciplines there are suggested core skills which are fundamental to ongoing development, but with no clear categories established yet. In science, there is a suggested distinction between processes and concepts (Samarapungavan et al., 2009); a five component model (Koerber et al., 2015); and Tolmie and colleagues' (2016) three core skills (prediction, description, and explanation). Maths research has led to the proposal of a two skill model consisting of ANS and tracking/representation (Feigenson et al., 2004); a single core number system (Spelke & Kinzler, 2007); digit knowledge as a foundational skill (Göbel et al., 2014; LeFevre et al., 2010); and the distinction between facts, procedures, and concepts (Cragg & Gilmore, 2014). The suggested core skills across science and maths show little overlap, although there is some similarity in the proposal that procedures and concepts are both important features. There remains debate about the precise nature of these skills in both disciplines, and their importance as applied to school-related outcomes.

### 1.2.2 Neural correlates of science and maths

Examination of the underlying neural correlates of science and maths may reveal more about the mechanisms through which success is reached. Again, science and maths have been considered separately so will be examined in turn here. In science, the neuroimaging data is limited but focusses on fMRI. In maths, there is a larger body of literature, with fMRI the most commonly used neuroimaging method. This section will therefore focus on what fMRI data have uncovered about the brain regions involved in science and maths, which will also be particularly useful in situating the thesis (which uses fMRI as a method) within the existing literature.

#### 1.2.2.1 Science

Work investigating the neural correlates of science is as fragmented as the behavioural research. Neuroimaging research has been conducted on adults and has focussed on the domain of physics. Nine young adults aged 19 to 25, who were majoring in physics or engineering, were shown 30 physics terms relating to a range of topics while inside an fMRI scanner (Mason & Just, 2016). The participants were tasked with thinking about each concept and its properties when shown the term. Factor analysis of activation data identified four physics factors: causality (e.g. gravity, potential energy), periodicity (e.g. wavelength, diffraction), equation representation (e.g. velocity, acceleration), and energy flow (e.g. electric field, heat transfer). There was also a final non-physics factor associated with word length. Each factor was associated with distinct neural correlates, which were common across participants (Mason & Just, 2016). While the networks were distinct, they each covered frontal and parietal regions, with different factors located adjacently, and the majority of the activation was left-lateralised. Regions included parts of the inferior frontal gyrus (IFG), intraparietal sulcus (IPS), the postcentral sulcus, superior temporal sulcus, and the superior frontal gyrus. The authors argued that each factor included regions associated with executive functions, spatial processing, and linguistic processing, indicating that despite their specificity they reflect domain-general skills (Mason & Just, 2016). These results are compelling in showing distinct neural representations for different types of physics concepts, but the study is limited in only examining relative experts, and using a task that engages thinking skills as opposed to reasoning skills. The very small sample size is also a major limitation, although the consistent recruitment of the same regions across participants is persuasive.

An fMRI study of 20 young adults with a mean age of 24 involved the observation of moving balls and judgment of their physical causality (Blos, Chatterjee, Kircher, & Straube, 2012). Physical causality judgments were associated with activation in the insulae (Brodmann Area (BA) 47), right angular gyrus (AG) (BA 39), right IFG (BA 44), and bilateral supplementary motor area (SMA) (BA 8). These neural correlates were thought to reflect visual perceptual processing, moving objects processing, and higher order processing (Blos et al., 2012). These regions showed some overlap with those reported by Mason and Just (2016) when physics students thought about causality concepts (BA 7/8/27/39). Common regions of activation were the AG, which may relate to spatial manipulations, and the SMA which is thought to play a role in time perception (Blos et al., 2012). The recruitment of similar regions suggests that thinking about causal processes and observing causal events may be supported by similar cognitive mechanisms.

Finally, the tracking of moving balls over moving three-dimensional floor planes was examined in 21 adults with a mean age of 25 years (Jahn, Wendt, Lotze, Papenmeier, & Huff, 2012). Tracking of the balls was associated with broad bilateral activation in the superior parietal lobule (SPL) (BA 7/40), IPS (BA 5/7/19), precuneus (BA 5), frontal eye fields (BA 6), SMA (BA 6), inferior precentral sulcus (BA 6), insula (BA 48/25), lateral occipital cortex (BA 37), and subcortical regions (including BA 37). These activations were taken to reflect spatial memory and imagery (Jahn et al., 2012). The task was similar to the task involving observation of moving balls (Blos et al., 2012), and yet a much broader brain network was implicated with no noticeable overlap. This is surprising because such similar tasks would be expected to lead to activation of some common regions. This may be due to the control condition used by Jahn and colleagues (2012), which required passive observation of the same stimuli, but without any need to actively track to complete the task. In contrast, the study by Blos and colleagues (2012) had a control condition with the same stimuli but they were 'social' in the sense that the balls were named and participants were instructed to think of their movements as social interactions. This condition still required participants to consider the movements of the balls and their positions in relation to one another, which may have led to less of a contrast between the experimental and control conditions. These differences indicate that the control conditions of the different studies which seem similar visually were in fact calling on different cognitive processes. Much more research is needed to clarify the role of different brain regions in observing and thinking about physics concepts and events, perhaps with control stimuli that show non-physical events.

Within the discipline of science there is a lack of research into the development of brain regions supporting science through childhood and adolescence. The studies so far also focus on a narrow interpretation of science, only addressing physics concepts and not covering the posited core skills in science.

#### 1.2.2.2 Maths

As in the behavioural studies, the literature pertaining to maths is more developed, although the focus has been on arithmetic, with little reference to other types of mathematical reasoning, such as geometry or probability, and a lack of focus on concepts. Nonetheless, there are three consistently observed parietal regions considered key for number and arithmetic. The IPS is activated bilaterally during magnitude representation, and as such has been shown to be involved in most tasks that require numbers (Butterworth & Varma, 2013). Left-lateralised AG supports the retrieval of number facts, and the posterior SPL is recruited when relating numbers to space (Butterworth & Varma, 2013). While these brain regions are commonly observed in adult samples, some studies suggest changes in brain regions from childhood to adulthood. Developmental fMRI studies will now be considered.

A meta-analysis of seven neuroimaging maths studies which included 88 participants in total reported a shift from recruitment of the frontal cortex in childhood to parietal regions in adulthood (Houdé, Rossi, Lubin, & Joliot, 2010). This finding corroborates a similar conclusion from a study of 8- to 19-year-olds that was not included in the meta-analysis. Seventeen participants were shown arithmetic equations and were required to judge whether or not they were correct (Rivera, Reiss, Eckert, & Menon, 2005). With increasing age there was decreased activation in frontal regions (BA 8/9/11/24/32/46/47), basal ganglia, left medial temporal lobe, brainstem, and left anterior insula. There was increased activation with age in inferior and middle temporal gyri (BA 21/37), middle and inferior occipital gyri (BA 21/37), left supramarginal gyrus (BA 40), and left IPS (BA 7) (Rivera et al., 2005). Overall, there was a large shift across a range of brain regions, mainly characterised by less recruitment of frontal regions and greater recruitment of temporal, parietal, and occipital regions. These findings appear to demonstrate a reduction in the recruitment of cognitive control processes during arithmetic with age, alongside increased reliance on declarative and procedural memory.

A more recent review of 38 studies of arithmetic in the developing brain suggested that the decrease in prefrontal cortex (PFC) and increase in posterior parietal cortex reflect a change in strategy from calculation to fact retrieval (Peters & De Smedt, 2018). The review also suggested that maths training leads to a shift in parietal activation from the IPS to the AG, which might suggest increased used of retrieval strategies and reduced use of backup strategies (Peters & De Smedt, 2018). Backup strategies might include counting or decomposition, whereby a problem is broken down into smaller parts. A meta-analysis of children’s number and calculations highlighted the potential importance of deeper brain regions that are rarely considered (Arsalidou, Pawliw-Levac, Sadeghi, & Pascual-Leone, 2018). The authors suggested that for children under 14 years the insula might have a particular role in intrinsically motivated behaviour within maths, while the claustrum might particularly support cross-modal integration of top-down and bottom-up processes (Arsalidou et al., 2018). Additionally, within this age range, the inferior parietal lobule (IPL, BA 40) appeared to serve different functions in the right versus the left hemisphere. The right IPL was more involved in number tasks with no calculation (such as comparing arrays of digits or dots) while the left IPL was recruited for calculation tasks, demonstrating increased hemispheric specialisation for different aspects of maths (Arsalidou et al., 2018).

Fifteen professional mathematicians and 15 non-mathematicians (who were of equal academic standing in humanities) listened to maths statements relating to analysis, algebra, topology, and geometry, and non-maths control statements, and judged whether they were true or false (Amalric & Dehaene, 2016). In terms of behavioural data, both groups performed equally well on the non-maths statements (above chance), while mathematicians performed above chance on maths statements, and non-mathematicians performed close to chance level on maths statements (Amalric & Dehaene, 2016). Mathematicians recruited a consistent network of bilateral intraparietal, inferior temporal, and dorsal prefrontal regions that supported all types of maths thinking but not thinking about non-mathematical statements. Since accuracy between trial types was the same for mathematicians, the recruitment of this network cannot be attributable to a performance-related process such as error-monitoring. In addition, participants observed a range of images including formulae and numbers, and non-maths items like faces and houses. There was selective recruitment of the inferior temporal gyrus to numbers, consistent with previous work indicating that this site is involved in processing visual number form. The authors argued that the recruitment of dorsal prefrontal and intraparietal regions reflect a complex integration of many requirements, including numerical, ordinal, logical, and spatial reasoning (Amalric & Dehaene, 2016). It is interesting to compare these results with those from the study of different physics concepts, which found distinct neural correlates for different aspects of physics in scientists (Mason & Just, 2016). This may indicate that at the expert level, science draws on more domain-specific skills that are relevant to individual topics, whereas maths requires more domain-general skills that are applied more broadly.

There is a clear gap in the literature for research into regions associated with educationally-relevant science reasoning in childhood and adolescence. While there have been investigations into the developing brain regions underlying maths reasoning, these have typically focussed on number and basic arithmetic, and there remains a lack of broader investigations into school-related stimuli, particularly as related to the wide range of topics covered in secondary school. Nonetheless, these investigations have been fruitful and revealed increasing automaticity and learning of number facts with age. A better understanding of the neural underpinnings of science and maths through development might help the emergence of a clearer picture of the cognitive mechanisms involved. In turn this would lead to a better understanding of the skills that should be focussed on in school and in instances where individuals struggle.

### 1.2.3 Misconceptions in science and maths

While science and maths are typically studied separately, a common thread that pulls them together is the presence of counterintuitive concepts that lead to misconceptions. A wide body of research has shown that misconceptions are prevalent in science and maths in all learners (not just those who struggle), and research has shown them to persist into adulthood, even when the correct concepts have been taught in school. Stavy and Tirosh (2000) brought the evidence relating to science and maths misconceptions together, and observed that many similar responses are present in incorrect reasoning about counterintuitive concepts. They proposed a theory of three intuitive rules that are common to science and maths and lead to misconceptions throughout the life course.

The first of these rules they labelled ‘More A—More B’. This refers to the intuition that more of one substance corresponds to more of the other. In **Figure 1.3a**, an intuitive response to the stimuli is that angle *x* is larger than angle *w*, when in fact they are the same size. Stavy and Tirosh (2000) suggest that this arises because the quantity of one item (in this case, the arc) is incorrectly assumed to relate to the quantity of another (here, the size of the angle). The second intuitive rule, called ‘Same A—Same B’, is a similar intuition regarding the quantity of a substance or dimension, this time that when two quantities are the same in one manner, they are the same in another. This rule is exemplified by **Figure 1.3b**, where an intuitive response to the stimuli is to reason that shapes *y* and *z* have the same perimeter, because their areas are the same, when in fact shape *y* has a larger perimeter. The final intuitive rule is termed ‘Everything Can Be Divided’. **Figure 1.3c** demonstrates an example of an incorrectly applied understanding of infinity; the intuitive response here is to think that a piece of paper can infinitely be divided in half, when this would not be possible. This framework is a useful way of thinking about misconceptions in science and maths; that they arise from intuitions that everyone experiences.

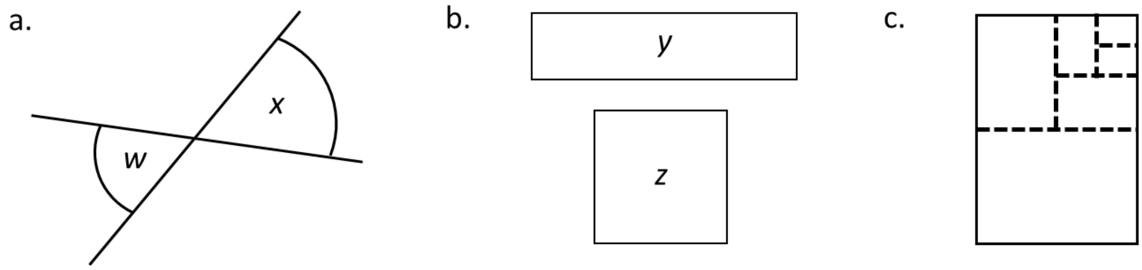



**Figure 1.3.** Example intuitive rules leading to misconceptions (adapted from Stavy & Tirosh, 2000). (a) More A—More B. An intuitive reasoning is that angle *x* is larger than angle *w* because angle *x*'s arc is longer. (b) Same A—Same B. An intuitive reasoning is that shapes *y* and *z* have equal perimeters because their areas are equal. (c) Everything Can Be Divided. An intuitive reasoning is that a piece of paper can be cut in half infinitely.

The origins of these intuitive rules may lie in previous encounters with the world. For example, it is increasingly argued that misconceptions arise through misleading perceptual information, experiences that appear to contradict scientific reality, or through simplified ideas that are taught at an early age but are later superseded by more detailed concepts (Dunbar, Fugelsang, & Stein, 2007; Houdé, 2000; Mareschal, 2016). **Figures 1.3a and b** are examples of misleading perceptual cues, whereby the visual information provided gives an impression that is contrary to reality. Experiences that appear to contradict scientific reality include the fact that the Earth revolves around the Sun, and yet it appears as though the Sun revolves around the Earth, 'rising' and 'setting' every day. Even the language used to describe our view of the Sun is misleading. In other cases, school pupils may be taught a simplified theory that constitutes an early introduction to a challenging and complicated topic. A few years later, a more accurate version of events may be taught, perhaps contradicting the original information. One example is that at an early age, we are taught that the number 2 is smaller than the number 4; 2 is always less than 4. However, when we start learning about fractions, ½ is larger than ¼; now the number containing a 4 is smaller than the number that contains a 2. All of these factors can make reasoning and learning new information in science and maths challenging.

Key Stage 3 curricula for science and maths in England (for 11- to 14-year-olds) acknowledge the presence of misconceptions but provide little information regarding the source of misconceptions and no examples of commonly held misconceptions (Department for Education, 2013b, 2013a). Both curricula encourage teachers to use discussion to probe and remedy misconceptions, with no reference to the best way to tackle them. A recent hypothesis considers that one way of overcoming misconceptions is through the use of inhibitory control; an executive function that enables the suppression of irrelevant ideas or information. The following sections will briefly introduce executive functions, before specifically focussing on inhibitory control, its behavioural and neural development, and its proposed link to science and maths.

## 1.3 Executive functions

Executive functions (sometimes referred to as cognitive control or executive control) are a set of effortful, top-down processes that enable purposeful behaviours. They are required when concentration and attention are necessary (Diamond, 2013), and in the formation of goals, the preparation for action, and the verification that plans have been adequately implemented (Anderson, 2002). The study of executive functions was historically focussed on patients with frontal lobe damage, who showed impairment on complex ‘frontal lobe’ tasks that heavily taxed executive functions (Miyake et al., 2000). More recently, research is also concerned with individual differences in executive functions in typical and atypical populations through development, as they are considered to be important across many aspects of life. Executive functions are thought to be essential for mental health, physical health, and quality of life, as well as school readiness and school success (Diamond, 2013). While executive functions are typically conceptualised as separable, they operate in a highly integrative manner, and are inter-related and inter-dependent (Anderson, 2002). Measuring executive functions is therefore a challenge, as it is difficult to tease apart the influence of different skills on test performance. An additional challenge of investigating executive functions is the malleability that these skills appear to possess, as demonstrated by their trainability (Blair, 2016).

Executive functions are affected by a range of additional factors, including stress, loneliness, lack of sleep and lack of exercise, each of which impair executive function (Diamond, 2013). Finally, executive functions appear to be a key contributor to the poverty-related gap in school success (Blair, 2016). Further establishing links between individual differences in executive functions and school-relevant measures may lead to practical suggestions for teaching and learning that encourage executive function development for school success. These investigations are especially appealing given that executive function training seems to be of particular benefit to those who start from a baseline level of poorer executive functions and social disadvantage (Diamond, 2013).

## 1.4 Inhibitory control

Inhibitory control, or inhibition, refers to the aspect of executive functioning that enables suppression of a prepotent response. Within the dominant model of executive functioning, inhibitory control is one of three separable executive functions; the second is shifting or flexibility (the switching between tasks or mental sets), and the third is working memory (the updating and monitoring of information in mind) (Diamond, 2013; Miyake et al., 2000). These three functions contribute to different degrees on complex tasks, but there is a moderate correlation between them (Miyake et al., 2000), showing that they are not entirely independent processes. Executive functions enable higher order cognitive skills, including problem-solving, planning, and reasoning (Diamond, 2013). While the three-factor model is dominant, developmental research indicates that the factors may be less differentiated early in childhood, and separate with development (Lee, Bull, & Ho, 2013). Preschool children appear to exhibit a single executive function factor, with protracted specialisation and differentiation until early adolescence where the three-factor model best describes performance (Lee et al., 2013). A less recent model of executive function proposed a four-factor model that includes attentional control (inhibitory control falls within this factor), cognitive flexibility, goal setting, and information processing (Anderson, 2002). Again, within this model, the factors are considered discrete but related. It is more common for planning to be categorised as a higher order skill (Diamond, 2013), and to consider information processing (including processing speed) a separate factor that contributes to all executive functions (Lee et al., 2013).

This thesis therefore conceptualises executive function according to the dominant three-factor model, which considers inhibitory control one component of three separate but related factors. Inhibitory control is given particular priority within the thesis, since it is thought to have a specific role in science and maths reasoning in relation to misconceptions, or reasoning about counterintuitive concepts (see section 1.3.3.3).

Inhibitory control is not generally considered a unitary executive function. Nigg (2000) outlined eight kinds of inhibition, which were split into three categories: executive inhibition, motivational inhibition, and automatic inhibition of attention. More commonly though, inhibitory control is grouped into two categories (Verbruggen, Liefooghe, & Vandierendonck, 2004). The first of the commonly used two categories is behavioural, or response inhibition, which is the suppression of a motor response. Simple response inhibition tasks aim to measure pure motor suppression, while complex response inhibition tasks aim to capture motor suppression in the context of a small cognitive load, thus requiring working memory (the ability to hold and manipulate information in mind). The second type of inhibitory control is interference control, or semantic inhibition, which refers to the suppression of meaning in the face of conflict. Many different types of task are used to measure these two types of inhibition through development (**Table 1.1, Figure 1.4**). A key debate considers the extent to which different inhibitory control tasks measure the same underlying construct; there is currently no consensus on this point (Khng & Lee, 2014; Verbruggen et al., 2004).

**Table 1.1.** Summary of commonly used inhibitory control tasks.

| Task | Description | Dependent variables | Example |
|---|---|---|---|
| *Response inhibition tasks (see **Figure 1.4a**)* | | | |
| Stop signal | The ‘primary task’ requires rapid response to stimuli, and infrequent non-response when a signal is given shortly after a stimulus (typically in another modality). | Stop signal RT (SSRT) which is estimated based on RT to the primary task (PTRT) and the probability of inhibiting correctly. | Left or right arrows appear on the screen, and the corresponding left or right key must be pressed in response (primary task). No response should be made when a beep sounds after an arrow is displayed (stop signal). |
| Simple Go/No-Go | Rapid response to frequent ‘Go’ stimuli (typically 75% of trials), and non-response to infrequent ‘No-Go’ stimuli (typically 25% of trials). | Commission errors (response to No-Go trials), omission errors (non-response to Go trials), and RT on correct Go trials. | Green squares appear on the left or right of the screen, and the corresponding left or right key must be pressed in response (Go trials). No response should be made to red squares (No-Go trials). |
| Complex Go/No-Go | As above, with a 1-back component so that ‘No-Go’ stimuli depend on both the current and previous trial. | As above. | Blue and yellow squares appear on the left or right of the screen, and the corresponding key must be pressed in response (Go trials). No response should be made to blue squares that follow yellow squares (No-Go trials). |

| | | | |
|---|---|---|---|
| Anti-saccade | An eye-tracking task, where the aim is to look away from the target stimulus (anti-saccade). | Looking accuracy (errors are pro-saccades, where the stimulus is looked at) and RT of anti-saccades. | Initially a fixation cross appears in the centre of the screen, and should be focussed on. A dot appears on the left or right of the screen, and the correct response is to look to the opposite side of the screen. |
| *Semantic inhibition tasks (see* ***Figure 1.4b****)* | | | |
| Stroop | Two competing representations are associated with a single stimulus. Only one of these representations is relevant for the choice of response. The representations match in congruent trials, and the representations do not match in incongruent trials. | Accuracy incongruency cost (equal to congruent accuracy – incongruent accuracy), and RT cost (equal to incongruent RT – congruent RT, for correct trials only). | Colour words appear on the screen and the text colour, which may conflict with the word, should be responded to. For example, the word blue appears in blue coloured text (congruent trial, correct answer is blue), or the word red appears in yellow coloured text (incongruent trial, correct answer is yellow). |
| Flanker | A speeded choice RT task to a target, while distractors on either side of the target either match the target (congruent trials) or provide conflicting information (incongruent trials). | As above. | On each trial five left or right arrows appear. Participants should respond to the central arrow with the corresponding left or right key. The arrows on either side (the flankers), point either in the same direction (congruent trials) or in the opposite direction (incongruent trials) to the central arrow. |

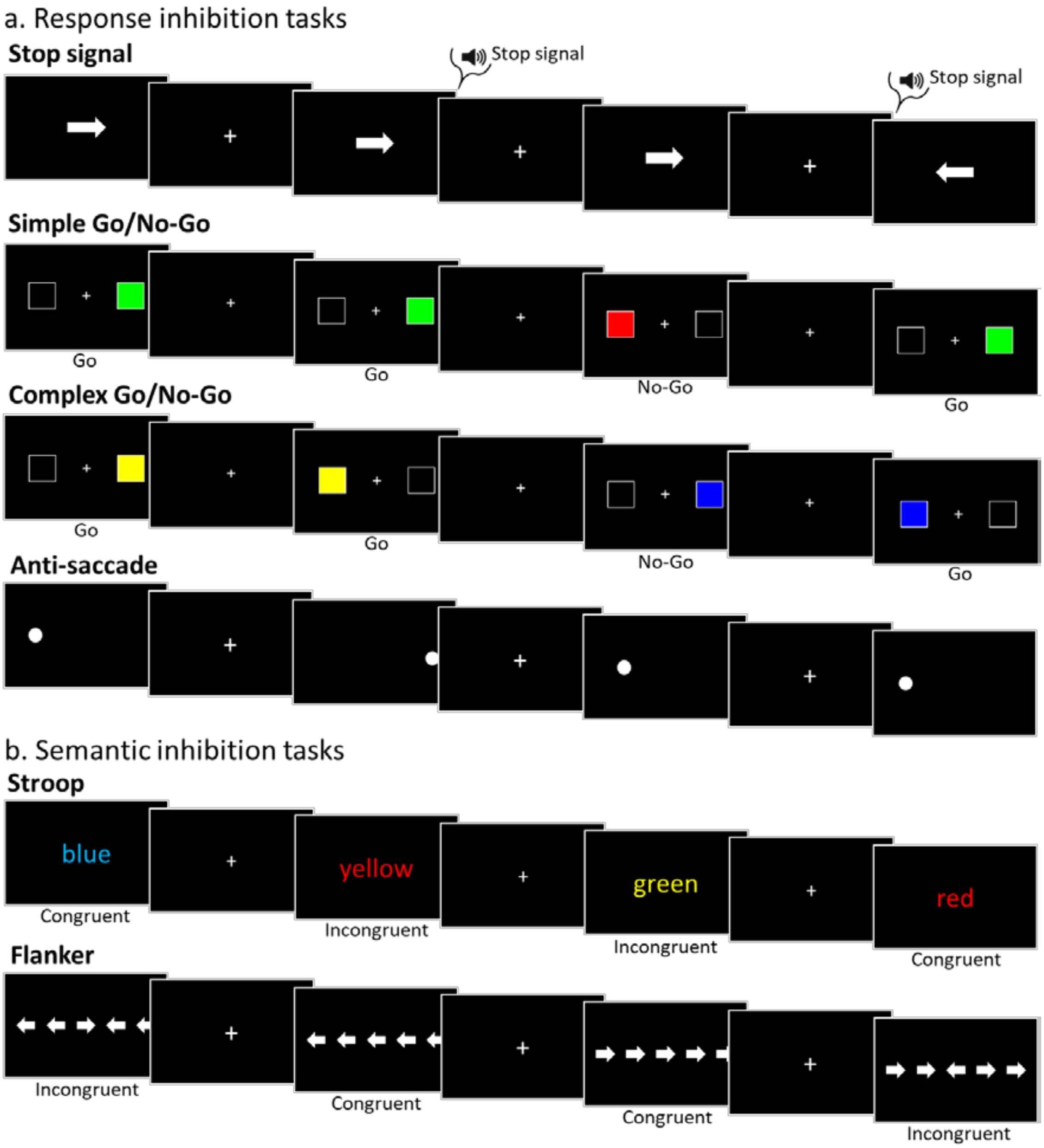



**Figure 1.4.** Example stimuli for (a) response and (b) semantic inhibition tasks.

Given the varied tasks used to measure inhibitory control, capturing developmental trends is a challenge. The tasks highlighted here are used frequently with adolescents and adults, but typically cannot be performed with younger participants as they are too difficult and lead to floor effects, where participants consistently perform poorly. Conversely, those used with young children are not appropriate for older participants because they lead to ceiling effects of very high performance. In cases of floor and ceiling effects, individual differences within age ranges are reduced, precluding in depth examination of trends and associations. A further challenge is that inhibitory control is thought to show heterotypic continuity: the manifestation of inhibitory control may be different in children compared to adults (Petersen, Hoyniak, McQuillan, Bates, & Staples, 2016). In practice this means that young children may engage their inhibitory control skills in a very different type of scenario to that of an adult: for example a child may inhibit the desire to snatch from another child, while an adult may inhibit the desire to have another drink with friends. These examples highlight the additional differences in inhibitory control manifestation that relate to social factors, which may have particular importance during adolescence, a time of social development (Burnett, Bird, Moll, Frith, & Blakemore, 2008), than at other ages. The social element of executive skills is investigated in tasks that manipulate how 'hot' or 'cool' a task is, whereby hot tasks involve social or reward-focussed motivation and cool tasks have no social factors or reward components. However, the role of social factors in inhibitory control is beyond the scope of this review and is likely only tangentially related to the overall topic of science and maths reasoning.

### 1.4.1 Behavioural development of inhibitory control

The exploration of the behavioural development of inhibitory control is embedded within a larger body of work that investigates the inhibitory control during overall cognitive development. It has been posited that inhibitory control, or 'resistance to interference', is a major component in cognitive development, influencing performance in a wide range of situations (Dempster, 1992). This section will briefly outline early development in inhibitory control, then take response and semantic inhibition in turn, finally bringing them both together.

#### 1.4.1.1 Early development

Petersen and colleagues (2016) conducted a comprehensive meta-analysis of 198 behavioural inhibitory control studies with children under 8 years. To cater for the young participants, tasks differed to those described in **Table 1.1** and **Figure 1.4**. For example, in a game of *Simon Says* instructions must only be followed when preceded with the words "Simon says…" and otherwise ignored (inhibited). In another example, in the Day/Night task, children are instructed to say "day" when they see a sun and "night" when they see a moon in congruent trials, and to say "night" when they see a sun and "day" when they see a moon in incongruent trials. A total of 14 different inhibitory control tasks were included in the meta-analysis. Note that the authors describe all of the tasks as complex response inhibition tasks, which is at odds with the definitions of inhibitory control given above. The Day/Night task in particular would be better described as a semantic inhibition task, as the participants must suppress the meaning associated with the pictures in order to succeed in the task. The tasks also appear to have no additional working memory component, so would not be considered complex in the same way as the complex Go/No-Go which incorporates a small cognitive load. These definitional complications speak to a larger field-wide issue, with terminology varying between papers. In part this may be due to the differing age groups, with child, adolescent, and adult research being described by researchers in inconsistent ways. Definitional problems are cited as one reason for difficulties in advancements in the field (Khng & Lee, 2014), making this a non-trivial issue.

Nonetheless, the findings from the meta-analysis revealed different developmental patterns for four different types of inhibitory control tasks in the early years, categorised by task demand (Petersen et al., 2016). The earliest development was seen in tasks where perceptual information needed to be inhibited, with the greatest change between 1.5 to 3.5 years of age, followed by development between 2.5 to 6 years in tasks where one response had to be inhibited and a different response generated. In terms of response versus semantic inhibition, these tasks appear to rely on semantic inhibition skills, where meaning needs to be inhibited in the face of conflict. The next to develop was what the authors termed performance inhibition, which closely matches a standard response inhibition definition, between 3 and 6 years. And finally, performance in motivational inhibition tasks, which included a hot component such as a win, developed until age 6.5 years. According to this extensive review, semantic inhibition skills appear to mature earlier than response inhibition.

#### 1.4.1.2 Response inhibition

A number of response inhibition tasks have been examined through early childhood, and into adolescence and adulthood. A study of performance in the stop signal task navigated the age issue by modifying the task according to participant age: longer presentation times and fewer trials were presented to younger participants in order to account for decreased attention and longer RTs (Carver, Livesey, & Charles, 2001). The 94 participants were separated into four age groups: pre-school children (4- to 5-year-olds), young primary children (5- to 7-year-olds), mid primary children (7- to 9-year-olds), and adults (17- to 43-year-olds). The two youngest groups did not perform significantly differently from each other, but performance improved with age between the other groups, reflected by faster RTs and more correctly inhibited responses (Carver et al., 2001). Other studies of stop signal performance through development, but with no age adjustments to the task, included older participants and those in adolescence. Speed of stopping improved through childhood and adolescence (6 to 17 years), and then diminished through adulthood (18 to 81 years) in a sample of 275 participants (Williams, Ponesse, Schachar, Logan, & Tannock, 1999). The same trend was seen in a study of 317 6- to 82-year-olds, with improved performance through childhood and adolescence, then a decrease in adulthood (Bedard et al., 2002). These studies suggest that performance in the stop signal does not improve much between the ages of about 4 and 6 years, then gets better through the rest of childhood and adolescence before decreasing in adulthood.

The Go/No-Go response inhibition task differs from the stop signal task in that the response to No-Go trials should not even be prepared, as opposed to being prepared and subsequently aborted on the signal. Task difficulty in the Go/No-Go was varied in one study of 20 6- to 10-year-olds and adults by manipulating the number of Go trials preceding No-Go trials (Durston, et al., 2002). The adults were faster and more accurate than the children, and the number of commission errors increased with the number of preceding Go trials for both age groups (Durston, et al., 2002). Adolescents have also been tested on the Go/No-Go, with one study of 90 participants finding lower error rates and faster RTs in 15-year-olds compared to 12-year-olds on both the simple and complex Go/No-Go, but no difference between 15- and 17-year-olds (Humphrey & Dumontheil, 2016). A less well powered study of 19 8- to 20-year-olds found high performance overall with no accuracy changes according to age, but with faster RTs with increasing age (Tamm, Menon, & Reiss, 2002). The lack of accuracy effects may be due to the 50:50 ratio of Go to No-Go trials, which was used in this neuroimaging study in order to increase the frequency of trials requiring inhibition inside the scanner.

The final commonly used response inhibition task in adolescence is the anti-saccade task, which differs from the aforementioned response inhibition tasks in requiring inhibition of an eye movement, rather than a finger movement. A study of 129 participants aged 9 to 26 years, with multiple testing sessions (totalling 312 sessions), found that behavioural performance increased with age, and rate of improvement decelerated with increasing age (Ordaz, Foran, Velanova, & Luna, 2013). A larger study, of 245 8- to 30-year-olds, similarly found large improvements from childhood to adolescence, followed by a plateau from late adolescence to early adulthood, with adult-level performance considered reached at 14 to 15 years of age (Luna, Garver, Urban, Lazar, & Sweeney, 2004). Overall the behavioural research into response inhibition across tasks presents a picture of protracted improvement throughout early childhood and adolescence, with slowed improvement through late adolescence and early adulthood, and decreasing performance in late adulthood.

#### 1.4.1.3 Semantic inhibition

Studies of semantic inhibition have also tracked longer term developmental trajectories beyond early childhood. Possibly the most well-known semantic inhibition task is the Stroop. A behavioural study of the colour-word Stroop task with 99 6- to 17-year-olds found that errors and RTs decreased linearly with age, when reading speed was held constant (Leon-Carrion et al., 2004). Without controlling for reading speed, distractor interference increased between 6 and 8 years of age for RT, and between 6 and 10 years for errors, before decreasing until the oldest age tested (Leon-Carrion et al., 2004). This study highlights the importance of skills that are not typically tested when measuring inhibitory control, but that may be hugely important in determining the patterns observed. In the traditional Stroop task, reading automaticity is clearly a moderating factor and so controlling for reading speed enabled a clearer picture of steady improvement with age to emerge. An earlier study of 235 7- to 80-year-olds found that performance in the Stroop improved through childhood, adolescence, and into adulthood, but then got worse in late adulthood (Comalli, Wapner, & Werner, 1962), mirroring findings in older adults with the stop signal.

A second commonly used semantic inhibition task, the flanker task, has no reading requirements and the conflict of information is in the contrast of the target stimulus to the surrounding distractors. The flanker task therefore does not have the verbal/reading component of the traditional Stroop task. A study of 62 5- to 12-year-olds and young adults showed a reduction in errors with age during childhood, but no difference in errors between 10- to 12-year-olds and adults (Ridderinkhof & van der Molen, 1995). A more recent flanker study of 7-year-olds, 10-year-olds, and adults showed that interference reduced between 7 and 10 years of age, but that 10-year-olds did not differ from adults (Cragg, 2016). Together these studies may indicate that the flanker task is not such a sensitive measure of semantic inhibition, as ability appeared to mature by the age of 10. Alternatively, this less verbal measure of semantic inhibition may better allow younger participants to perform well, unaffected by their still-maturing verbal ability. A final possibility is that these studies did not capture improvement that might otherwise be seen in adolescence, before a reduction in adulthood.

### 1.4.1.4 Comparing response and semantic inhibition

Overall, the studies discussed here have shown a relatively consistent picture of the development of inhibitory control through childhood and adolescence. Both response and semantic inhibition tasks show protracted maturation throughout early childhood and into adolescence, typically with a slowing or plateau in adulthood, followed by a reduction in ability in late adulthood. Some studies have compared performance on more than one inhibitory control task within the same sample in order to establish the degree to which these tasks tap the same underlying construct. A comprehensive study of adolescent executive function development measured performance on the stop signal task, the Stroop, and the flanker in 384 7-, 11-, 15- and 21-year-olds (Huizinga, Dolan, & van der Molen, 2006). Performance in both the stop signal and flanker showed improvement with age until age 15, while Stroop performance showed more protracted development, with improvement continuing up to the oldest group (21 years). A structural equation model did not find a common inhibition factor from the three inhibitory control tasks, as correlations between the tasks were very low. The authors argued that the lack of association between tasks may be due to the different task requirements (Huizinga et al., 2006), but they did not consider that the tasks may be tapping different types of inhibitory control. The use of discrete age groups, with many ages not tested (particularly during adolescence) may also mean that developmental trends are not well captured here. Nonetheless these results appear to indicate that inhibitory control is not a single factor, nor simply two factors that differ according to whether they require suppression of a motor response or semantic information.

The Stroop and Go/No-Go were compared in a study of 108 6- to 14-year-olds (Morooka et al., 2012). Performance on both tasks improved with age, but rates of maturation differed between tasks. Greater changes in Stroop performance were seen at younger ages, while greater Go/No-Go improvements were seen in the older participants. Significant correlations were found between different measures of the two tasks, although the coefficients were generally low. The authors argued that while there may be similarities between the tasks, they primarily capture different processes (Morooka et al., 2012). Performance in the Stroop and stop signal tasks were compared in a study of 237 13- to 15-year-olds (Khng & Lee, 2014). The study found minimal association between RT measures across tasks, which is noteworthy as we might expect these to be the most related factors of the tasks, reflecting an underlying general processing or response speed. Across a number of measures in the tasks, only Stroop intrusion errors on incongruent trials and SSRT were correlated. The authors argued that while there may be some association between the two tasks, it could be due to non-inhibition processes such as proactive, attention-based goal-directed behaviour (Khng & Lee, 2014). An alternative explanation not considered by the authors is that these tasks measure two related aspects of inhibitory control that develop at different rates, thus not showing much correlation. Given the research discussed so far, age effects would be likely in this group of 13- to 15-year-olds, yet age was not considered as an interacting factor, or even as a main effect, within this study.

To conclude, both response and semantic inhibition show improvement behaviourally from early childhood, through late childhood and adolescence, peaking around late adolescence or early adulthood, and plateauing before decreasing in late adulthood. Attempts to compare behaviour on more than one inhibitory control task have shown that there is little association between measures on different tasks. The reason for this disconnect remains unclear. The tasks used may require too many other skills to adequately tap one pure inhibition skill: for example, requiring working memory, reading, or processing speed, or depending on motivation. Alternatively, the tasks may in fact capture different aspects of inhibitory control, which seems to not be a single unified ability but separated between response and semantic inhibition, or perhaps even more subdivisions (for example, Nigg's (2000) eight types of inhibition). Neuroimaging research into inhibitory control is able to uncover more about the mechanisms involved in these different tasks.

### 1.4.2 Neural correlates of developing inhibitory control

Examining neural correlates may be able to highlight what drives development in inhibition, and expose the cognitive processes that underlie performance in different tasks. The neural underpinnings of both response and semantic inhibitory control appear to develop in line with improved performance. Neuroimaging evidence suggests that the PFC, and in particular the dorsolateral prefrontal cortex (DLPFC), anterior cingulate cortex (ACC) and IFG, show changes in activation associated with inhibitory control over the course of late childhood and adolescence. The remainder of this section will again consider response and semantic inhibition in turn, drawing on the relevant fMRI literature, and then consider the issues in teasing apart inhibition effects in the brain.

### 1.4.2.1 Response inhibition

The aforementioned longitudinal study of 123 9- to 26-year-olds examined changes in brain activation during the anti-saccade task (Ordaz et al., 2013). Activation in the right DLPFC (BA 44) decreased from childhood until adolescence, while error-related activation in the dorsal ACC (BA 24) decreased until adulthood and was associated with the percentage of inhibition errors. This study suggests that it is a change in error processing that is most important in reaching adult performance (Ordaz et al., 2013): the decrease in activation with age suggests that automation, increased efficiency, and reduced effort occur with improved inhibition performance. While this is the only longitudinal study, cross-sectional studies have revealed more about the specific roles of different brain regions during developing inhibition.

Tamm and colleagues (2002) similarly found greater activation in younger participants compared to older participants in the left PFC (BA 8) during the Go/No-Go, in their sample of 19 participants aged 8 to 20 years. This finding was taken to partly reflect the role of working memory, highlighting the difficulty (i.e. reduced automaticity) of such requirements for younger participants. Conversely, the study also found an increase in left IFG (BA 48) activation with age, thought to show a more specific role for this region in effective response inhibition. Frontal cortex activation was overall more diffuse in the younger participants, which the authors argued reflects less efficient organisation and monitoring, and use of less efficient strategies (Tamm et al., 2002). The role of the IFG in response inhibition was further supported in a stop signal task with 26 10- to 17-year-olds and 21 20- to 42-year-olds (Rubia, Smith, Taylor, & Brammer, 2007). The adult group showed increased activation in the right IFG (BA 45) compared to the child and adolescent group during successful inhibition; additional evidence that the IFG becomes increasingly specialised for response inhibition with age. When inhibition was unsuccessful, adults showed greater activation in the ACC (BA 24/32) in comparison to the child and adolescent group (Rubia et al., 2007). This further supports the finding from Ordaz and colleagues (2013) that ACC involvement in inhibitory control relates to error processing, since the activation occurred when adults were likely noticing and reflecting on their errors of commission. This may mean that the younger participants were less likely to notice their errors than the adults.

There remains debate surrounding the specific role of the right IFG in response inhibition. One hypothesis argues that it is involved in a global stopping system, where a signal from the right IFG to the subthalamic nucleus of the basal ganglia suppresses thalamocortical output, which prevents a motor response (Banich & Depue, 2015).

Although converging evidence from a number of different methods support this view, a recent hypothesis suggests that in fact, the right IFG does not have a specific role in inhibition, rather it monitors external information from the environment to implement goal-related actions (Banich & Depue, 2015). As with all neuroimaging research, uncovering the precise role of different brain regions is a challenge. Although it is clear that the right IFG plays a role in response inhibition, it remains unclear whether this role is as a unifying node for response inhibition.

#### 1.4.2.2 Semantic inhibition

The neural underpinnings of semantic inhibition are less studied but similarly show age-related changes. An fMRI study of 30 participants aged 7 to 22 years investigated brain activation during the Stroop task (Adleman et al., 2002). There was a positive correlation between age and left-lateralised activation in the ACC (BA 24/32) and IPL (BA 31/7), and no reduction in activations with age in any brain region. A larger fMRI study of 70 7- to 57-year-olds during the Stroop task found that activation in the right PFC (BA 46) increased with age and related to better performance. Conversely, activation in the left ACC (BA 24/32) decreased with age, as did activation in the right superior temporal gyrus (BA 22) (Marsh et al., 2008). The study found no evidence of a more diffuse pattern of activation in younger participants (Marsh et al., 2006). The authors acknowledged that reading ability may play a role in these findings, but concluded that these results show the brain regions which underlie the transition to better inhibition and control. Overall, the pattern of brain activation associated with developing semantic inhibition is less clear than with response inhibition. However, similar brain regions are at play; in particular the PFC and ACC, indicating that while the developmental trajectory of the use of these regions in semantic inhibition are unclear, they remain important in the development of inhibitory control.

#### 1.4.2.3 Issues in investigating the neural mechanisms of inhibitory control

Some of the neuroimaging studies examined here have emphasised the importance of other factors in attempting to measure inhibitory control. Inhibition is closely related to other cognitive control mechanisms, particularly working memory (Davidson, Amso, Anderson, & Diamond, 2006) and attention (Jaeger, 2013), since a rule has to be held in mind, and stimuli must be constantly focussed on and processed. Dissociating these factors is difficult, particularly when tasks utilise a complex version, aiming to involve a working memory component. It has been argued that simpler forms of inhibitory control tasks would help to provide a purer measure of inhibition (Criaud & Boulinguez, 2013). However, using complex tasks that require working memory might tell us more about real-world abilities, where inhibitory control is likely required in the context of other cognitive mechanisms. In science and maths for instance, inhibitory control is not thought to be the only executive mechanism involved. Criaud and Boulinguez (2013) also point out that within many of these neural investigations, it is assumed that No-Go trials uniquely require inhibition, while Go trials do not. In reality, Go trials likely require inhibition to some extent, as participants must inhibit the desire to respond immediately, before a decision is made. Nonetheless, greater inhibition is necessary for No-Go trials than Go trials since responses must be fully suppressed.

It is worth noting that while these tasks are designed to measure inhibition, the processes involved could equally be described as amplifying the intended process, as opposed to inhibiting unwanted processes (Aron, 2007). For example, rather than inhibition, these tasks may rely on the active process of successful goal activation (Khng & Lee, 2014). It seems likely that active and inhibitory processes work together in these tasks proactively and reactively (Khng & Lee, 2014). Many of the studies discussed here do acknowledge the ambiguity of the term ‘inhibition’ (e.g. Verbruggen et al., 2004), and seek to address this through studying associations between inhibitory control measures.

Neuroimaging studies of inhibitory control have not focussed on the distinction between response and semantic inhibition, leaving this an important area of ongoing investigation. Response inhibition seems to be studied more than semantic inhibition in behavioural and neuroimaging studies, and semantic inhibition tasks are sometimes referred to as measuring response inhibition: the literature so far conveys an inconsistent picture. This chapter will now move on to reviewing the literature associating inhibitory control and science and maths. In considering the extent to which inhibitory control associates with other measures, it is important to keep the potential difference between response and semantic inhibition in mind. It may help to explain any inconsistent findings, and is likely to influence any later recommendations for education.

### 1.4.3 Inhibitory control in science and maths

Inhibitory control is thought to be involved in effective science and maths reasoning throughout development, in particular to enable the suppression of incorrect responses (Mareschal, 2016). The link has been examined using correlational, priming, and neuroimaging studies. While most studies have explored this link in either science or maths, one study investigated inhibitory control in a joint science and maths task. One hundred and thirty-three participants aged 9 to 11 years were given a computerised task called the Re-Categorisation task (Re-Cat), which involved categorising science and maths concepts (Vosniadou, Pnevmatikos, Makris, Eikospentaki, & Chountala, 2018). Participants were also given the Sentence-Picture Verification task (SP-Ver), which involved determining whether a scientific or mathematical statement accompanied by an image was true or false. Both semantic inhibitory control and shifting were tested in a Stroop-like task, and age and general cognitive ability were controlled for in the analyses. Inhibition accuracy only predicted SP-Ver accuracy for inconsistent trials (which required rejection of an incorrect concept), while shifting accuracy predicted accuracy in consistent SP-Ver trials (which did not require rejection of a concept) and Re-Cat accuracy, which also did not require rejection of an incorrect concept. The authors argued that this indicates that shifting has a general role in complex science and maths tasks, while inhibition has a *specific* role that is selectively recruited when a conflicting concept needs to be rejected (Vosniadou et al., 2018). The study did not consider science and maths separately, so it is unclear whether the same process applied to both disciplines. To address this question, science and maths studies will next be considered in turn.

#### 1.4.3.1 Science

Evidence from correlational studies suggests that children with better inhibitory control perform better on science problems requiring counterintuitive reasoning. In the domain of physics, the tubes task had an experimenter drop a ball down one of three opaque tubes that crossed over (Baker, Gjersoe, Sibielska-Woch, Leslie, & Hood, 2011). Thirty-six 3- and 4-year-olds were asked to point to the opaque container where the ball landed. The most common error was to choose the container directly beneath the opening of the tube, suggesting that the toddlers' gravity theory (that the ball would fall in a straight trajectory) was not successfully inhibited in favour of an object solidity theory (that the ball would follow the solid tube). Since looking time paradigms reveal an understanding of object solidity at this age (Hood, Cole-Davies, & Dias, 2003), the authors argued that the task engaged a number of strategies, and that selection of the correct strategy depended on the inhibition of incorrect strategies. This interpretation was further supported by a positive correlation between tubes task performance and inhibitory control as measured by a gift delay task (Baker et al., 2011). This task likely most closely reflects response inhibition as the measure records the time before the participant physically moves towards the gift. In biology, mature biological understandings of life, death, and bodily functions were predicted in 79 5- to 7-year-olds by an aggregate measure of executive function that captured performance on tasks requiring a combination of shifting, working memory, and response and semantic inhibition (Zaitchik et al., 2014). Although the executive function measure was not purely inhibitory control, the authors suggested that inhibitory control is one skill that enabled the suppression of naïve biological theories when a mature conceptual understanding was shown.

The picture is, however, inconsistent, as inhibitory control was not related to science performance in other studies. In a study of 155 10-year-olds, scientific reasoning was measured through a pencil and paper test which covered the design of experiments, interpreting data, scientific theories, and understanding the nature of science (Mayer, Sodian, Koerber, & Schwippert, 2014). Performance on the scientific reasoning test was not related to performance on a pencil and paper version of the Stroop (Mayer et al., 2014), although this was an unusual measure of semantic inhibition in that it was not computerised so may not have been able to capture inhibition with great enough sensitivity. Trial by trial information regarding RT was not captured, and it may be that an RT cost measure would better reflect inhibitory control ability. A study of 56 12- to 13-year-olds specifically considered the *learning* of biological concepts, as opposed to simply reasoning about them, through using a pre- and post-test design (Rhodes et al., 2014). This differs from the other studies described, in capturing change in performance over time following a period of teaching, rather than a single snapshot of performance that cannot adequately measure learning. Response inhibition as measured by the stop signal did not predict memory for the biological facts taught or understanding of information learnt in the practical session. Similarly, stop signal performance in 63 12- to 13-year-olds did not predict the learning of chemistry or performance in a school science exam (Rhodes et al., 2016). The inconsistencies across studies highlight that relatively little is currently known about the role of executive functions in science as compared to maths or reading (Tolmie et al., 2016). Nonetheless, the limited evidence that exists does suggest that inhibitory control is related to science performance at least in certain contexts.

#### 1.4.3.2 Maths

In the discipline of maths, overlapping strategy use in problem-solving demonstrates the maintenance of multiple concepts and theories over the course of learning. In depth examinations of strategy-use in 4- and 5-year-olds found that even when new, more sophisticated strategies had been learnt, children continued to use old strategies (Siegler, 1998; Siegler & Jenkins, 1989). The concurrent existence of multiple strategies suggests that inhibitory control is likely to be involved in maths reasoning from a young age, enabling selection of the best strategy through suppression of the alternatives.

There is also more direct evidence for the involvement of inhibitory control skills in maths. In a study that measured semantic inhibition through a Stroop task, performance in 3- to 6-year-olds was associated with scores on a standardised maths test in a sample of 58 children, and with magnitude comparison in a separate sample of 48 children (Merkley, Thompson, & Scerif, 2016). Similarly, a study of 209 11- to 14-year-olds found that maths achievement was correlated with both numerical and non-numerical semantic inhibitory control (Gilmore, Keeble, Richardson, & Cragg, 2015). A different maths study investigated the role of inhibitory control in the learning of a new algebra strategy for solving word problems in 157 14-year-olds (Khng & Lee, 2009). Better response (stop signal) and semantic (Stroop) inhibitory control predicted higher accuracy as well as fewer intrusions of the previous strategy (Khng & Lee, 2009), indicating that inhibitory control was used to suppress the previous strategy in effective maths reasoning. As in science, inhibitory control does not always relate to maths performance, as evidenced by a study where semantic inhibition as measured by both the Day/Night task and the Stroop did not predict school readiness for maths in a sample of 85 5- to 6-year-olds (Monette, Bigras, & Guay, 2011).

A meta-analysis of the association between inhibitory control and academic skills in 2- to 6-year-olds reported a modest overall association between inhibition and maths, and noted that inhibitory control was more associated with maths than literacy (Allan, Hume, Allan, Farrington, & Lonigan, 2014). The meta-analysis also highlighted the moderating impact of the type of task (the distinction between ‘hot’ and ‘cool’) and method of assessment (behavioural task or adult report) (Allan et al., 2014). However, the meta-analysis did not comment on the distinction between response and semantic inhibition. Overall, inhibitory control does seem to relate to maths performance.

#### 1.4.3.3 Counterintuitive reasoning

Inhibitory control is thought to have a specific role in counterintuitive science and maths reasoning, since the intuitive response to a problem needs to be suppressed in order to reach the correct answer (Dunbar et al., 2007; Houdé, 2000; Mareschal, 2016). Inhibitory control is thought to enable the suppression of any misleading or incorrect information that may interfere with the appropriate reasoning. This means that prior knowledge can be a barrier to the acquisition of new knowledge, since it may conflict with newly presented concepts or data, and thus influence learning (Heit, 1994).

A few studies have used priming paradigms to probe the role of inhibitory control in counterintuitive maths reasoning. In a study of 40 9-year-olds, participants performed better on a counterintuitive number conservation trial if they were primed to inhibit through the successful inhibition of an incongruent Stroop trial (Linzarini, Houdé, & Borst, 2015). Similarly, in a study of 20 10-year-olds and 20 young adults, participants performed better on a number conservation or class inclusion task when primed by a trial from the other task requiring the inhibition of a misleading strategy (Borst, Poirel, Pineau, Cassotti, & Houdé, 2013). Finally, in a study of 11-year-olds, 14-year-olds, and young adults, response times were slower when a probe problem with congruent relational term and arithmetic operation ("more than" > addition) followed a prime problem with incongruent relational term and arithmetic operation ("more than" > subtraction). This negative priming was interpreted as reflecting that successfully solving these arithmetic counterintuitive problems required the inhibition of an incorrect strategy (Lubin, Vidal, Lanoë, Houdé, & Borst, 2013).

In addition to the behavioural research linking inhibitory control to science and maths reasoning, there have been a number of neuroimaging studies aiming to discover more about the specific neural and cognitive processes underlying effective reasoning. These studies have generally shown that brain regions involved in inhibitory control are also involved in counterintuitive reasoning, although the research so far relates only to adults. In two related studies of 23 adults who were classed as either physics experts or novices, participants were presented with misconception-related electric circuits (Masson, Potvin, Riopel, & Brault Foisy, 2014) and falling objects (Brault Foisy, Potvin, Riopel, & Masson, 2015) inside an fMRI scanner. Experts who accurately judged the stimuli to be non-scientific showed increased activation in the dorsolateral and ventrolateral PFC (electric circuits: BA 9/45, falling objects: BA 47/46/10), with additional increased ACC (BA 32) and AG (BA 39) activation in the electric circuit study only. Since the PFC and ACC are known to be associated with inhibitory control, these activations were taken to reflect the experts' inhibition of misconceptions or naïve responses. In maths, an fMRI study of a perimeter misconception with 14 adults showed increased PFC (BA 47) activation in incongruent trials that were counterintuitive compared to congruent trials that were not counterintuitive (Stavy & Babai, 2010). As in the science studies, this suggests that participants may be using inhibitory control brain regions in order to suppress the incorrect answer.

A limit of these neuroimaging studies is the reliance on reverse inference, attributing activation in the PFC and ACC to inhibitory control mechanisms. No research has yet compared activation during counterintuitive reasoning to activation during inhibitory control within the same participants. This would be a better way of discovering whether or not inhibitory control really is contributing. It is also not possible to tell from these studies what kind of inhibition might be involved. Comparing neural activation of response and semantic inhibition to counterintuitive science and maths reasoning might reveal which type of inhibitory control is enabling suppression of incorrect theories.

Note that these proposed inhibitory mechanisms are not necessarily specific to science and maths. It is likely that the same cognitive and neural underpinnings apply to reasoning in any discipline that requires learning new knowledge or skills, particularly where these may be counterintuitive. Science and maths remain the topic of focus as they have been a key focus of the literature so far, and are considered particularly difficult subjects to learn (Lortie-Forgues et al., 2015; Zaitchik et al., 2014).

Overall the evidence suggests that both response and semantic inhibitory control contribute to science and maths reasoning, and that they may have a specific role in the inhibition of pre-existing beliefs, naïve theories, misleading perceptual-biases, and intuitive heuristics during counterintuitive reasoning. This differs from a more traditional understanding of learning, which describes the replacement, reorganisation, or restructuring of previously held knowledge (e.g. Posner, Strike, Hewson, & Gertzog, 1982; Vosniadou, 2007). The relative importance of response and semantic inhibition remains unknown. Response inhibition may have a role in enabling the physical stopping of providing a response, while semantic inhibition may be more related to suppressing a theory. There is a clear gap in the literature for research comparing the role of response and semantic inhibition in counterintuitive reasoning, and for establishing their shared neural underpinnings in adolescence. In addition, while the adult neuroimaging research has focussed on the repetition of a small set of problems, within adolescence it would be of particular interest to examine school-related problems covering science and maths curricula broadly. Focussing on educationally-relevant stimuli would enable greater links between research and educational practice.

## 1.5 Relational reasoning

The second cognitive ability to be considered in relation to science and maths is relational reasoning. Relational reasoning is the ability to consider relations between multiple mental representations (Crone et al., 2009) and to detect meaningful patterns in presented information (Alexander, Dumas, Grossnickle, List, & Firetto, 2016). It is considered one component of fluid reasoning, the broader cognitive ability to solve problems independently of knowledge (Ferrer, O'Hare, & Bunge, 2009). Relational reasoning differs from instinctive, spontaneous relational *thinking* in requiring intentional processing over a period of time (Alexander, 2016). There are different domains of relational reasoning, the most commonly studied of which is analogical reasoning. Analogical reasoning is the ability to observe similarities between items based on shared associations, and it is particularly important in the context of education as it is thought to underlie knowledge acquisition (Whitaker, Vendetti, Wendelken, & Bunge, 2017). This section will first consider the development of relational reasoning skills, then their relation to executive functions, next, their potential role in science and maths, and finally, links between relational reasoning and education.

### 1.5.1 Development of relational reasoning skills

Non-verbal analogical reasoning tasks are suitable for use with young children as simple pictures of familiar objects are typically presented. A series of studies have used scene analogy problems with children as young as 3 years (Richland, Morrison, & Holyoak, 2006). Within these problems, participants are asked to point to the object in a scene picture that is equivalent to the target object in a different scene picture. For instance, a scene might show a cat (target) chasing a mouse in one scene, and the second scene might show a boy (correct answer) chasing a girl. Within a series of these problems, the number of relations is manipulated (e.g. the addition of a dog chasing the cat and a woman chasing the boy), as is the presence of a distractor (e.g. a cat in the second scene). Performance improved with age across all configurations of the task in a sample of 68 3- to 14-year-olds (Richland et al., 2006). The 3- and 4-year-olds were strongly influenced by the presence of a distractor and relational complexity, with lowest performance where there was a distractor and two relations rather than one. A similar pattern was seen in the 6- to 7-year-olds, while the 13- to 14-year-olds were not affected by the distractor, but performance was lower with greater relational complexity. In a follow up study of 44 3- to 4- and 9- to 11-year-olds, the younger participants were again negatively influenced by the distractor and relational complexity, while the older participants showed no effect of distractor or relational complexity (Richland et al., 2006). The authors attributed this difference to either motivational factors, since the 13- to 14-year-olds were tested in a group while the 9- to 11-year-olds were tested individually, or to academic differences, since the 9- to 11-year-olds attended a school with higher achievement. Nonetheless, the studies showed overall improvement with age.

Other non-verbal analogical reasoning tasks have taken an A:B::C:? format. In this type of task, participants are usually shown two related pictures such as a dress (A) and a wardrobe (B), alongside a third picture such as a milk carton (C) and a question mark (e.g. Whitaker et al., 2017). The participant must choose the picture that is analogous to the wardrobe – in this case a fridge – that appears alongside a semantic lure (here, a cow), a perceptual lure (here, a clock that is similar in colour to the milk carton), and an unrelated lure (a tennis racket in this example). Ninety-five participants aged between 6 and 18 years completed this task inside an fMRI scanner (Whitaker et al., 2017). Large improvements in performance were seen between the ages of 6 and 10, with continued improvement until the age of 14 years where performance plateaued at almost ceiling level. Overall, the same network of frontal, parietal, and temporal regions were recruited across ages, indicating shared common processes through development. However, there was a positive correlation between age and activation in the left PFC (BA 47/45), which the authors argued reflected increased controlled semantic retrieval with age (Whitaker et al., 2017). This finding highlights that while fluid reasoning is considered a knowledge-free ability, analogical reasoning tasks of this kind which purport to measure a type of fluid reasoning (Ferrer et al., 2009), rely heavily on semantic knowledge since the stimuli used are not abstract. It is also likely that participants use inner speech to solve the problems, labelling the stimuli and the relation (e.g. "a dress goes in a wardrobe, so a milk carton goes in a fridge"), which would be influenced by verbal ability. Nonetheless, relational reasoning tasks that rely on knowledge remain an important area of investigation, since real-life problem-solving, for instance in science and maths, often requires the retrieval of knowledge.

A similar non-verbal analogical reasoning task was performed in an fMRI scanner by 33 participants who were grouped into 6- to 13-year-olds and 19- to 26-year-olds (Wright, Matlen, Baym, Ferrer, & Bunge, 2008). Children made more errors and had longer RTs than adults. Within the group of 6- to 13-year-olds there was greater activation of the left RLPFC (BA 10), left ventrolateral prefrontal cortex (VLPFC) (BA 45) and right SPL (BA 7) with age (Wright et al., 2008). There were no differences with age in the adult group, indicating that by the age of 19, the brain regions recruited for reasoning analogically in this task have matured. It is possible that an examination of older participants would have shown a decline, as observed in the inhibitory control studies. An interesting timecourse analysis showed that the RLPFC was recruited much later in the child group than the adult group, with children activating that region after response selection. The authors argued that this demonstrates more impulsive responding in children, as the region was not recruited in time to influence behavioural responses (Wright et al., 2008).

Further evidence of the role of the left RLPFC in the development of relational reasoning skills comes from a study of 37 11- to 30-year-olds, performing a non-verbal reasoning task with abstract rather than familiar stimuli (Dumontheil, Houlton, Christoff, & Blakemore, 2010). In this task, participants assessed whether items had the same shape or texture (1-relational problem, control condition) and whether two pairs of items changed along the same dimension (2-relational problem, relational condition). Activation in the left RLPFC (BA 46) in the 2-relational condition increased from early- to mid-adolescence, then decreased from mid-adolescence to adulthood (Dumontheil et al., 2010). This non-linear change in neural activation corroborated the results of their behavioural study that showed an increase in RT during mid-adolescence, suggesting that RLPFC activation reflects speed of neural processing rather than accuracy (Dumontheil et al., 2010). Overall the authors argued that these non-linear findings demonstrate the development of a number of different processes, including grey and white matter maturation and changing neurocognitive strategies.

Alexander and colleagues (Alexander et al., 2016) conceptualised relational reasoning in terms of four distinguishable categories based on an extensive analysis of the literature and validation through a series of studies. The first category, analogical reasoning, refers to similarity of relations, as described above. The second, anomalous reasoning, refers to discrepancy between items. Antinomous reasoning, the third category, refers to incompatibilities, and finally antithetical reasoning, the fourth category, refers to dichotomies. Sixty-one participants ranging from 5 to 17 years of age took part in semi-structured conversations which required them to reason out loud about different objects (Jablansky et al., 2015). All four categories of relational reasoning were observed across the age groups, with frequency of reasoning type in the following order: analogical (the most used), anomalous, antinomous, and then antithetical (the least used). The oldest children displayed antinomous and antithetical reasoning proportionally more than the younger children. The authors argued that analogy and antinomy reflect judgements of similarity and difference, maturing earlier than antinomy and antithesis which may be more complex forms of relational reasoning (Jablansky et al., 2015). This study is particularly noteworthy as it demonstrates changing uses of relational reasoning with age in a task that is more closely linked to school-based reasoning (since it involved reasoning about objects using vocabulary). However, the authors pointed out that the older children who were less likely to verbalise analogical and anomalous reasoning may have been using these processes without verbalising them.

These studies show that children are able to reason about relations from a young age. Perceptually and semantically distracting information can hinder reasoning ability, with distractors having less influence as participants reach adolescence. The change from reasoning based on surface features to reasoning based on relations is sometimes termed a relational shift (Gentner, 1988). While relational reasoning is typically considered a knowledge-free skill, knowledge is often a constraining factor on tasks that rely on an understanding of different items (Goswami, 1992; Leech, Mareschal, & Cooper, 2008), particularly in tasks for children that depict real objects rather than abstract stimuli. RLPFC appears to play an important role in the development of relational reasoning, although it should be noted that the RLPFC supports a range of cognitive processes which require retrieval and manipulation of abstract thoughts (Dumontheil, 2014). Relational reasoning may therefore rely on executive function skills, enabling the manipulation of information in mind, and the suppression of distracting information.

### 1.5.2 Relational reasoning and executive functions

A number of studies have investigated the association between relational reasoning abilities and executive functions explicitly. Richland and Burchinal (2013) investigated longitudinal associations between the two in a large study of 1,364 participants. The Day/Night task measured semantic inhibition when participants were 4 years old, and the Tower of Hanoi was given as a measure of general executive function at age 6 to 7 years. A verbal analogical reasoning task was administered at age 15 in the A:B::C:? format, for example “dog is to puppy as cat is to… ?”, with pictures to select from. Both executive function measures predicted analogical reasoning performance when vocabulary, likely an important contributing factor in this verbal task, was controlled for (Richland & Burchinal, 2013). The authors do not speculate on the mechanism of this association. It is possible that executive functions allowed the participants to inhibit any distracting information among the pictures presented, and to hold the problem in mind while integrating information. An alternative explanation is that better executive function skills helped those children to learn better, and the knowledge learnt helped in the analogical reasoning task which drew on semantic knowledge.

In a study of 71 young adults, the non-verbal Test of Relational Reasoning (TORR), measuring the four components of relational reasoning defined by Alexander and colleagues (Alexander et al., 2016) was administered alongside visuospatial and phonological working memory tasks (Grossnickle, Dumas, Alexander, & Baggetta, 2016). VS-WM, but not phonological working memory, was positively related to all subscales of the TORR, most strongly with analogy and anomaly. The results support the idea that the ability to hold visual information in mind is important in visual relational reasoning tasks. The authors also posited the possibility that those with better relational reasoning create better mental models and thus do not need to draw on their working memory so much (Grossnickle et al., 2016). According to this argument, the association with working memory may be down to shared reliance on mental models, without working memory directly influencing relational reasoning capacity.

One hundred and twenty-six 5- to 6-year olds were encouraged to look at the A:B part of a pictorial A:B::C:? task (Glady, French, & Thibaut, 2017). The aim was to help children focus their attention on the part of the task that might be overlooked when completing the task, and indeed performance improved when children were supported in this manner (Glady et al., 2017). The authors argued that children found it less of a struggle to inhibit C when A:B was highlighted. This finding suggests that inhibitory control enables suppression of the part of the problem presented that might lead to a semantic or surface-feature match rather than a relational match. This is a noteworthy addition to the field's understanding of the role of inhibitory control in relational reasoning, since others tend to focus on the need to ignore (or inhibit) distractors in the answer set presented (Vendetti, Matlen, Richland, & Bunge, 2015). Glady and colleagues (2017) also argued that in scene analogies (as described above), children might struggle to find the right answer due to the requirement to inhibit the target scene and focus on the initial scene. The authors called this the 'unbalanced attentional focus hypothesis'.

Relational reasoning ability appears to be constrained by executive function ability. In particular, working memory enables the holding and manipulation of task features in mind, while inhibitory control enables distracting information within the question and the answer set to be suppressed. Since these executive functions mature throughout childhood and adolescence, they likely limit the extent to which individuals can reason effectively in relational tasks during development. Given that executive functions have a role in science and maths, it is likely that executive functions and relational reasoning interact in influencing science and maths performance. While the interplay between relational reasoning, executive functions, and science and maths has not been studied, there are examples of research into relational reasoning in science and maths.

### 1.5.3 Relational reasoning in science and maths

Links between relational reasoning and science and maths have been investigated separately. As in the other literatures presented in this chapter, studies in maths are more developed than those in science. Nonetheless, the research in both science and maths has led to particular recommendations for education, which will be discussed at the end of this section. First, the association between relational reasoning and science will be considered, followed by the association between relational reasoning and maths.

#### 1.5.3.1 Science

In science, the association with relational reasoning has primarily been probed through attempts at teaching scientific concepts by analogical reasoning. In one example, 77 9- to 10-year-olds were taught geoscience concepts either by visual analogy or through use of a target picture (Matlen, Vosniadou, Jee, & Ptouchkina, 2009). In both cases, the children received the same analogy-enhanced text, but the 9-year-old children who were shown the visual analogy pictures learnt more than those who were not. While this suggests that the visual analogies were an especially helpful aid, the authors acknowledged that the increased performance may be down to better image processing since more images were seen by the visual analogy group (Matlen et al., 2009). The 10-year-old children did not show the same effect, which the authors attribute to better domain-general skills such as reading. This explanation is plausible, since the older children were perhaps more able to access and process the analogy text than the younger children. Nonetheless, the age gap between the children was very small, which suggests that this effect is not robust. A series of similar studies in adults found that visual analogies in geoscience helped learners form accurate conceptual representations (Jee et al., 2013). While these studies go some way in highlighting the role of relational reasoning in science learning, they are less helpful in determining the extent to which individual differences in relational reasoning are related to successful performance in science. They also focus on analogical reasoning, revealing less about different kinds of relational reasoning that might also play a role.

A different line of enquiry has investigated the skills involved in relational reasoning within science, which may reveal more about in-the-moment processes, as opposed to longer-term learning. Utterances from 18 novice and expert scientists during an anomalous meteorology task were compared (Trickett, Trafton, & Schunn, 2009).

Expert scientists used conceptual simulations when addressing anomalies, a strategy whereby a situation is mentally played out and the implications are considered, and they also used pure spatial transformation, whereby a mental image is manipulated to create a new one. Novices, on the other hand, used a less sophisticated strategy termed ‘comparison spatial transformation’, where two images are compared (Trickett et al., 2009). This research revealed differences between novice and expert scientists, suggesting that relational reasoning skills continue to develop through adulthood, becoming more specialised through practice. However, this is less useful in considering the role of relational reasoning skills in development through childhood and adolescence as applied to school. A strength of this research is in considering performance on tasks where relational reasoning is embedded within science.

#### 1.5.3.2 Maths

In maths, studies have sought to look at individual differences. A sample of 26 4- and 5-year-olds performed the Test of Analogical Reasoning in Children (TARC), which is a non-verbal test in the A:B::C:? format using coloured shapes of different sizes (White, Alexander, & Daugherty, 1998). A maths test required sorting sets, counting, making comparisons, and extending patterns. Performance on the TARC and maths test correlated positively, and 35% of the variance in analogical reasoning was explained by maths. The authors argued that maths requires individuals to associate objects and symbols to abstract concepts, and to recognise patterns; these skills are considered to require analogical reasoning (White et al., 1998). In a different maths study, 92 5- to 8-year-olds solved addition problems and undertook a range of relational reasoning tasks (Farrington-Flint, Canobi, Wood, & Faulkner, 2007). Better performance on the maths problems was associated with better relational reasoning in the older children only, indicating that early addition skills may be domain-specific, but that as children develop reasoning abilities these enhance their addition skills (Farrington-Flint et al., 2007). Following the creation of the TORR, Alexander and colleagues (Alexander et al., 2016) investigated its association with 13 problems from the maths Scholastic Assessment Test (SAT) in 42 adults. Scores on the TORR explained 13.4% of the variance in maths SAT scores (Alexander et al., 2016), further evidence that the two types of reasoning have common requirements, and indicating that the association holds until adulthood.

A study of fluid reasoning more broadly investigated the predictive role of reasoning in later maths (Green, Bunge, Briones Chiongbian, Barrow, & Ferrer, 2017). Sixty-nine participants aged 6 to 21 years undertook three measures of reasoning that required pattern completion, logic puzzles, and rule identification, combined to form a single measure of fluid reasoning. Fluid reasoning predicted performance in standardised maths tests one and a half and three years later, and moreover, fluid reasoning was a better predictor of maths performance than previous performance on a maths test (Green et al., 2017). The study provides support for a link between reasoning skills and maths ability throughout development in childhood and adolescence. The authors argued that a common factor in fluid reasoning and maths reasoning was relational reasoning, and that this ability may be the foundation for maths development (Green et al., 2017). Finally, the authors also suggested that while fluid reasoning is likely important for many disciplines, it may have a particularly important role in maths which requires the solving of novel problems.

#### 1.5.3.3 Education

The role of relational reasoning within science and maths *misconceptions*, particularly those relating to concepts learnt at school, is unknown, although Vendetti and colleagues (2015) argued that extending analogical reasoning in incorrect instances could lead to misconceptions. An understanding of the role of relational reasoning in education has led to recommendations for teaching and learning. Richland, Zur and Holyoak (2007) argued that since maths requires understanding abstract concepts which may be hard to comprehend, teaching by analogy (drawing parallels between similar problems) may support learning. This would be similar to the approach described in the science studies (Jee et al., 2013; Matlen et al., 2009). Vendetti and colleagues (2015) highlighted the importance of supporting relational reasoning within science, arguing that explicit explanation of comparisons is essential, as teachers may assume that analogous relations are obvious, when they are not to learners, as exemplified by studies showing improvement in analogy through the school years (Richland et al., 2006). Richland and Burchinal (2013) argued that improving executive functions, and inhibitory control in particular, might lead to improvements in school related to reasoning, although while efforts at executive function training were initially promising (Diamond & Lee, 2011), they have so far seen limited success (Melby-Lervåg & Hulme, 2013).

Relational reasoning skills develop throughout childhood and adolescence, at a time when executive function skills continue to develop and when understanding of science and maths content at school is of great importance for compulsory exams. While there are associations between these different skills, the precise interplay is unknown. Many studies investigating relational reasoning aim to reduce the role of knowledge, but knowledge is undoubtedly an important factor in science and maths reasoning. These are knowledge-dependant disciplines, but relational reasoning appears to be an important contributing factor. A more comprehensive understanding of the interplay of executive functions, relational reasoning, and science and maths reasoning could have important implications for learning and development in science and maths.

## 1.6 Thesis overview

In this section I will describe the research approach taken throughout the PhD, and then summarise each following chapter.

### 1.6.1 Research approach

The overarching aim of this thesis was to take an educational neuroscience approach to understand more about the cognitive and neural bases of science and maths reasoning in adolescence. The key dimensions of this approach were to:

a. Regularly meet with teachers to continually assess the educational relevance of the studies and findings.
b. Spend a substantial amount of time in school to ensure a rounded understanding of the typical science and maths classroom, and the school context more widely.
c. Use stimuli that are relevant to science and maths content that 11- to 15-year-olds frequently encounter and need to understand for exam success.
d. Adopt a suite of methods to address science and maths reasoning and learning at different levels of scientific understanding.
e. Feed results back to teachers and pupils who took part in the research to ensure that their participation saw rewards and that the research had impact.

Using this approach, and given the literature presented above, the specific aims of the research were to investigate the role of inhibitory control and relational reasoning in science and maths during adolescence. In particular, the thesis aimed to examine general science and maths reasoning, in addition to science and maths reasoning in the context of misconceptions, or counterintuitive concepts. Behavioural, classroom-based, and fMRI studies were used in order to capture different levels of description, and to link findings to real-world settings. The research also aimed to examine specific effects of inhibitory control and relational reasoning, by controlling for other contributing factors such as language ability and working memory.

### 1.6.2 Chapter summaries

In **Chapter 2** I summarise key aspects of the methods used in the experimental chapters. First I describe the creation of a novel science and maths misconceptions task that was designed specifically for use with 11- to 15-year-olds, drawing on the curricula for England. Second, I describe the selection of and experimental detail regarding the inhibitory control tasks used throughout the thesis. Third, I review literature pertaining to fMRI, briefly describing how it works, with a focus on the benefits and limitations of this approach, in particular as applied to young participants.

In **Chapter 3** I present the results of a first behavioural study, which sought to investigate the association between inhibitory control and counterintuitive science and maths reasoning. Note that the data was collected for my MSc dissertation, and the results presented here constitute a reanalysis of this data, and thus a more nuanced picture is presented than in the original analysis. It was hypothesised that adolescents with better inhibitory control as measured by response and semantic inhibition tasks would perform better on the science and maths misconceptions task, when controlling for age and general cognitive ability.

In **Chapter 4** I describe two classroom-based studies that aimed to establish the association between inhibitory control and the learning of new counterintuitive concepts in the classroom. **Study A** investigated the learning of two new physics concepts, and **study B** investigated the learning of a new maths concept. In both cases, the classroom teachers taught the new concepts, and learning was tracked one week after the lesson and one month later. It was hypothesised that those with better inhibitory control would perform better in the tests following the lessons in which they were taught the counterintuitive material.

In **Chapter 5** I present the first behavioural and neuroimaging results of an fMRI study. Participants completed the science and maths misconceptions task and two inhibitory control tasks (measuring response and semantic inhibition) inside the scanner. In contrast to previous research, inhibitory control brain activation was directly compared to counterintuitive reasoning brain activation to establish the extent of overlap. It was hypothesised that brain regions recruited during inhibitory control would also be recruited during counterintuitive concept reasoning.

In **Chapter 6** I examine the data pertaining to relational reasoning from the fMRI study. Exploratory analyses aimed to link verbal analogical and non-verbal relational reasoning to science and maths behaviour and brain data. It was hypothesised that those with better relational reasoning skills would perform better on the science and maths misconceptions task, and that brain activation would correlate with relational reasoning ability.

In **Chapter 7** I describe a second behavioural study, which aimed to establish the relative importance of inhibitory control and relational reasoning in science and maths performance. As in the previous studies, both response and semantic inhibitory control were measured, and verbal analogical reasoning was measured, along with a more comprehensive measure of relational reasoning. It was hypothesised that inhibitory control would explain unique variance in counterintuitive reasoning, and that relational reasoning would explain unique variance in general science and maths reasoning.

In **Chapter 8** I discuss the overall findings from each experimental chapter and describe what the studies have revealed about the cognitive and neural bases of science and maths reasoning in adolescence. I consider the ways in which I was able to meet my aims in terms of educational neuroscience approach, evaluating educational neuroscience as a discipline based on my experiences, and describing the challenges that I faced along the way. I describe the novel contributions of the results to our understanding of science and maths reasoning, and finally consider what the findings might mean for education.

The thesis now proceeds with the Methods section, which describes the key tasks used throughout the thesis, and reviews fMRI as a tool to examine cognitive and neural processes.

# Chapter 2 Methods

In this chapter I introduce aspects of the methods in the thesis that are relevant to several experimental chapters. First, I describe the ethics procedures and the selection of schools who provided participants. Second, I describe the design of a novel computerised science and maths task aiming to assess the ability to answer questions related to misconceptions. This task was first used for the behavioural study in **Chapter 3**, then modified for the fMRI study presented in **Chapters 5 and 6**, and modified further for the second behavioural study in **Chapter 7**. Third, I describe the two inhibitory control tasks used in all of the studies presented in the thesis to measure response and semantic inhibition. Fourth, I briefly present the use of fMRI as a neuroimaging tool able to study the developing brain, and discuss study design, analysis pathways, and the strengths and limitations of this method.

## 2.1 Ethics and school selection

All of the studies presented in this thesis underwent the strict ethical clearance procedures required by the local ethics committees (Birkbeck and/or UCL) to ensure the safety and wellbeing of participants. All information sheets and consent forms are provided in **Appendix 1**. The behavioural and classroom studies reported in **Chapters 3, 4, and 7**, received ethical clearance from Birkbeck. **Chapters 3 and 7** were opt-in studies, where parents received an information sheet, and provided written informed consent if they wished for their child to take part. A condition stated on the information sheet was that participants should not have any known neurological or developmental disorders. Parents confirmed on their consent form that their child had no known neurological or developmental disorders, and this was not formally checked. Participants aged 11 or 12 years verbally consented, while 13- to 15-year-olds provided written consent. **Chapter 4** was an opt-out study, where parents received an information sheet, and provided a written opt-out request form if they did not wish for their child to take part. There were no exclusionary criteria for this classroom study, since it took place during normal school class times and it was considered that nobody should be excluded. All participants provided written consent.

Schools were recruited in a non-systematic manner. Participants from **Chapter 3** were recruited from a school that was approached through networks. Participants from **Chapter 4** were recruited from a school that was a partner throughout my PhD, and were thus happy to be involved in a more intensive classroom study. In this case, the school had been involved in the project since its inception so were not recruited just for this study. Half of the participants from **Chapter 7** were recruited through the same school as in **Chapter 3**, and the other half were recruited through a personal contact who is a teacher at the school. Information relating to each school is reported within the relevant chapters.

The analyses reported in **Chapters 5 and 6** refer to a single study that received ethical clearance from UCL (due to use of the fMRI scanner on UCL premises). This was an opt-in study, where parents received an information sheet, and provided written informed consent if they wished for their child to take part. Again, parents confirmed on their consent form that their child had no known neurological or developmental disorders. To ensure the safety of participants, parents were present for the testing procedure (in a nearby room), and provided screening on the day. All participants provided written consent. Participants from **Chapters 5 and 6** were recruited through opportunity sampling and were thus from a range of schools: some participants were from the schools described above, others were recruited from schools that were approached due to proximity to UCL, and others were friends of those who had already taken part in the study.

## 2.2 Science and maths misconceptions task

A novel computerised science and maths misconceptions task was designed to measure adolescents' ability to give the correct (although counterintuitive) answer when faced with problems known to be associated with misconceptions. The aim was to cover a broad set of concepts across science and maths curricula for 11- to 14-year-olds (from Key Stage 3 for England, covering school Years 7 to 9) to make the task relevant to education. In order to pull together a variety of problems, the literature was searched for examples of misconceptions that are present to some extent within this age range. The sources of the problems included books that had compiled the scientific literature on this topic (Driver, Squires, Rushworth, & Wood-Robinson, 2015; Ryan & Williams, 2007; Stavy & Tirosh, 2000), and discussions with teachers. In addition, study guides (Parsons, 2014; Parsons & Gannon, 2014) and the Key Stage 3 curricula (Department for Education, 2013a, 2013b) were consulted to ensure that only misconceptions related to these curricula were included. The difficulty of the problems varied in order to ensure that younger and older participants would find some problems easy and some challenging; this ensured a range in accuracy levels to explore individual differences. The oldest participants in the study were 15-years-old and in Year 10, so should have covered all of the topics addressed by the problems at school in Years 7 to 9. While the task was designed according to topics from science and maths curricula, it is possible that participants in Year 10 may not have covered each problem presented, for example, if they had missed school due to illness. Nonetheless, it was expected that they would have caught up on any missed material in preparation for exams. It is also possible that participants in the same Year group had not covered the same aspects of the curriculum at the time of testing.

The problems found in the literature were modified to fit into the problem-set format that was adopted for these studies. The format was adapted from adult neuroimaging studies into misconceptions in science (Brault Foisy, Potvin, Riopel, & Masson, 2015; Masson, Potvin, Riopel, & Brault Foisy, 2014, see **Introduction** section 1.3.3.3 for a summary of their findings). Each of these adult studies focussed on a single science misconception. Participants had to judge whether the stimuli were scientific or not (**Figure 2.1**). Some stimuli were scientific, in that they were correct (e.g. showing balls of different sizes falling at the same speed), while others were non-scientific, i.e. incorrect and in line with the misconception often held by individuals (e.g. showing a larger ball falling faster than a smaller ball). The science experts recruited for the study judged both scientific and non-scientific stimuli correctly, while the science novices judged them incorrectly. There were also control problems, which were non-scientific, covering a similar aspect of science, but not related to the misconception (e.g. a smaller ball falling faster than a larger ball). Both experts and novices in the studies accurately judged the control stimuli to be non-scientific.

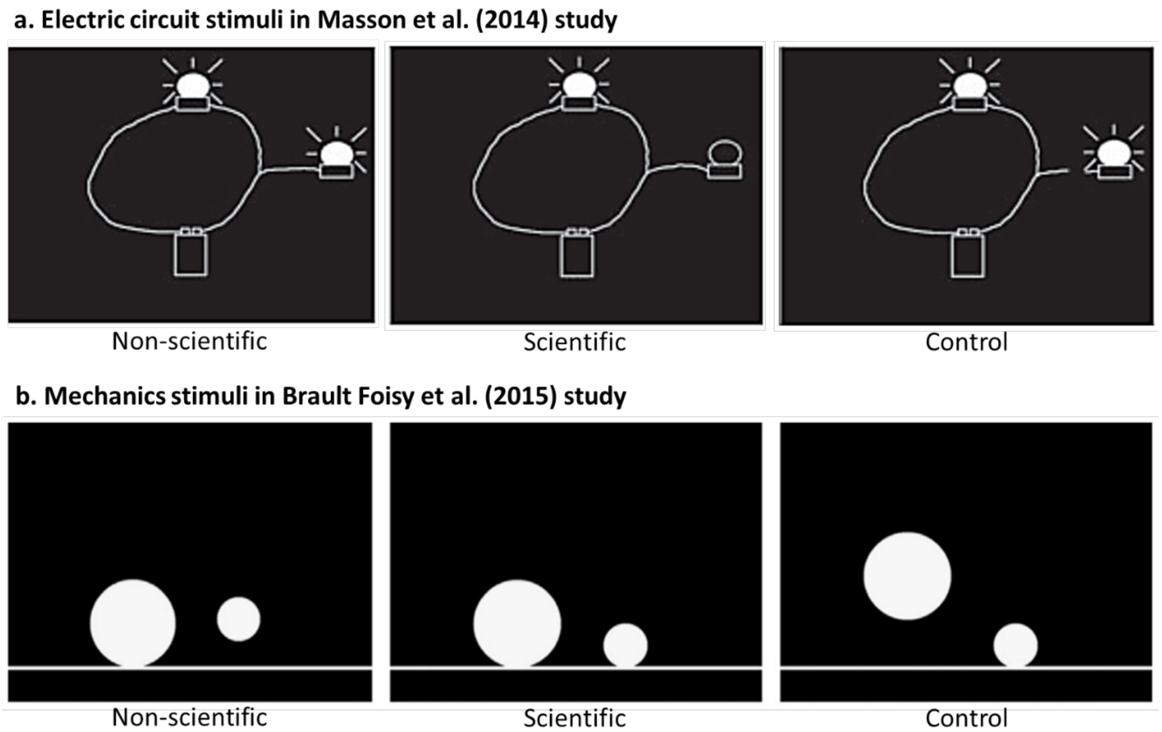



**Figure 2.1.** Stimuli presented inside the fMRI scanner for (a) the electric circuit study (Masson et al., 2014) and (b) the mechanics study (Brault Foisy et al., 2015). Electric circuit stimuli were still images, and mechanics stimuli were short videos showing the balls falling. Permission to reproduce these images has been granted by John Wiley and Sons, and Elsevier respectively.

The version of the science and maths misconceptions task developed for the study presented in **Chapter 3** consisted of 48 problem-sets each addressing different misconceptions. This differs from the adult neuroimaging studies, which concerned just two misconceptions that were shown repeatedly. It was considered that showing the same problems might lead participants to overthink problems and change their response, and that findings would not be generalizable to the range of counterintuitive concepts encountered at school. Another significant change from the adult work was to include two control problems per misconception, so that the control question was not always unscientific. This meant that the problem-types (misconception and control) were matched, and that there were equal numbers of control problems and misconception problems. Control problems controlled for knowledge and interest in the topics, and task-general factors such as processing speed and attention. Efforts were made to ensure that misconception and control problems were matched on statement length, positive versus negative wording, and terminology. Each problem-set thus contained a scientific problem relating to a misconception (*Misconception-True*), a non-scientific problem relating to the same misconception (*Misconception-False*), a scientific problem where no counterintuitive reasoning was required (*Control-True*), and a non-scientific problem where no counterintuitive reasoning was required (*Control-False*) (**Figure 2.2**). Rather than a single image (Masson et al., 2014) or short video (Brault Foisy et al., 2015), here participants were shown more varied stimuli including sentences and equations. This variety meant that the problems were more similar to those shown at school. For each problem, participants were asked to judge whether the stimuli shown were correct (i.e. true, scientific) or incorrect (i.e. false, unscientific) by clicking one of two buttons with their dominant hand.

Twenty-six of the 48 problem-sets were science-based. Biology topics included living organisms, cells, inheritance, genetics, and plants (8 problem-sets). Chemistry topics included pollution, atoms, pure and impure substances, heating, and melting (7 problem-sets). Physics topics included force, the solar system, electricity, gravity, waves, and temperature (11 problem-sets). Maths topics included fractions, decimals, angles, algebra, shape, transformations, statistics, probability, and graphs (22 problem-sets). One hundred and twenty-six problems were accompanied by images (biology: 28/32; chemistry: 18/28; physics: 36/44; maths: 44/88), which were sometimes essential for the problem, sometimes provided further explanation, and sometimes simply relevant to keep the task engaging. In all cases the *Misconception-True/False* and *Control-True/False* problems were matched as well as possible in terms of text and images presented, ensuring all participants saw similar stimuli across conditions, and allowing the comparison of participants' performance across conditions.

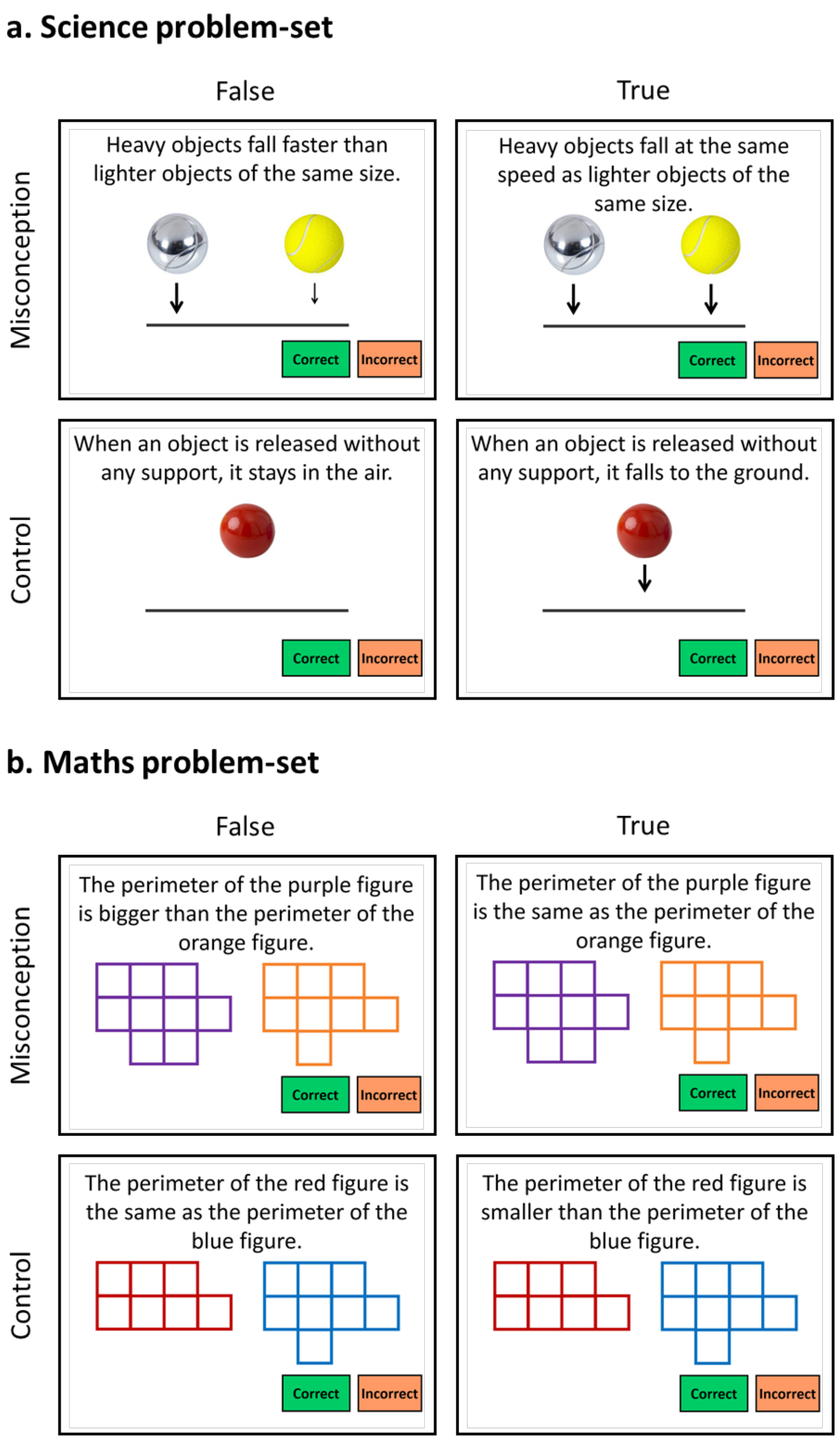



**Figure 2.2.** Example (a) science and (b) maths problem-sets. Note that text and image size has been increased to enhance legibility here.

With 48 problem-sets each containing four problems, there were a total of 192 problems. Each participant was shown one misconception problem per set, and one control problem per set, totalling 96 problems per participant, with an equal distribution of True and False problems. Each participant thus completed 24 of each of the four problem types (*Misconception-False, Misconception-True, Control-False, Control-True*). Two sequences of 96 trials were created by randomly distributing problems such that each sequence contained one misconception trial and one control trial from each problem-set. Science, maths, misconception, and control trials were all mixed within the sequence. Two further sequences were created by reversing the presentation order of the original sequences. Participants were assigned to a sequence pseudo-randomly, ensuring each age group contained the range of sequences.

Participants were instructed to respond as quickly as possible, and stimuli, along with Correct and Incorrect 'buttons' (**Figure 2.2**), remained on the screen until a response was made. There was no time limit for participants' responses. The task was programmed in Cogent (www.vislab.ucl.ac.uk/cogent_graphics.php) running in MATLAB (The MathWorks, Inc., Natick, MA). Self-timed breaks were inserted ¼, ½, and ¾ of the way through the task. The task lasted approximately 11 min on average.

Changes were made to the science and maths misconceptions task across the different studies. **Chapter 3** contains the first iteration of the task, and as such was presented exactly as described above. The task was not used in **Chapter 4**. The task was adapted as follows for the fMRI study described in **Chapters 5 and 6**.

The problem-sets were changed in some cases, based on discussions with teachers about how related the concepts were to school and how clear each problem was: wording was tweaked and pictures were made clearer. Some problems were dropped and new ones added in order to have an even split between disciplines: there were now 24 science problem-sets and 24 maths problem-sets, and within science there were eight biology problem-sets, eight chemistry problem-sets, and eight physics problem-sets. Each trial now had a grey background rather than white background, so that the screen was not too bright in the dark scanner room. The task had an event-related design, and on each trial participants again had to judge whether the statement was true or false. This time there were four response options that appeared on the screen in the following order: definitely true, probably true, probably false, and definitely false (**Figure 2.3**). It was explained that the definitely/probably distinction referred to the participant's confidence in their response. Responses were made through two button boxes, and the index and middle fingers of both hands rested on four response buttons. Each participant saw 24 of each of the four problem types, as before, with one misconception problem and one control problem from each problem-set.

Unlike in **Chapter 3**, a time limit was added to each trial, to constrain the duration of each scanning run. Based on the behavioural data from **Chapter 3**, a maximum of 12 s was allowed for a response to be made on each trial, and if a response had not been made within 9 s, a red border appeared around the response options on the screen to prompt a response (**Figure 2.3**). Each trial lasted a total of 16 s, and the remaining time following a response was filled with a fixation cross on a third of trials or a simple arrows task on two thirds of trials. In the arrows task, participants pressed the left or right key with their index fingers according to the direction of arrows on the screen. This constituted an active baseline, and was used to encourage participants to stop thinking about the science and maths task, allowing the BOLD signal to decrease, but participants to remain engaged.

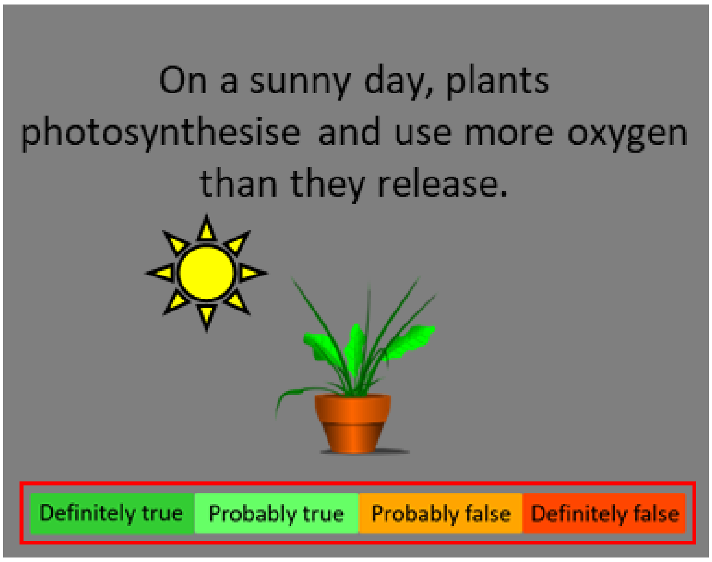



**Figure 2.3.** Example science problem with red border to prompt participants to answer as quickly as possible, which appeared when there were just 3 s remaining.

While science and maths stimuli were intermixed in the first version of the task, the fMRI task consisted of four runs: two runs of science and two runs of maths, alternating, and the starting run was counterbalanced across participants. Science and maths were separated to maximise the extent to which science- or maths-related regions were engaged during the run as a whole. Each run started with an instruction screen that told the participant whether they would be given science or maths questions. A fixation cross appeared in the centre of the screen for 10 s at the start and end of each run, and 15 s in the middle of each run. Eight new fixed trial orders were created, and participants were assigned to one of the orders according to their school year group and gender, ensuring representation of each order across year groups and genders. Within each run of 24 trials, each participant saw six problems of each trial type. Short breaks were given between each run, so that participants had a chance to relax and to say if they were unsure about anything or uncomfortable inside the scanner. The task lasted a total of approximately 30 min.

In **Chapter 7**, which presents the second behavioural study, the science and maths problems were intermixed again and the sequences were prepared as in **Chapter 3**. The problems used were those from **Chapters 5 and 6**, with further small tweaks to improve stimuli, and a 12 s time limit per trial (with red border prompt at 9 s) as in **Chapters 5 and 6** to ensure that the maximum time spent on this task was under 20 min. This was important for the study in **Chapter 7** because one school that was hosting the research imposed a strict time limit for the involvement of each participant in order to reduce disruption to the school day. Four response options were given, as in **Chapters 5 and 6**, but there was no arrows task and there were no fixation blocks. Self-timed breaks were given at three time points as in **Chapter 3**. All versions of the stimuli are freely available online (osf.io/ytcwk).

Accuracy and RT were measured in all versions of the task. It was anticipated that accuracy would be higher and RTs faster for control trials compared to misconception trials. Mean RTs are reported for all trials. While it is typical in cognitive tests to only include RTs for correct responses, in this case incorrect trials are of interest since it was expected that participants would get many of the misconception problems wrong. Although it is possible that on some trials participants may have been distracted and not responding to the task as anticipated, RTs for incorrect trials were expected to mostly reflect the time spent reasoning about a counterintuitive (or control) concept, even if the resulting answer is incorrect.

## 2.3 Inhibitory control tasks

Two inhibitory control tasks were selected in order to measure both response and semantic inhibition, and were used, in slightly different formats, throughout the PhD. From the previous research linking inhibitory control to science and maths, it is unclear whether response or semantic inhibition is more important in counterintuitive reasoning, and no previous studies have focussed on this distinction (see **Introduction** section 1.3.3.3).

### 2.3.1 Response inhibition

The Go/No-Go task measured response inhibition, and was adapted from Watanabe and colleagues (2002) and Humphrey and Dumontheil (2016). This task was chosen in part as is has shown improvement with age from 12 to 15 years (Humphrey & Dumontheil, 2016), and as such would be appropriate to investigate age effects in the 11- to 15-year-olds studied in these experiments. In addition, the task differs from other Go/No-Go tasks in requiring a response choice on Go trials, which makes the task more challenging and engaging. It was therefore hoped that participants would remain engaged throughout the task. Both simple and complex versions of the task were included to investigate the possibility that response inhibition in the context of a higher cognitive load, namely a 1-back working memory load, would be more associated with complex science and maths reasoning, where information may need to be maintained and manipulated while the answer is worked out, compared to inhibitory control within a simpler task.

Familiarisation and practice phases consisted of ten trials for each version of the Go/No-Go. In the familiarisation phase, coloured squares (green squares in the simple task; yellow or blue squares in the complex task) appeared on the left or right of the screen, and participants pressed the corresponding key, using the index or middle finger of their dominant hand, to indicate the location of the square. This was repeated until participants made two errors or fewer. In the practice phase, No-Go trials, where the response must be withheld, were introduced (red squares in the simple task; a blue square following a yellow square in the complex task) (**Figure 2.4**). The practice phase was repeated until participants made no more than one No-Go error out of three No-Go trials. Test phases followed, and included 80 trials, with 25% No-Go trials, and a self-timed break half-way through. The square's location and fixation duration were pseudo-randomised so that for every set of ten trials, 50% of stimuli appeared on the left, and fixation duration was randomly chosen from a uniform distribution between 600 and 800 ms. Responses were not recorded for the first 100 ms after stimulus presentation because such a quick response would most likely relate to the previous trial. There were never two No-Go trials in a row. The task was programmed in Cogent (www.vislab.ucl.ac.uk/cogent_graphics.php) running in MATLAB (The MathWorks, Inc., Natick, MA). The task lasted a total of 6 min on average.

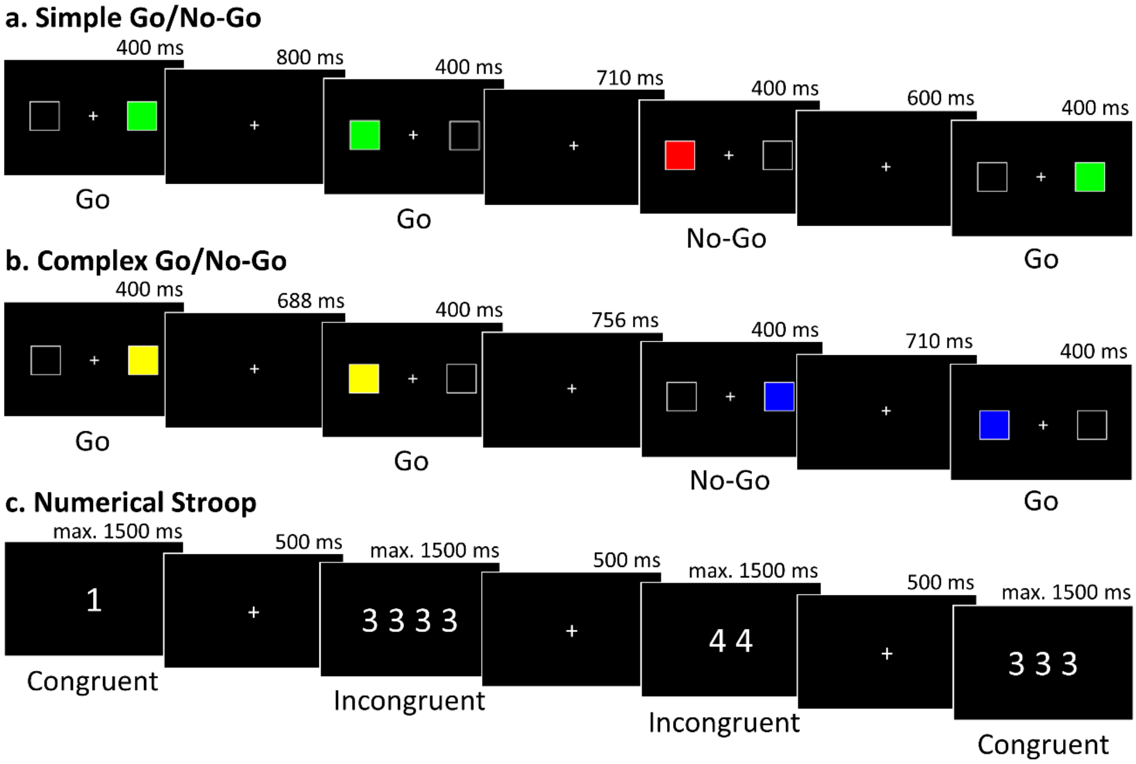



**Figure 2.4**. Example time course of events in the (a) simple Go/No-Go, (b) complex Go/No-Go, and (c) numerical Stroop.

The Go/No-Go was adapted over time to suit the needs of different studies. In **Chapter 3**, the task was administered as described above, in a one-to-one testing session where the experimenter could ensure the participant understood the task. In **Chapter 4**, the task was reprogrammed in PHP and JavaScript, running in an Ubuntu web server. The task was administered in a group testing session and so a description of the task was given to the group rather than individually (see further information in section 4.3.1.3.2).

In **Chapters 5 and 6**, the task was adapted for the fMRI study. The task had a block design containing three block types. *Go blocks* contained only Go trials, where a beige square appeared on the left or right of the screen, and participants used their left or right index finger on the button boxes to indicate the location of the square. *Simple Go/No-Go blocks* contained 50% Go trials, where participants indicated the location of the beige square, and 50% No-Go trials where no response was given to a blue square. *Complex Go/No-Go blocks* contained 50% Go trials and 50% No-Go trials, with a one-back component to tax working memory. Squares were pink and yellow, and a response indicating the square's position was required when the colour of the square matched the previous trial (Go trial). No response was required when the colour of the square did not match the previous trial (No-Go trial). The percentage of No-Go trials was increased from 25 to 50 so as to increase the difference between Go and Go/No-Go blocks, in line with previous research (e.g. Tamm, Menon, & Reiss, 2002). While this optimised the task for the scanner, in increasing the inhibitory control required during the block, it made the task less optimal for behavioural testing. The task was now easier for participants because No-Go trials were more common, and thus less difficult to inhibit. It was anticipated that this would have an impact on the behavioural results, such that accuracy would be higher. Differences in findings across studies might therefore reflect modifications in the task rather than genuine differences between participants. The 50% No-Go trials meant that two No-Go trials could now appear in succession. Stimuli remained on the screen for 400 ms, followed by a fixation cross in the centre of the screen, which remained on the screen between 600 and 800 ms. A fixation cross appeared in the centre of the screen for 10 s at the start and end of each run, and 15 s in the middle of each run. The task was performed in a single run, with four repeats of each block type. Each block contained 20 trials, with location, interstimulus interval (ISI), and trial type pseudo-randomised per block. Each block lasted 22 s and the task lasted approximately 6 min in total. In **Chapter 7**, for the second behavioural study, the Go/No-Go consisted of only the complex version, in order to reduce the length of the task, and was administered as in **Chapter 3**. Accuracy and RT were recorded in all iterations. RT analyses in this task considered only correct Go trials, since it is the speed of correct responses that is of interest.

### 2.3.2 Semantic inhibition

The numerical Stroop task, adapted from Khng and Lee (2014) provided a computerised measure of semantic inhibition (**Figure 2.4**). The Stroop has shown improvement with age in the age range of interest here (Comalli et al., 1962; Leon-Carrion et al., 2004), while the flanker for example, has shown performance to mature by age 10 (Cragg, 2016; Ridderinkhof & van der Molen, 1995). The Stroop therefore appears to be a more sensitive measure of semantic inhibition. This numerical version of the Stroop task allowed a more intuitive mapping between numbers and keys as compared to non-ordinal stimuli, such as colours, in the traditional colour-word Stroop task (Stroop, 1935).

The aim in this task was to press the key corresponding to the number of elements on the screen (between one and four). In the familiarisation phase, the elements were asterisks. The four number keys were located in the centre of the keyboard, labelled with stickers, and participants were instructed to use the index and middle fingers of their left and right hands, such that the fingers corresponded to numbers one to four from left to right. The familiarisation phase was repeated until participants responded correctly on 11 out of 12 trials. This conservative threshold ensured that mapping between fingers, keys and responses had been achieved. In the first practice phase, consisting of 24 trials, the stimuli were single digit numbers, and participants continued to respond to the number of elements (digits) shown. These were all congruent trials, where the digit matched the number of elements (e.g., ‘1’ or ‘3 3 3’), and there was no performance threshold. The second practice phase contained only incongruent trials, where the digit did not match the number of elements (e.g., ‘4 4’ or ‘1 1 1 1’). The second practice contained just two trials, and was repeated (with changing stimuli) until participants got both correct. This ensured that participants understood the task but did not get too much practice.

The test phase contained 50% congruent and 50% incongruent trials presented in four blocks of 24 trials. The same pseudo-random trial order was used across participants, such that each block contained one of each possible incongruent trial and three of each congruent trial. There was a self-timed break between each block. Accuracy and RT were recorded, and the task took 4.5 min on average. The task was programmed in Cogent (www.vislab.ucl.ac.uk/cogent_graphics.php) running in MATLAB (The MathWorks, Inc., Natick, MA).

The numerical Stroop was also adapted over time. In **Chapter 3**, the task was administered as described above, with the experimenter administering the task in a one-to-one testing session. In **Chapter 4**, as with the Go/No-Go, the task was carried out in a group testing session, and so was reprogrammed in PHP and JavaScript, running in an Ubuntu web server. In **Chapters 5 and 6**, the task was adapted for the fMRI study using a block design. *Congruent blocks* contained only congruent trials, and *Mixed blocks* contained 50% congruent trials and 50% incongruent trials. The same four fingers were used to respond using the button boxes. Stimuli remained on the screen until a response was made or a maximum of 1.1 s had passed. Each trial lasted 1.5 s and the remainder of the trial was filled with a fixation cross in the centre of the screen. Blocks alternated and there were a total of five blocks of each type, with a fixed trial order across participants. Before the start of a block, participants were shown an instruction screen for 2 s indicating which block type they would be completing. There was 10 s of fixation at the beginning and end of a run, with 15 s fixation roughly in the middle of a run. Accuracy and RT were recorded and the task lasted approximately 5 min. In **Chapter 7**, for the second behavioural study, the numerical Stroop was administered as in **Chapter 3** but with half as many trials (48 trials) presented in two blocks, in order to reduce the length of the task. RT analyses for all iterations of the task only considered correct trials.

## 2.4 Functional magnetic resonance imaging (fMRI)

Blood oxygenation level-dependent (BOLD) fMRI was used in **Chapters 5 and 6**, alongside behavioural data, to reveal more about the cognitive and neural underpinnings of science and maths reasoning. BOLD fMRI is a non-invasive method of imaging the functioning human brain, developed in the 1990s, with very good spatial resolution of a few millimetres. Here I will provide a brief overview of how BOLD fMRI works, consider experimental design using fMRI, describe analysis pathways, and discuss the issues that arise when using fMRI with young participants.

### 2.4.1 Principles of BOLD fMRI

An MRI scanner produces a strong magnetic field, causing some of the protons in the brain's hydrogen atoms to align with the field (Dick, Lloyd-Fox, Blasi, Elwell, & Mills, 2013). The protons spin constantly, and at constant speed, but they are not in phase with each other. The beaming of a pulse of radiofrequency (RF) causes the protons to all spin in phase, pointing in the same direction, perpendicular to the main magnet of the scanner (Dick et al., 2013). This causes an electric current in the scanner's surrounding coil of wire which is detected by the scanner. The protons eventually dephase and realign with the magnetic field, and so the current reduces and disappears. The speed with which the current reduces depends on the chemicals in the surrounding tissues, and this is how tissues are differentiated (Dick et al., 2013). The use of three perpendicular magnetic coils in the MRI scanner allows spatial information to be recorded so that an image of the brain can be created, comprised of many slices.

BOLD imaging relies on the detection of transient local changes in oxygenated haemoglobin in the brain's blood. During the activation of neurons in a given region, there is increased demand of oxygen from the blood, causing an expansion of arterioles and capillaries (Turner, 2016). This gives rise to an overall increase in oxygen (despite the concurrent consumption of oxygen). Since oxygenated haemoglobin is less magnetic than deoxygenated haemoglobin, the proton dephasing process takes longer (deoxygenated haemoglobin causes dephasing to happen more quickly). Therefore, when there has been increased demand for oxygen, the electric current measured by the scanner is slower to reduce, and thus the BOLD signal shows a net increase. The BOLD signal reflects changes in both blood volume and oxygen extraction (Turner, 2016), and as such is not a direct measure of neuronal activation.

## 2.4.2 Experimental design using fMRI

There are two commonly used designs for fMRI, each with associated advantages and disadvantages. The first, most used design, is the block design. In a block design, there are alternating blocks of experimental conditions. For example, in the numerical Stroop (see section 2.2.2) there were blocks of congruent trials only, alternating with mixed blocks of congruent and incongruent trials. Within the block design, a subtractive method is used in order to isolate activation associated with a specific cognitive function (Corr, 2006). In the case of the numerical Stroop, congruent blocks were subtracted from mixed blocks, to remove any common activation associated with congruent trials, leaving activation that is specific to incongruent trials. The block design is considered to have superior statistical power and to be simpler to implement than the alternative (Chee, Venkatraman, Westphal, & Siong, 2003). However, the design is also limited because it cannot distinguish between trial types within a block, which may be particularly important in the context of correct versus incorrect trials (Petersen & Dubis, 2012). Where in behavioural data, error trials might be discarded, this is not possible in fMRI studies using a block design, although if errors are infrequent an event-related regressor can remove brain activity related to errors. In addition, averaging over different trial types (such as in the mixed blocks in the numerical Stroop) may result in a cancelling effect that does not reflect positive and negative neural responses (Petersen & Dubis, 2012). Finally, participant awareness of the block they are currently completing may introduce confounds due to predictability (Chee et al., 2003).

Event-related designs present mixed trials, allowing random presentation of stimuli (Chee et al., 2003). This design was adopted for the science and maths misconceptions task (see section 2.1). Misconception and control trials were mixed, which meant that participants would less likely be aware of any difference in trial type. A strength of this approach more generally is that specific trials can also be sorted or selected after the experiment has taken place, allowing trials with error responses to be dropped, for instance (Chee et al., 2003). Modelling of individual trials occurs when analysing event-related data, using a similar subtractive approach, whereby individual trials are subtracted rather than full blocks. This allows for additional specificity, as models may include only the time spent on the trial, as opposed to the trial and the inter-trial duration. For instance, in the science and maths misconceptions task, control trials were subtracted from misconception trials, such that activation specific to misconception trials was leftover, and only the time spent responding was included in the model. However, due to the decreased signal-to-noise ratio, event-related designs are less powerful and less efficient than block designs (Petersen & Dubis, 2012). Null events are required even if not modelled explicitly (The FIL methods group, 2015), and the inclusion of such events can make the task feel very long and less engaging, which may be a particular issue for younger participants.

### 2.4.3 Analysis of BOLD fMRI data

The first stage of analysis for either experimental design is preprocessing, which ensures that voxels can be compared over time by reducing variability in the data not associated with the task. This occurs in four steps. The first step is realignment, which aims to correct for head motion inside the scanner by realigning each scan to a reference scan (Friston, 2007) such as the mean image as in **Chapter 5** (see section 5.3.5). Since motion can still account for much of the variance after realignment, for example because movement occurred between slice acquisition (Friston, 2007), the second step in preprocessing is adjusting for these movement effects. A composite measure of head motion across the six realignment estimates is calculated, giving a framewise displacement (FD) estimate (Siegel et al., 2014). Individual scans with a FD estimate greater than 0.9 mm are modelled out of the general linear model (GLM) by including a regressor of no interest for each censored volume. The third step in preprocessing is spatial normalisation, which matches each scan to a common template (such as the Montreal Neurological Institute, MNI registered maps used in **Chapter 5**) of standard anatomical space (Friston, 2007). This step is necessary because of the large individual differences in brain size and shape. In order to assess effects over groups of participants, each brain must be normalised. Note that other approaches can be used, such as in retinotopic mapping of visual cortex, where conventional normalisation is considered inappropriate because of the particularly large variability in size, shape and position of these areas (Henriksson, Karvonen, Salminen-Vaparanta, Railo, & Vanni, 2012). Finally, the scans are spatially smoothed. Spatial smoothing reduces spatial precision but makes errors more normally distributed, thus ensuring validity of parametric tests (Friston, 2007). Standard practice smooths images by 8 mm (as in **Chapter 5**), allowing for group averaging and statistics across normalised images (Turner, 2016).

Following the preprocessing procedure, there are two further stages, based on the GLM, modelled at two levels (The FIL methods group, 2015). The first level implements within-participant analyses, by modelling regressors which represent the different conditions or trial types of the task, including fixation periods where relevant, censored volumes that have been excluded due to movement, and the session mean. Contrasts are then calculated for each participant (for example, Misconception > Control, see **Chapter 5** section 5.3.6), and then entered into the second level model (The FIL methods group, 2015). It is from the second level analyses that maps are created and inferences drawn. Thresholds are put into place to determine which voxels and clusters are to be reported, and these are determined by the researcher. The chance of false positives due to multiple comparisons can be managed through applying a family-wise error (FWE) corrected threshold, either at the voxel level, which is a conservative approach similar to Bonferroni, or at the cluster level, typically at the default $p < .05$ threshold, with an uncorrected voxel level threshold of $p < .001$. There remains debate in the field regarding whether or not FWE correcting is sufficient, and what the best thresholds are to adequately control for false positives (Eklund, Nichols, & Knutsson, 2016; Flandin & Friston, 2017). The researcher can also determine an extent threshold, which produces maps of clusters of a minimum number of voxels.

In this thesis, SPM12 (www.fil.ion.ucl.ac.uk/spm/ software/spm12/) was used to preprocess and analyse the data. MNI coordinates were used to define regions. Region labelling was completed with Automatic Anatomical Labeling (AAL) (Tzourio-Mazoyer et al., 2002). BA labelling was completed with MRIcron (Rorden & Brett, 2000). MarsBaR (Brett, Anton, Valabregue, & Poline, 2002) was used to calculate average parameter estimates per participant in different clusters.

### 2.4.4 FMRI with young participants

As a non-invasive method of brain imaging with no known side effects, fMRI is well suited to non-clinical research, and particularly to use with children; repeated MRI has been shown to have no negative impact on children (Holland et al., 2014). A key challenge in using fMRI with young participants is the critical requirement that participants are still while inside the scanner. This can be difficult for children and adolescents, but it also introduces a possible confound because the extent of movement may be different across age groups (Dick et al., 2013). It is also possible that the amount of movement would be related to the participant’s inhibitory control ability, whereby those with poorer inhibition would find it harder to stop their desire to move inside the scanner. This confound would be particularly problematic for the fMRI study described in this thesis, where inhibitory control is a key variable of interest. Where there is too much movement, scans are removed from analyses, and where a participant has moved too much within a single session, this whole session is dropped from the analyses (see **Chapter 5** section 5.3.5). Another issue with imaging young participants is that anatomical atlases may not be appropriate for the wide variety in head sizes; using standard spaces based on young adult participants, as is typical, may lead to misclassification of brain tissue and spurious age differences, for instance (Richards, Sanchez, Phillips-Meek, & Xie, 2016).

An additional age-related issue is that attention and task compliance may differ according to age, but this cannot easily be detected when participants are inside the scanner. For example, a child who is particularly struggling to pay attention in the unusual scanner environment may not be attending to the task as much as an older participant who is more comfortable inside the scanner and able to focus on the task. This could mean that the task being performed is essentially different for the two, and yet this is not taken account of in the analyses. However, age differences in the contrasts of interest are examined (although it is not necessarily clear what age effects reflect), and participants are monitored as well as possible from the console room. The fMRI operators can speak to the participant regularly to ensure they are as comfortable as possible, and cameras inside the scanner ensure operators can observe behavioural cues.

A final challenge with regards to conducting fMRI with young participants is recruitment. The requirement that participants have no metal in their body can be a particular challenge when the target age group is adolescents, many of whom have braces at this age. In addition, parents can sometimes have concerns about the participation of their child, and therefore not agree to their child's participation, even if the child is keen. Finally, the loud scanner or the clinical atmosphere can be off-putting to young participants who may not have been fully aware of what the study entailed until arrival at the scanner suite, and who may then decide not to take part. This possibility can be reduced by describing the procedure in detail to the parent or guardian, and providing videos for the participant to watch in advance of making their decision. Some labs also have pretend scanners, where participants can get used to the environment before going into the main scanner, and practice lying still to the required degree. The youngest participants tested for this thesis were 11 years of age, so this process was not used.

Having described the features of the main tasks, and considered the use of fMRI, the thesis will now move on to the experimental chapters. The first experimental chapter describes a behavioural study seeking to establish to the association between science and maths reasoning and inhibitory control.

# Chapter 3 Behavioural study 1: Inhibitory control in science and maths

## 3.1 Overview

When reasoning about counterintuitive concepts, inhibitory control is thought to enable the suppression of incorrect concepts. This study investigated the association between inhibitory control and counterintuitive science and maths reasoning in adolescents (*N* = 90, 11-15 years). The novel science and maths task measured reasoning about counterintuitive concepts, where misconceptions may be held, that are part of the school curriculum. Both response and semantic inhibition were associated with counterintuitive science and maths reasoning, when controlling for age, general cognitive ability and performance in control science and maths trials. Better response inhibition was associated with longer reaction times (RTs) in counterintuitive trials, while better semantic inhibition was associated with higher accuracy in counterintuitive trials. This novel finding suggests that different aspects of inhibitory control may offer unique contributions to reasoning about misconceptions during adolescence and provides support for the hypothesis that inhibitory control plays a role in science and maths reasoning.

## 3.2 Introduction

According to the traditional view of learning, prior knowledge is replaced or reorganised when new information is learnt (Posner et al., 1982; Vosniadou, 2007). A more recent theory suggests that in fact prior knowledge remains, and must be inhibited, along with any misleading perceptual cues or naïve theories, in order to reason effectively in science and maths (Dunbar et al., 2007; Houdé, 2000; Mareschal, 2016). Inhibitory control requirements are thought to be especially great during counterintuitive reasoning, where intuitive responses may interfere with reasoning. Brain regions associated with inhibitory control are recruited in adults when reasoning about misconceptions (Brault Foisy et al., 2015; Masson et al., 2014), while the mechanisms are not yet well understood in adolescents. Behavioural research has linked individual differences in inhibitory control to performance in science and maths in adolescence (Gilmore et al., 2015; Khng & Lee, 2009) but there has been no examination of its association with reasoning about counterintuitive concepts specifically.

An understanding of the mechanisms underlying counterintuitive reasoning during adolescence is important because adolescents are faced with increasingly complex science and maths concepts (some of which can be counterintuitive) through compulsory school curricula (Department for Education, 2013b, 2013a). This is also a time when inhibitory control abilities and their neural correlates undergo protracted development (Humphrey & Dumontheil, 2016; Ordaz et al., 2013). It is unclear whether response or semantic inhibition is most important for effective reasoning about counterintuitive concepts in adolescence, since studies tend to focus on one measure of inhibitory control. This distinction is important in considering how best to inform teaching and learning practices. If response inhibition is most important, it may be that encouraging physical stopping of a response would help most, while an approach that increases consideration of different possible answers might help most if semantic inhibition is most important.

The current study therefore aimed to investigate the association between inhibitory control and counterintuitive science and maths reasoning in 11- to 15-year-olds, roughly corresponding to the Key Stage 3 curriculum. The youngest participants were in Year 7 which is the first year of secondary school and the start of the Key Stage 3 curriculum. The oldest participants were in Year 10 and had completed the Key Stage 3 curriculum. The problems presented could therefore relate to Key Stage 3 and be suitable for all participants, with improvements in performance expected with age.

It was hypothesised that better semantic inhibitory control, evidenced by less interference effect on accuracy and RT in the numerical Stroop task, would allow participants to better solve the conflict between the intuitive response and the correct answer, and that they would therefore show more accurate and faster responses on misconception problems, relative to control problems, which were designed to not involve misconceptions but target similar science and maths topics. It was also hypothesised that better response inhibition, evidenced by higher accuracy in simple and complex No-Go trials, would be associated with better misconception performance, by limiting impulsive responses. It was hypothesised that these associations would be specific to counterintuitive reasoning, with lower associations between science and maths performance in control trials and inhibitory control. There were no hypotheses as to which type of inhibitory control would be most predictive of misconception performance, so this was examined in an exploratory manner. As science and maths problem-solving typically requires the maintenance of some information in working memory, and as misconceptions in particular may elicit competition between, and comparison of, intuitive and counterintuitive responses in working memory, it was finally hypothesised that performance in complex No-Go trials would show a greater association with science and maths misconception performance than performance in simple No-Go trials.

# 3.3 Methods

## 3.3.1 Participants

Ninety pupils with no known neurological or developmental disorders, from an English state secondary school where most students are from minority ethnic heritages, and the proportion of free school meals (determined by parental income-related benefits) is well above average, took part. Letters were sent to parents of 11- to 15-year-olds (in Years 7 to 10), inviting their children participate. Written informed parental consent was obtained, where parents confirmed that their children had no known neurological or developmental disorders. Participants aged 11 or 12 years verbally consented, while 13- to 15-year-olds provided written consent, in accordance with the guidelines of the local ethics committee, which approved the study. For simplicity of reporting, age groups are referred to according to the mean age of the group (for example, 12y refers to 12-year-olds, the Year 7 participants whose ages ranged between 11.75 and 12.67).

## 3.3.2 Tasks

### 3.3.2.1 Wechsler Abbreviated Scale of Intelligence (WASI)

The Vocabulary and Matrix Reasoning subtests of the WASI-II (Wechsler, 2011) were administered using the stimulus book, to control for the contribution of general cognitive ability to science and maths performance (**Table 3.1**). The Vocabulary subtest requires participants to explain the meaning of words, while the Matrix Reasoning subtest requires participants to choose the picture that completes a pattern.

**Table 3.1.** Participant characteristics. Age groups did not differ in WASI Vocabulary, *p*'s > .2.

| | | **Age (years)** | | **Girls : Boys** | **WASI Vocabulary** | **WASI Matrix Reasoning** |
|---|---|---|---|---|---|---|
| **Age group** | ***n*** | ***M (SD)*** | ***Range*** | ***n*** | ***M (SD)*** | ***M (SD)*** |
| 12y | 25 | 12.14 | 11.75-12.67 | 13:12 | 33.01 (3.51) | 19.00 (3.46) |
| 13y | 25 | 13.26 | 12.75-13.75 | 17:8 | 33.12 (3.69) | 18.12 (3.14) |
| 14y | 21 | 14.32 | 13.92-14.75 | 9:12 | 33.71 (5.02) | 19.67 (2.92) |
| 15y | 19 | 15.21 | 14.75-15.75 | 12:7 | 35.26 (3.21) | 18.32 (4.06) |

#### 3.3.2.2 Science and maths misconceptions

The science and maths misconceptions task was administered on a laptop (see Methods section 2.1 for details). Participants were presented with science and maths statements, mostly accompanied with a picture, and were asked to indicate whether the statement was correct or incorrect. Half were misconception problems, and half were control problems, where no counterintuitive reasoning was anticipated. In addition, half of the problems were true and half were false. Note that it was anticipated that the true problems might be easier to answer correctly. No practice was given but participants were shown an example screen containing one control problem that was not included in the main trial.

#### 3.3.2.3 Inhibitory control

The simple and complex Go/No-Go and numerical Stroop were administered on a laptop (see Methods section 2.2 for details). The Go/No-Go, measuring response inhibition, required participants to press the left or right key in response to Go trials, and to withhold their response for No-Go trials. In the complex version, a small working memory load was introduced as the previous trial had to be kept in mind in order to respond correctly. In the numerical Stroop, participants pressed a key to indicate how many numbers were shown on the screen, not responding to the number represented by the digits themselves. In congruent trials, the number matched the digit (e.g. ‘2 2’), and in incongruent trials they did not match (e.g. ‘3 3 3 3’).

### 3.3.3 Procedure

Participants were tested in a quiet space in school for approximately 45 min during the school day. The experimenter described each computerised task, emphasising that responses should be given as quickly and accurately as possible. The tasks were performed in the following order: simple Go/No-Go, complex Go/No-Go, numerical Stroop, science and maths misconceptions, WASI Vocabulary, and WASI Matrix Reasoning. Participants were given no results, and no rewards for taking part, and it was explained that their responses would remain anonymous and independent of school assessments.

### 3.3.4 Statistical analysis

Examination of boxplots across tasks showed outliers, so exclusionary criteria were put in place before analysis commenced. Participants whose mean accuracy or RT was further than ±3.29 standard deviations (*SD*s) away from the group mean were excluded from analyses of the task on which they were an outlier, as standardised scores outside that range are cause for concern (Field, 2012). Effect sizes are reported as partial eta squared ($\eta_p^2$). Main effects of Age group were followed up with three planned tests assessing differences between 12y and 15y, 13y and 15y, and 14y and 15y, since the greatest differences were anticipated in comparison to the oldest group who had covered all concepts.

In the science and maths misconceptions task analysis, three participants were excluded due to low accuracy (one 13y) or slow RT (one 12y, one 13y), leaving a final $N$ = 87 participants. Two (Trial type: control, misconception) x two (Discipline: science, maths) x two (Statement type: true, false) x four (Age group: 12y, 13y, 14y, 15y) mixed model repeated measures analyses of variance (ANOVAs) were performed on accuracy and RT.

Three participants were excluded from the simple Go/No-Go analysis because of low accuracy (one 12y, one 13y) or high RT (one 12y) leaving a final $N$ = 87 participants. Two participants were excluded from the complex Go/No-Go analysis, because of low accuracy (one 12y) or an inability to pass the practice (one 12y), leaving a final $N$ = 88 participants. Four participants were excluded from the Stroop analysis because of low accuracy (one 12y, one 13y) or because they were unable to perform the two-handed task due to a hand injury (one 13y, one 15y), leaving a final $N$ = 86 participants. Two (Trial type: Go, No-Go or congruent, incongruent) x four (Age group: 12y, 13y, 14y, 15y) mixed model repeated measures ANOVAs were run on accuracy scores in each of the three tasks and on RT in the Stroop task. One-way ANOVAs examined the effect of Age group (12y, 13y, 14y, 15y) on Go RT in the simple and complex Go/No-Go tasks separately.

Participants excluded from any individual task analysis were also excluded from the regression analyses (final *n*s: 12y: $n$ = 20, 13y: $n$ = 22, 14y: $n$ = 21, 15y: $n$ = 18), leaving a total $N$ = 81 participants. Correlations were run between the variables of interest to examine collinearity and assess associations between measures across the whole sample. Hierarchical multiple regressions first investigated the contribution of inhibitory control to general science and maths performance, and then investigated whether inhibitory control variables could explain individual differences in science and maths misconception accuracy and RT specifically, while accounting for performance in control trials.

The first regression models included science and maths overall accuracy and RT as dependent variables. The control variables were inserted using the enter function in block 1: age in months, and WASI Vocabulary and Matrix Reasoning raw scores. These variables were expected to have an influence on the outcome measure, but were not the primary predictive variables of interest. Raw WASI scores were entered rather than standardised scores so that scores were directly comparable across ages. Standardised scores were not included since age was controlled for separately through entering age in months, and the inspection of standardised scores was not of specific interest in this study of participants with no known disorders. Go/No-Go variables were entered stepwise in block 2: simple No-Go accuracy, complex No-Go accuracy, simple Go accuracy, complex Go accuracy, simple Go RT, complex Go RT. Stroop variables were entered stepwise in block 3: accuracy cost (congruent minus incongruent), RT cost (incongruent minus congruent), congruent accuracy, and congruent RT.

The second regression models investigated misconception accuracy and RT. Again the control variables were inserted using the enter function in block 1: age in months, WASI Vocabulary and Matrix Reasoning raw scores, and this time including science and maths control performance (accuracy or RT). The same Go/No-Go variables were entered stepwise in block 2, and Stroop variables were entered stepwise in block 3.

Inclusion of separate Go/No-Go and Stroop blocks allowed for investigation of variance explained individually by response and semantic inhibition. Stepwise entry and the inclusion of variables that do not necessarily reflect inhibition (such as Go accuracy or congruent Stroop RT) enabled examination of the possibility that general processing speed or accuracy alone were the most important predictors of performance, rather than inhibition per se.

Follow up exploratory regressions were run on science and maths separately where there were significant inhibitory control predictors, to examine possible discipline-specific effects and to explore whether directions of association were consistent. All follow up models included the control variables and the inhibitory control variables identified in the science and maths combined regressions, using the enter method.

## 3.4 Results

### 3.4.1 Science and maths misconceptions

In line with the design of this task, participants tended to give the correct answer in control trials, with a mean accuracy of 82.2%, while they made more errors on misconception trials, where the mean accuracy was 54.7%. While this is close to chance performance (50%), the range of mean accuracy on individual misconception problems suggests that participants did not consistently guess across all problems, demonstrating that participants answered correctly more often on some trials than others (**Figure 3.1**). This indicates that the accuracy in misconception trials is not attributable to chance performance (guesses) on all problems.

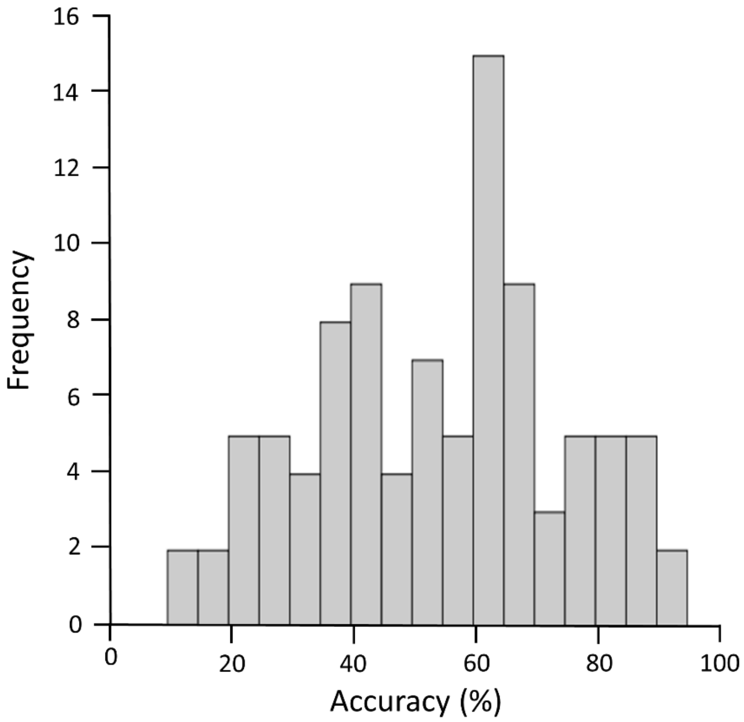



**Figure 3.1.** Histogram of mean accuracy in the 96 science and maths misconception problems.

A two (Trial type: control, misconception) x two (Discipline: science, maths) x two (Statement type: true, false) x four (Age group: 12y, 13y, 14y, 15y) mixed model repeated measures ANOVA performed on accuracy showed main effects of Trial type and Statement type, with greater accuracy in control compared to misconception trials, and true compared to false statements (**Table 3.2)**. There was no main effect of Discipline.

**Table 3.2.** Accuracy and RT estimated marginal means in the science and maths misconceptions task.

| | | Accuracy (%) | RT (ms) |
|---|---|---|---|
| | | ***M (SE)*** | ***M (SE)*** |
| **Main effects** | | | |
| *Trial type* | | $F(1, 83) = 816.73$, $p < .001$, $\eta_p^2 = .908$ | $F(1, 83) = 310.32$, $p < .001$, $\eta_p^2 = .789$ |
| Control | | 82.2 (0.8) | 5156 (134) |
| Misconception | | 54.7 (0.9) | 6683 (190) |
| *Discipline* | | n.s., $p = .367$ | $F(1, 83) = 55.73$, $p < .001$, $\eta_p^2 = .402$ |
| Science | | 68.1 (0.9) | 5598 (147) |
| Maths | | 68.9 (0.7) | 6240 (180) |
| *Statement type* | | $F(1, 83) = 38.64$, $p < .001$, $\eta_p^2 = .318$ | $F(1, 83) = 5.26$, $p = .024$, $\eta_p^2 = .060$ |
| True | | 72.5 (0.8) | 5837 (158) |
| False | | 64.4 (1.1) | 6002 (168) |
| *Age group* | | $F(3, 83) = 5.61$, $p = .001$, $\eta_p^2 = .169$ | n.s., $p = .631$ |
| 12y | | 65.3 (1.3) | 6149 (301) |
| 13y | | 66.5 (1.3) | 5856 (307) |
| 14y | | 69.2 (1.4) | 6073 (322) |
| 15y | | 72.8 (1.5) | 5601 (338) |
| **Interaction effects** | | | |
| *Trial type* | *Statement type* | $F(1, 83) = 11,48$, $p = .001$, $\eta_p^2 = .121$ | n.s., $p = .076$ |
| Control | True | 84.8 (0.9) | 4986 (136) |
| | False | 79.6 (1.1) | 5327 (153) |
| Misconception | True | 60.2 (1.1) | 6688 (199) |
| | False | 49.3 (1.4) | 6678 (204) |
| *Discipline* | *Statement type* | $F(1, 83) = 67.73$, $p < .001$, $\eta_p^2 = .449$ | $F(1, 83) = 14.15$, $p < .001$, $\eta_p^2 = .146$ |
| Science | True | 75.8 (1.1) | 5406 (146) |
| | False | 60.3 (1.4) | 5791 (159) |
| Maths | True | 69.3 (0.9) | 6268 (181) |
| | False | 68.5 (1.1) | 6214 (193) |

These main effects were modulated by a significant interaction between Trial type and Statement type (**Table 3.2**), which was followed up with two repeated measures

ANOVAs on control and misconception accuracy. The interaction was attributable to less difference in accuracy between true and false statements in control trials, $F(1,83) = 11.33$, $p = .001$, $\eta_p^2 = .120$, compared to misconception trials, $F(1,83) = 28.57$, $p < .001$, $\eta_p^2 = .256$. There was an additional significant interaction between Discipline and Statement type, whereby the difference between true and false statements was significant for science trials, $F(1,83) = 84.75$, $p < .001$, $\eta_p^2 = .505$, but not maths trials, $p = .629$ (**Table 3.2**). There was a main effect of Age group on accuracy, $F(3, 83) = 5.61$, $p = .001$, $\eta_p^2 = .169$. Follow-up planned comparisons revealed significant differences between 12y and 15y, $p < .001$, 13y and 15y, $p = .002$, and marginal differences between 14y and 15y, $p = .077$, each of which demonstrated increasing accuracy with age (**Table 3.2**).

There was a significant interaction between Discipline and Age group $F(3, 83) = 3.68$, $p = .015$, $\eta_p^2 = .117$. Follow-up repeated measures ANOVAs performed separately in each Discipline showed a significant effect of Age group for science trials $F(3, 83) = 4.95$, $p = .003$, $\eta_p^2 = .152$ and maths trials $F(3, 83) = 5.15$, $p = .003$, $\eta_p^2 = .157$. Bonferroni-corrected post-hoc comparisons (**Figure 3.2**) revealed significant increases in accuracy between 12y and 15y in science, $p = .043$, and maths, $p = .002$, and between 13y and 15y in science, $p = .002$.

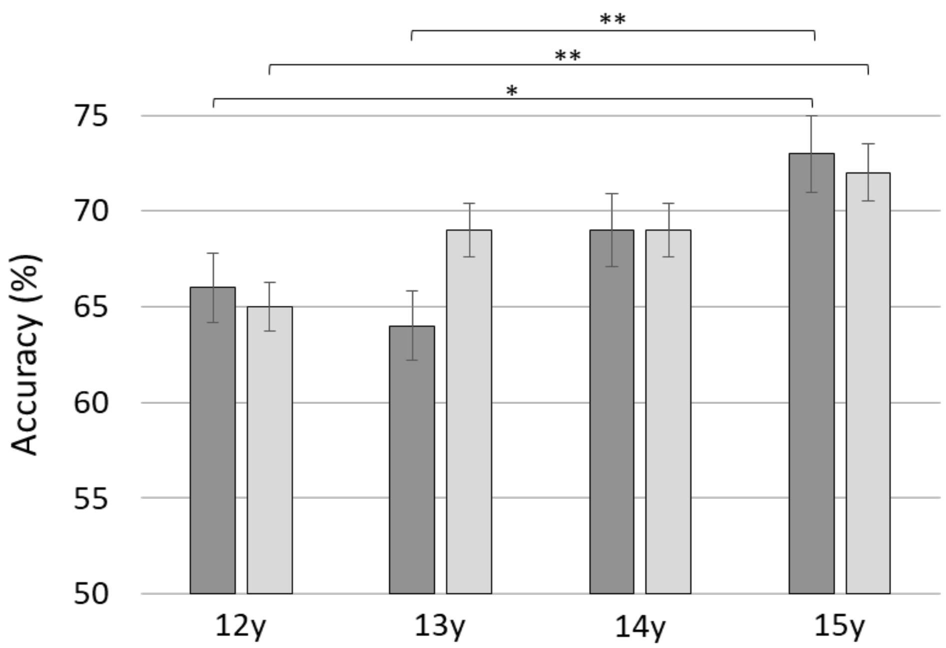



**Figure 3.2.** Accuracy estimated marginal means in science and maths trials by age. Dark bars represent science, light bars represent maths. *,** indicate $p < .05$ and $p < .01$ respectively.

The same ANOVA was performed on RT. There were main effects of Trial type, Discipline, and Statement type, with longer RTs in maths compared to science trials, misconception compared to control trials, and false compared to true trials (**Table 3.2**). There was no main effect of Age group on RT. As with the accuracy analyses, there was a significant interaction between Discipline and Statement type, which was followed up with two further repeated measures ANOVAs. There was a significant difference between true and false statements for science, $F(1,83) = 20.50$, $p < .001$, $\eta_p^2 = .198$, with longer RTs for false trials (**Table 3.2**), while there was no significant difference between true and false trials for maths, $p = .589$. No other interaction was significant, *p*'s > .1.

In summary, the main finding of interest was lower accuracy and longer RTs in misconception compared to control trials. This is consistent with the hypothesis and design of the paradigm, since it was anticipated that intuitive responses would be incorrect and reasoning would take longer in misconception trials. Lower accuracy and longer RTs were also found for false statements compared to true statements in science, but not maths. Maths RTs were longer than science RTs overall. Finally, improved performance with age was reflected in accuracy only and more prolonged in science than maths.

### 3.4.2 Inhibitory control

Mixed repeated measures ANOVAs performed on accuracy in the Go/No-Go tasks revealed a main effect of Trial type for both the simple and complex tasks, with higher accuracy for Go trials than No-Go trials (**Table 3.3**). On average, 53% of Go errors were omissions. There was a main effect of Age group on accuracy for the simple Go/No-Go task only, and planned post-hoc comparisons revealed a significant difference between 12y and 15y, $p = .029$ (other *p*'s > .70).

Similarly, one-way ANOVAs revealed a marginal effect of Age group on RT in the simple task only (**Table 3.3**). Planned post-hoc comparisons showed a significant difference between 12y and 15y, $p = .020$, and between 13y and 15y, $p = .023$, but only marginally between 14y and 15y, $p = .090$. In all cases, RTs were faster in the older age group.

In the numerical Stroop task, the repeated measures ANOVAs revealed a significant main effect of Trial type for both accuracy and RT (**Table 3.3**) with greater accuracy and faster RTs for congruent trials than incongruent trials. There was also a main effect of Age group on accuracy, but not RT. Post-hoc comparisons showed significantly poorer accuracy at 12y than 15y only, $p = .013$ (all other *p*'s > .92). The interaction between Trial type and Age group was not significant, $p = .622$.

**Table 3.3.** Accuracy and RT estimated marginal means in the inhibitory control tasks. Note that RTs are for correct trials only, therefore there are no RTs for No-Go trials.

| | **Simple Go/No-Go** | **Complex Go/No-Go** | **Numerical Stroop** |
|---|---|---|---|
| | ***M (SE)*** | ***M (SE)*** | ***M (SE)*** |
| **Accuracy (%)** | | | |
| *Trial type* | $F(1, 83) = 93.37$, $p < .001$, $\eta_p^2 = .529$ | $F(1, 84) = 183.31$, $p < .001$, $\eta_p^2 = .686$ | $F(1, 82) = 224.29$, $p < .001$, $\eta_p^2 = .732$ |
| Go/Congruent | 96.4 (0.4) | 84.5 (1.0) | 96.1 (0.4) |
| No-Go/Incongruent | 84.7 (1.3) | 53.6 (2.0) | 80.1 (1.2) |
| *Age group* | $F(3, 83) = 3.16$, $p = .029$, $\eta_p^2 = .102$ | n.s, $p = .725$ | $F(3, 82) = 3.58$, $p = .017$, $\eta_p^2 = .116$ |
| 12y | 86.7 (1.4) | 69.2 (2.2) | 84.3 (1.3) |
| 13y | 91.5 (1.4) | 67.7 (2.1) | 89.2 (1.3) |
| 14y | 92.4 (1.5) | 68.2 (2.3) | 89.5 (1.4) |
| 15y | 91.4 (1.6) | 71.3 (2.5) | 89.4 (1.5) |
| **RT (ms)** | | | |
| *Trial type* | - | - | $F(1, 82) = 426.67$, $p < .001$, $\eta_p^2 = .839$ |
| Go/Congruent | 346 (3) | 400 (6) | 671 (9) |
| No-Go/Incongruent | - | - | 779 (11) |
| *Age group* | $F(3, 83) = 2.36$, $p = .078$ | n.s., $p = .530$ | n.s., $p = .122$ |
| 12y | 354 (7) | 400 (11) | 745 (18) |
| 13y | 353 (7) | 399 (10) | 752 (19) |
| 14y | 348 (7) | 413 (11) | 700 (20) |
| 15y | 330 (7) | 388 (12) | 703 (21) |

In summary, No-Go errors (commissions) were more common than Go errors (omissions and incorrect side judgements) in both Go/No-Go tasks. Accuracy in the Stroop task was higher and correct responses faster in congruent than incongruent trials. There were age effects for accuracy on the simple Go/No-Go task, and the Stroop, with better performance in the oldest age group compared to the youngest age groups.

### 3.4.3 Regression analyses

Correlations between the variables of interest were examined (**Table 3.4**) and assumptions regarding multicollinearity were met.

**Table 3.4.** Pearson correlation coefficients of regression variables for science and maths combined. Statistically significant (two-tailed) correlations are highlighted in bold; [a,b,c] indicate $p$ <.05, $p$ < .01 and $p$ < .001 respectively.

| **Variable** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Misconception accuracy | | | | | | | | | | | | | | | | |
| 2. Misconception RT | .04 | | | | | | | | | | | | | | | |
| 3. Control accuracy | **.39**[c] | .22 | | | | | | | | | | | | | | |
| 4. Control RT | -.01 | **.89**[c] | .05 | | | | | | | | | | | | | |
| 5. Age (months) | **.38**[c] | -.11 | **.27**[a] | -.16 | | | | | | | | | | | | |
| 6. WASI Vocabulary | **.31**[b] | -.04 | **.51**[c] | -.14 | **.21**[a] | | | | | | | | | | | |
| 7. WASI Matrix Reasoning | .13 | **.21**[a] | .21 | .14 | -.02 | .15 | | | | | | | | | | |
| 8. Simple No-Go accuracy | .17 | **-.24**[a] | .02 | -.18 | .21 | -.01 | -.20 | | | | | | | | | |
| 9. Complex No-Go accuracy | .03 | .09 | .07 | .01 | -.04 | .09 | -.09 | .18 | | | | | | | | |
| 10. Simple Go accuracy | .13 | -.14 | .14 | -.12 | .15 | .06 | .00 | **.48**[c] | .10 | | | | | | | |
| 11. Complex Go accuracy | **.24**[a] | -.04 | .12 | -.08 | .18 | .06 | .13 | .00 | -.09 | .10 | | | | | | |
| 12. Simple Go RT | -.13 | **.24**[a] | **-.24**[a] | **.35**[b] | **-.33**[b] | -.16 | **-.22**[a] | .11 | .07 | -.19 | **-.21**[a] | | | | | |
| 13. Complex Go RT | -.04 | **.22**[a] | -.10 | **.31**[b] | -.07 | -.15 | -.10 | **.28**[b] | .13 | .07 | **-.23**[a] | **.66**[c] | | | | |
| 14. Stroop accuracy cost | .03 | -.03 | -.04 | -.05 | -.04 | -.12 | .06 | **-.25**[a] | -.19 | -.03 | .13 | **.28**[b] | -.21 | | | |
| 15. Stroop RT cost | -.16 | .06 | -.05 | .09 | .06 | -.09 | -.05 | .14 | -.18 | -.05 | -.07 | .15 | .06 | -.11 | | |
| 16. Stroop congruent accuracy | .20 | -.10 | **.23**[a] | -.08 | **.42**[c] | .05 | -.07 | **.48**[c] | .10 | **.36**[c] | -.01 | .11 | **.24**[a] | **-.27**[a] | **-.29**[b] | |
| 17. Stroop congruent RT | -.06 | .14 | -.14 | **.22**[a] | **-.23**[a] | -.15 | -.13 | -.02 | -.04 | **-.23**[a] | -.12 | **.56**[c] | **.44**[c] | **-.26**[a] | .01 | .05 |

An initial hierarchical multiple regression (**Table 3.5**) investigated whether inhibitory control measures could account for variance in science and maths accuracy overall. Model 1 was significant, explaining 34% of the variance, with age and WASI Vocabulary both significant predictors of accuracy. No inhibitory control variables were selected by a second model. The second hierarchical multiple regression (**Table 3.5**) examined whether inhibitory control could account for variance in science and maths RT overall. Model 2a was not significant, while model 2b was significant, explaining 14% of the variance, with both WASI Matrix Reasoning and Simple Go RT significant predictors. Simple Go RT uniquely predicted 10% of the variance in science and maths RT. Since this measure is thought to reflect general RT as opposed to inhibitory control per se, this association was not followed up.

**Table 3.5.** Regression models for science and maths overall performance. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science and maths accuracy | | | | |
| Model 1<br>$F(3, 77) = 13.06$,<br>$p < .001$, $R^2 = 34\%$ | Constant | | 1.87 | .064 |
| | **Age (months)** | **.30** | **3.16** | **.002** |
| | **WASI Vocabulary raw** | **.40** | **4.20** | **< .001** |
| | WASI Matrix Reasoning raw | .13 | 1.36 | .177 |
| Science and maths RT | | | | |
| Model 2a<br>$F(3, 77) = 1.07$,<br>$p = .366$, $R^2 = 4\%$ | **Constant** | | **2.74** | **.008** |
| | Age (months) | -.07 | -0.58 | .566 |
| | WASI Vocabulary raw | -.07 | -0.59 | .555 |
| | WASI Matrix Reasoning raw | .18 | 1.60 | .114 |
| Model 2b<br>$F(4, 76) = 3.19$,<br>$p = .018$, $R^2 = 14\%$,<br>$\Delta R^2 = 10.4\%$ | Constant | | -0.56 | .577 |
| | Age (months) | .03 | 0.26 | .799 |
| | WASI Vocabulary raw | -.04 | -0.32 | .748 |
| | **WASI Matrix Reasoning raw** | **.27** | **2.43** | **.017** |
| | **Simple Go RT** | **.35** | **3.03** | **.003** |

A third hierarchical multiple regression (**Table 3.6**) investigated whether inhibitory control measures could account for science and maths misconception accuracy. The first model (3a) with age, WASI Vocabulary and Matrix Reasoning raw scores, and science and maths control accuracy as predictors, was significant, explaining 26% of the variance. Age and science and maths control accuracy were significant predictors of misconception accuracy. Stroop RT cost was selected using a stepwise approach in model

3b, uniquely accounting for 5% of the variance. Greater Stroop RT cost was associated with lower misconception accuracy. No Go/No-Go variables were selected by the model.

**Table 3.6.** Regression models for science and maths misconceptions. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science and maths misconception accuracy | | | | |
| Model 3a<br>$F(4, 76) = 6.61$,<br>$p < .001$, $R^2 = 26\%$ | Constant | | -0.62 | .535 |
| | **Age (months)** | **.29** | **2.83** | **.006** |
| | WASI Vocabulary raw | .10 | 0.85 | .397 |
| | WASI Matrix Reasoning raw | .09 | 0.87 | .389 |
| | **Science and maths control accuracy** | **.26** | **2.22** | **.029** |
| Model 3b<br>$F(5, 75) = 6.68$,<br>$p < .001$, $R^2 = 31\%$,<br>$\Delta R^2 = 5.0\%$ | Constant | | -0.24 | .812 |
| | **Age (months)** | **.30** | **3.02** | **.004** |
| | WASI Vocabulary raw | .08 | 0.71 | .482 |
| | WASI Matrix Reasoning raw | .09 | 0.87 | .389 |
| | **Science and maths control accuracy** | **.26** | **2.28** | **.026** |
| | **Stroop RT cost** | **-.22** | **-2.33** | **.023** |
| Science and maths misconception RT | | | | |
| Model 4a<br>$F(4, 76) = 101.52$,<br>$p < .001$, $R^2 = 84\%$ | Constant | | -1.37 | .174 |
| | Age (months) | -.01 | -0.23 | .819 |
| | WASI Vocabulary raw | .09 | 1.89 | .063 |
| | WASI Matrix Reasoning raw | .07 | 1.46 | .148 |
| | **Science and maths control RT** | **.911** | **19.61** | **< .001** |
| Model 4b<br>$F(5, 75) = 86.61$,<br>$p < .001$, $R^2 = 85\%$,<br>$\Delta R^2 = 1.0\%$ | Constant | | -1.73 | .088 |
| | Age (months) | -.01 | -0.12 | .908 |
| | WASI Vocabulary raw | .08 | 1.66 | .101 |
| | WASI Matrix Reasoning raw | .07 | 1.48 | .143 |
| | **Science and maths control RT** | **.91** | **20.05** | **< .001** |
| | **Complex No-Go accuracy** | **.10** | **2.26** | **.027** |

The fourth regression investigated misconception RT. The first model (4a) with age, WASI Vocabulary and Matrix Reasoning raw scores, and science and maths control RT as predictors, was significant, explaining 84% of the variance. Only science and maths control RT was a significant predictor of misconception RT. Complex No-Go accuracy was selected in model 4b, uniquely accounting for 1% of the variance. Greater complex No-Go accuracy was associated with higher misconception RT. No Stroop variables were selected by the model.

Follow up exploratory regressions (**Table 3.7**) examined the extent to which these associations held for science and maths individually. Stroop RT cost was not a significant predictor of science (model 5) or maths (model 6) misconception accuracy, although the *p*-values were at trend and the coefficients were in the same direction as the combined analyses. Complex Go/No-Go accuracy was not a significant predictor of science (model 7) or maths (model 8) misconception RT. This time the coefficient was positive for maths, as with the combined analyses, but negative for science.

**Table 3.7.** Regression models for science and maths misconceptions separately. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| Dependent variables | Independent variables | β | *t* | *p* |
|---|---|---|---|---|
| Science misconception accuracy | | | | |
| Model 5 $F(5, 75) = 4.21$, $p = .002$, $R^2 = 22\%$ | Constant | | -0.21 | .904 |
| | **Age (months)** | **.21** | **2.00** | **.049** |
| | WASI Vocabulary raw | -.02 | -0.19 | .849 |
| | WASI Matrix Reasoning raw | .00 | 0.00 | .999 |
| | **Science control accuracy** | **.35** | **2.98** | **.004** |
| | Stroop RT cost | -.18 | -1.77 | .080 |
| Maths misconception accuracy | | | | |
| Model 6 $F(5, 75) = 3.98$, $p = .003$, $R^2 = 21\%$ | Constant | | 0.05 | .961 |
| | **Age (months)** | **.27** | **2.52** | **.014** |
| | WASI Vocabulary raw | .15 | 1.30 | .199 |
| | WASI Matrix Reasoning raw | .15 | 1.42 | .160 |
| | Maths control accuracy | .09 | 0.76 | .452 |
| | Stroop RT cost | -.18 | -1.76 | .083 |
| Science misconception RT | | | | |
| Model 7 $F(5, 75) = 32.84$, $p < .001$, $R^2 = 69\%$ | Constant | | -0.11 | .915 |
| | Age (months) | .08 | 1.28 | .203 |
| | WASI Vocabulary raw | .03 | 0.37 | .711 |
| | WASI Matrix Reasoning raw | .01 | 0.14 | .890 |
| | **Science control RT** | **.84** | **12.43** | **< .001** |
| | Complex No-Go accuracy | -.10 | -1.55 | .124 |
| Maths misconception RT | | | | |
| Model 8 $F(5, 75) = 43.02$, $p < .001$, $R^2 = 74\%$ | Constant | | -0.79 | .430 |
| | Age (months) | -.08 | -1.31 | .195 |
| | WASI Vocabulary raw | .11 | 1.78 | .079 |
| | WASI Matrix Reasoning raw | .11 | 1.86 | .067 |
| | **Maths control RT** | **.83** | **13.96** | **< .001** |
| | Complex No-Go accuracy | .063 | 1.05 | .296 |

In summary, the regression analyses revealed unique roles for response and semantic inhibition in reasoning about science and maths misconceptions. Both response inhibition (complex No-Go accuracy) and semantic inhibition (Stroop RT cost) were predictors of performance when science and maths misconceptions were combined. Proficiency in semantic inhibition was more important for predicting misconception accuracy, while proficient response inhibition was more important for predicting longer RTs when addressing misconception problems. When disciplines were analysed separately the association with complex No-Go accuracy was similar, albeit at trend level, for science and maths, while the association with Stroop RT cost was inconsistent.

## 3.5 Discussion

The current study investigated the role of inhibitory control in counterintuitive science and maths reasoning in adolescence. It was hypothesised that better inhibitory control would be associated with better performance in science and maths misconception problems, when controlling for performance on related problems, age, and general cognitive ability. Ninety adolescents were tested on response and semantic inhibition and a novel science and maths misconceptions task. Both response and semantic inhibition were associated with performance in science and maths misconception trials, beyond performance in control trials and individual differences in general cognitive ability or age. This was the first study to consider the unique roles of response and semantic inhibition in this context, demonstrating that response inhibition may be more related to RTs in counterintuitive reasoning, while semantic inhibition may be more related to accuracy.

As anticipated, accuracy was lower and RTs slower for misconception trials, indicating that reasoning about counterintuitive curriculum-related concepts leads to misconceptions in this age group, even in the oldest participants who should have a good understanding of these concepts having covered them all at school. Only small age effects were observed, in line with standardised assessment findings that only small improvements are made in maths within this age range (Ryan & Williams, 2007).

The reduction in accuracy in false trials compared to true trials was greater for misconception than control trials, which may be due to increased cognitive demand in false trials. To arrive at the correct response, the participant must first read the statement and detect an error, then possibly generate the true statement internally before deciding that the statement presented is false. This may explain why it is easier to answer a true statement correctly, especially if it is counterintuitive. This pattern of performance was observed in science trials only, which may be explained by the inclusion in maths of nine problem-sets containing equations, where both true and false trials require a mental calculation, which should limit any specific increase in cognitive demand for false trials. It should also be noted that a higher proportion of science trials were accompanied by a picture (79% versus 50%) which cannot be ruled out as a source of difference between the two disciplines.

Overall, these findings support the previous literature that misconceptions due to intuitive reasoning exist in this age range (Driver et al., 2015; Stavy & Tirosh, 2000). Although the task was novel and thus has not been extensively validated, the inclusion of problems that cover the curriculum broadly is a strength of the study, allowing greater generalisation and relevance for education.

The inhibitory control tasks showed a degree of improvement with age, echoing findings in the literature (Leon-Carrion et al., 2004). Some measures of inhibitory control were moderately correlated with each other, with the highest correlations between RT measures, likely representing processing speed (Kail, 1993) rather than inhibition per se. There was a marginal negative correlation between the two inhibitory control measures that were selected by the regression model: higher complex response inhibition accuracy was associated with lower semantic inhibition RT cost. This suggests that response and semantic inhibition are partially related, in keeping with the previous literature (Verbruggen et al., 2004), whereby the ability to make less impulsive motor responses is linked to the ability to suppress irrelevant stimuli with less interference.

Both response and semantic inhibition were associated with science and maths misconception performance when controlling for age, general cognitive ability, and control performance. In line with the hypothesis, a smaller difference in RT between incongruent and congruent Stroop trials, suggesting less interference and better semantic inhibition, was associated with higher accuracy on misconception trials. These results fit with the proposal that semantic inhibition may allow suppression of naïve beliefs or irrelevant perceptual information in order to reach the correct answer to counterintuitive problems. Although the amount of variance explained was small, it is still meaningful given that the model controlled for age, general cognitive ability (verbal and non-verbal), and performance in related science and maths control trials. The fact that the association is observed after inclusion of control trials as a covariate in the analyses is consistent with the idea that semantic inhibitory control may play a specific (or more important) role in science and maths counterintuitive reasoning rather than science and maths reasoning more broadly. This interpretation is bolstered by the regression analyses of overall science and maths performance which showed no significant inhibitory control predictors for accuracy. For RT, Simple Go RT was a significant predictor, but this measure is not thought to be an inhibitory control measure per se, rather it appears to represent an overall speed of response factor.

The ability to withhold a response in the complex Go/No-Go task was associated with longer RTs on misconception trials. Therefore, contrary to the hypothesis, good response inhibition was not associated with better performance in the science and maths misconceptions task. However, a possible interpretation is that good response inhibition may afford more time for consideration of the response, with a less impulsive pattern of responding. Individuals may not necessarily eventually choose the correct response, but they may be able to spend more time thinking about their response and evaluate competing alternative answers.

The complex Go/No-Go measured inhibitory control within the context of a cognitive load. The regression model's selection of a complex rather than simple Go/No-Go variable implicates individual differences in the ability to manage combined response inhibition and working memory demands. The use of this ability is exemplified by a maths misconception problem that requires counting items and calculating probabilities, and holding this information in mind while considering how it applies to the statement. This account is consistent with suggestions that beliefs must be held in working memory during reasoning, before the incorrect response is inhibited (Zaitchik et al., 2014). Measuring a purer form of working memory in a separate task would enable assessment of the extent to which working memory makes a unique contribution outside of the context of inhibitory control (see **Chapters 4, 6 and 7**).

The discipline-specific analyses had reduced power due to the smaller number of trials, in addition to the different number of pictures within the stimuli, so must be interpreted cautiously. There were also fewer maths problems than science problems. The results overall suggest that misconceptions in science and maths show similar associations to semantic inhibition, with potentially different associations with response inhibition: higher complex No-Go accuracy was associated with shorter responses in science but longer responses in maths. However, these exploratory analyses did not reach significance so further research would be necessary to determine discipline-specific associations.

The associations seen in this novel study suggest that both response inhibition and semantic inhibition play unique roles in counterintuitive science and maths reasoning problems that are curriculum-related. The study provides further evidence that poor reasoning partly reflects poor inhibitory control as opposed to simply poor logic or understanding. An examination of the neural correlates of counterintuitive reasoning in adolescence would reveal more about the cognitive mechanisms involved (see **Chapter 5**).

The thesis will next consider *learning* of counterintuitive concepts in science and maths, as opposed to reasoning about them. **Chapter 4** describes two classroom-based studies that attempted to establish the extent to which inhibitory control is related to the acquisition of new counterintuitive concepts in science and maths.

# Chapter 4 Classroom studies: Inhibitory control in science and maths learning

## 4.1 Overview

Behavioural and neuroimaging research suggests that inhibitory control is an important factor in counterintuitive science and maths reasoning, and by extension, learning. The current studies sought to investigate the role of inhibitory control in learning new counterintuitive concepts in the classroom, in an attempt to bring the research closer to educational practice. Working memory was also considered, to establish whether it would have a unique role in learning counterintuitive material, beyond inhibition. In **study A**, 12- and 13-year-olds (*N* = 48) were taught two new counterintuitive physics concepts in their normal school lessons, and in **study B**, 13- and 14-year-olds (*N* = 69) were taught a new counterintuitive maths concept in their normal school lesson. Inhibitory control did not associate with misconception presence one week before the lessons, one week following the lessons, nor one month later in either study. However**,** response inhibition and working memory were associated with better performance on control maths questions. The studies highlight the difficulties in conducting classroom-based research, and tentatively suggest that inhibitory control may be more important for reasoning about counterintuitive concepts that are already learnt as opposed to during the acquisition of new counterintuitive concepts.

## 4.2 Introduction

The behavioural study presented in **Chapter 3** demonstrated that both response and semantic inhibitory control were predictors of performance in misconception problems. This provided further evidence for the idea that inhibitory control enables the suppression of incorrect responses during science and maths reasoning. In addition to assisting in-the-moment reasoning, inhibitory control is thought to have a role in the learning of new counterintuitive concepts through the same mechanism of suppressing conflicting prior knowledge or perceptual information (Mareschal, 2016). While there are a number of studies investigating links between inhibitory control and science and maths reasoning, there have been relatively few attempts to examine inhibitory control in the learning of new information.

Notable examples investigated inhibitory control and other executive functions in the learning of new biology (Rhodes et al., 2014) and chemistry (Rhodes et al., 2016) concepts. Participants were aged 12 to 13 years and both studies took the same approach. In biology, participants were taught about DNA, and in chemistry, participants were taught about acids and alkalis. Each teaching session lasted 45 min, and included taught material and a practical session. Following the lesson, participants completed an assessment that covered facts taught in the lesson as well as conceptual understanding that required problem-solving. Inhibitory control (the stop signal task), spatial working memory, attention set-shifting, and planning were all measured prior to the lesson to examine how each of these could predict science outcomes. In biology, planning predicted performance on the factual part of the assessment, while working memory and planning both predicted the conceptual part of the assessment (Rhodes et al., 2014). In chemistry, there were no significant predictors of the factual part of the assessment, while working memory and attention set-shifting predicted the conceptual part of the assessment (Rhodes et al., 2016). These findings suggest that inhibitory control may not be important in learning new concepts, and that there may be different patterns of association for different science disciplines. Working memory may be a common important factor that contributes to both biology and chemistry learning. However, these studies did not look at counterintuitive concepts, so it remains possible that inhibitory control would contribute to learning new science and maths material where misconceptions may be present. There is also the possibility that semantic inhibition might play a more important role in learning than response inhibition, which was measured in the studies from Rhodes and colleagues (2014, 2016).

The aim of the current studies was to assess the extent to which inhibitory control associates with the learning of new counterintuitive concepts in a typical classroom setting. In order to ensure that the studies were close to a normal school learning exercise, close consultation with teachers occurred during the design of the studies, which also ensured that participants had not previously encountered the material at school. **Study A** aimed to investigate the role of inhibitory control in learning new counterintuitive physics concepts, while **study B** aimed to investigate the same associations in maths. As in **Chapter 3**, both response and semantic inhibition were investigated to examine the possible specificity of these measures. Taking into account the studies described above, working memory was also measured. Since working memory was a last minute addition, in **study A** this occurred in a later testing session after the initial pre-test. The working memory measure was added due to the publication of other research that indicated working memory was an important component in learning new material in science (Rhodes et al., 2014). It was considered that even though testing was at a later date, individual differences in working memory would likely be stable enough to relate meaningfully to performance in the physics tests. Finally, it was expected that the extent to which participants pay attention in lessons might influence performance, so observations of participants in their lessons took place. In both studies, it was hypothesised that those with better inhibitory control would be better able to suppress their naïve theories and reach the correct answer following a lesson on the counterintuitive concept. Exploratory analyses examined the role of working memory, and attention in classroom lessons on general measures of physics and maths performance.

## 4.3 Study A: Physics

Discussions with a school physics teacher identified two counterintuitive physics concepts that a) the students had not learned yet; b) students were likely to get wrong before receiving taught material on; c) some students were likely to keep getting wrong after instruction; and d) were relevant to the age group and going to be taught that year (**Figure 4.1**). These criteria were selected to ensure that the concepts were relevant to the participants' normal learning, allowing conclusions to relate to the classroom, and to maximise variability in responses to enable assessment of individual differences. The first counterintuitive concept concerned gravity, and the misconception that heavier objects fall to the ground faster than lighter objects of the same size (the gravity misconception). The second counterintuitive concept was related to pressure change, and the misconception that a burning candle under a jar uses up oxygen to create a vacuum (the pressure misconception).

### 4.3.1 Methods

#### 4.3.1.1 Participants

Two Year 8 classes of participants at an English fee-paying independent secondary school where students come from a variety of ethnic backgrounds and performance is above the national average, took part (*N* = 48, 22 girls, 26 boys). The age range was 12.75 to 13.75 years (*M* = 13.28, *SD* = 0.24). Letters were sent home to parents inviting their children to take part, accompanied by opt-out consent forms. Written consent was obtained from participants before taking part. There was no exclusion criterion, and the local ethics committee approved the study.

#### 4.3.1.2 Physics lessons

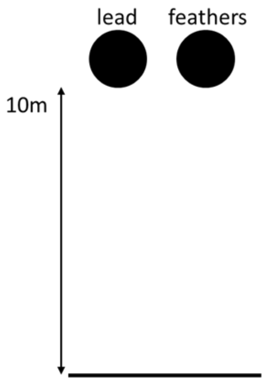
Each counterintuitive concept was addressed in a separate physics lesson. The lessons were designed and carried out by the teacher who usually teaches these classes (the same teacher for all pupils). Each lesson lasted 45 min and included taught material, group experiments, and worksheets, in the format of a normal school lesson. There was no measure of fidelity, since the teacher was asked to simply teach their lesson as they normally would, in order to make the lesson as close to normal practice as possible. This means that the pupils from one class may have had a slightly different learning experience to the pupils from another class, as the teacher may have adapted the teaching approach to better suit the pupils.

#### 4.3.1.3 Measures

##### 4.3.1.3.1 Physics tests

A physics test was administered at three time points to track changes in counterintuitive concept understanding, as well as general physics performance (at pre-test, post-test, and follow up). The 10 min paper and pen physics test consisted of nine questions: the two counterintuitive concept questions (**Figure 4.1**), and a range of questions addressing other physics topics at an appropriate level for the participants. The questions were a mixture of multiple choice and those requiring written answers, designed with the help of the teacher, including three that the teacher thought constitute counterintuitive concepts and four that did not. The physics teacher thought that this range of questions was close to mimicking a typical class test which would include different types of questions examining different types of content (i.e. counterintuitive and non-counterintuitive, and different topics in the curriculum). The order of the questions and of answers for the multiple choice questions of the physics test varied between the three time points. Responses to the gravity and pressure questions were coded with a 1 or 0 to indicate whether a misconception was shown in the answer.

a. Two rubber balls are identical in shape and size. One rubber ball is filled with lead and the other is filled with feathers. Here they are held at a height of 10 metres from the ground.

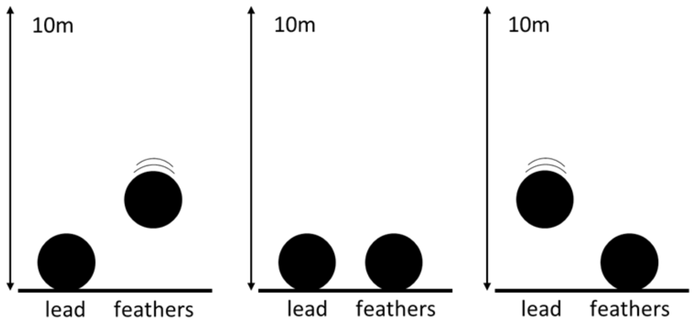



The rubber balls are dropped at the same time from a height of 10m. What will happen? Tick **ONE** box from the options below.

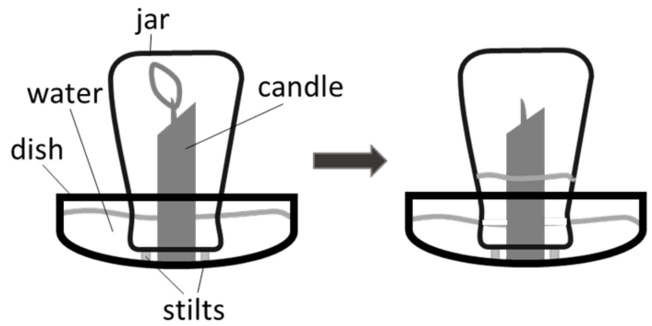
☐ The ball of lead will reach the ground first

☐ Both balls will reach the ground at the same time

☐ The ball of feathers will reach the ground first




b. A jar is placed over a candle that is in a dish of water. Stilts are used to ensure water can flow between the dish and the jar. The flame goes out and the water level rises up the jar.




Why does the water level rise inside the jar as the candle goes out?

.......................................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

**Figure 4.1.** Questions in the physics test addressing the (a) gravity and (b) pressure misconceptions.

An average score for the other seven questions was also calculated and the school provided average physics scores and end of year physics exam results for more general measures of physics performance.

#### 4.3.1.3.2 Executive function tasks

The executive function tasks were programmed in PHP and JavaScript, running in an Ubuntu web server with data stored using mysql. The simple and complex Go/No-Go tasks and numerical Stroop were administered as in the first behavioural study, but in a group testing session as a class, in an IT room, where each participant carried out the tasks on a separate computer. A brief introduction was given by the researcher, who explained which website to go to, as well as the response keys required for each task. Participants were asked to complete the tasks in silence and to read the instructions carefully before each task. They were also told to raise their hand if they had any problems or questions, and the researcher spoke to any individuals who required help (e.g. if they had not understood the instructions and were not passing the practice). Accuracy and RT were recorded for all three tasks, and together they took approximately 10 min to complete.

A computerised dot matrix task, adapted from the Dot Matrix test of the Automated Working Memory Assessment (Alloway, 2007), was administered in a separate group testing session to measure VS-WM. In a practice phase, three dots appeared in turn for 600 ms each in a white four by four grid, with a 300 ms delay between dots. The grid turned orange for a delay of 1500 ms, and when the grid turned white again, participants used the computer mouse to click where the dots had appeared, in the order that they had appeared. The practice repeated, with a different order of dots, if the participant failed. The main trial continued with a load of three dots, and the load increased by one after four trials, until two incorrect answers were given at any one load, or the highest level of load eight was completed. The total number of correct trials was recorded. The task took approximately 3 min.

Participant behaviour during the two taught lessons was observed by the researcher, using an adapted version of a system developed by Blatchford, Bassett and Brown (2011), for a very coarse measure of general attention in lessons. Each participant was observed in turn. Each observation lasted 10 s, followed by 20 s for coding the observed behaviour into two categories: either as on-task or off-task (or half and half if applicable). The process immediately continued with the next child, and paused only when observations became impossible (e.g. when participants moved round the room to collect equipment and were no longer identifiable based on the location of their chair). Observations continued throughout the whole lesson, so that multiple observations were recorded per participant per lesson. The mean number of observations per participant was 6 (*SD* = 1). The percentage of observations that were classed as on-task was calculated for each participant.

#### 4.3.1.4 Procedure

During the inhibitory control assessment, participants first performed the simple Go/No-Go, then the complex Go/No-Go, and then the numerical Stroop. The physics pre-test took place in a physics lesson within three days of the inhibitory control assessment. The two 45 min counterintuitive concept lessons took place within a week of the pre-test. The physics post-test was two weeks after the pre-test, and the follow up occurred approximately four weeks later. The VS-WM task was completed approximately seven months later. Note that six participants also took part in **study B**, three of whom had not completed the executive function tasks and so their executive function data were taken from **study B**.

#### 4.3.1.5 Statistical analysis

##### 4.3.1.5.1 Physics tests

Participants who missed a counterintuitive concept lesson or a physics test were excluded on the relevant analyses (those who attended the lesson and answered all three tests: gravity final *N* = 33, pressure final *N* = 32). Cochran's *Q* was calculated to examine change over time in misconception responses for the gravity and pressure questions individually, and these were followed up with McNemar Tests.

##### 4.3.1.5.2 Executive function tasks

Participants whose mean accuracy or RT was further than ±3.29 *SD*s away from the group mean were excluded from analyses of the task on which they were an outlier. One participant was excluded from simple Go/No-Go analyses due to low accuracy and a further participant was excluded for high RT (final *N* = 46). One participant was excluded from complex Go/No-Go analyses due to low accuracy (final *N* = 47). Two participants opted out of completing the numerical Stroop task, and one participant was excluded from Stroop analyses due to low accuracy and RT (final *N* = 45).

Repeated measures ANOVAs were performed on accuracy in the simple and complex Go/No-Go with the within-subject factor Trial type (Go, No-Go). Two further repeated measures ANOVAs were performed on accuracy and RT in the numerical Stroop with the within-subject factor Trial type (congruent, incongruent).

Five participants did not complete the VS-WM task: Since this measure was collected later, three participants had since left the school, and a further two were not able to attend a testing session (final $N = 43$). There were no outliers on the VS-WM task. One participant was absent from both lessons so had no observation data (final $N = 47$).

#### 4.3.1.5.3 Association between executive functions and physics

First, point-biserial partial correlations, controlling for age, were used to examine associations between inhibitory control (simple No-Go accuracy, complex No-Go accuracy, Stroop accuracy cost, and Stroop RT cost) and presence of a gravity and pressure misconception at each of the three time points. Second, point-biserial partial correlations, controlling for age, were used to examine associations between the same inhibitory control variables and presence of a misconception at post-test and follow up only in those who held the misconception at pre-test. Participants who missed a lesson were excluded from the analysis relating to that lesson.

Third, partial correlations controlling for age were run to examine the association between executive function (simple No-Go accuracy, complex No-Go accuracy, Stroop accuracy cost, Stroop RT cost, VS-WM, percentage on-task) and general physics performance (average number of correct responses across the physics tests, average physics score at school, end of year physics exam).

## 4.3.2 Results

### 4.3.2.1 Physics tests

Cochran's $Q$ indicated a significant change in the presence of the gravity misconception over time, $Q(2) = 38.00$, $p < .001$ (**Figure 4.2a**). McNemar's tests showed significant differences between pre-test and both post-test and follow up, $p$'s $< .001$, but no difference between post-test and follow up where no misconception was held by any participant ($p = 1$).

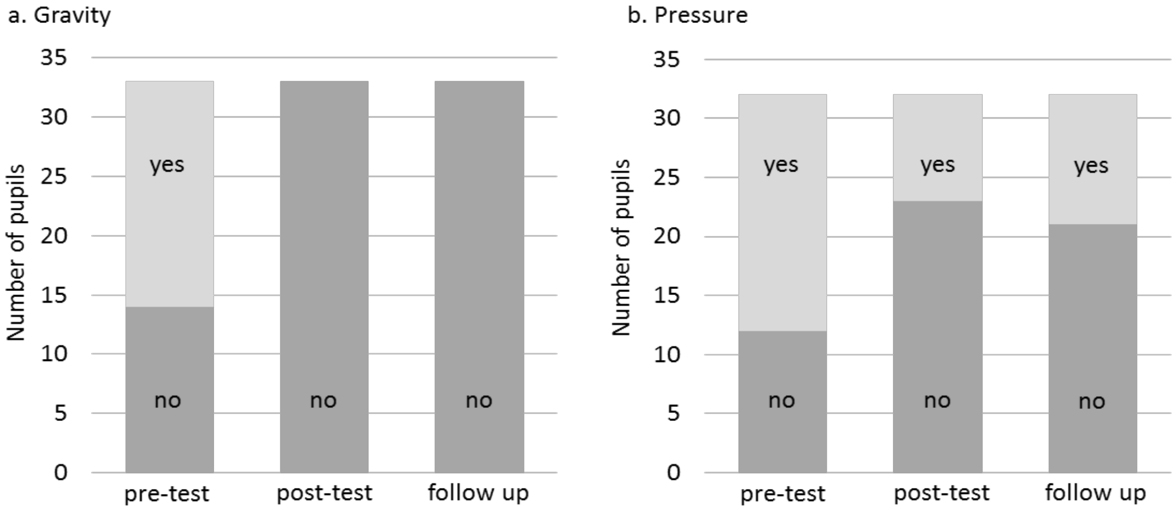



**Figure 4.2.** Presence of a misconception in response to the (a) gravity and (b) pressure questions.

Cochran's *Q* also showed a change over time in presence of the pressure misconception, $Q(2) = 9.81$, $p = .007$ (**Figure 4.2b**). McNemar's tests showed a marginal reduction in presence of the misconception between pre- and post-test, $p = .064$, and no significant difference between post-test and follow up or between pre-test and follow up, *p*'s > .1.

Overall there was a reduction in the number of pupils who held a misconception after the counterintuitive concept lesson, as anticipated. In the gravity misconception, all participants learnt the correct answer and maintained this even at follow up.

### 4.3.2.2 Executive function tasks

Repeated measures ANOVAs on accuracy in both the simple and complex Go/No-Go showed a significant effect of Trial type, with higher accuracy in Go trials than No-Go trials (**Table 4.1**). Repeated measures ANOVAs on accuracy and RT in the numerical Stroop showed significant effects of trial type with higher accuracy and faster RTs in congruent trials than incongruent trials (**Table 4.1**). The inhibitory control tasks showed the anticipated effects of reduced accuracy and longer RTs in trials that require inhibition.

**Table 4.1.** Accuracy and RT estimated marginal means in the inhibitory control tasks.

| | **Simple Go/No-Go** | **Complex Go/No-Go** | **Numerical Stroop** |
|---|---|---|---|
| | ***M (SE)*** | ***M (SE)*** | ***M (SE)*** |
| **Accuracy (%)** | | | |
| *Trial type* | $F(1, 45) = 78.48$, $p < .001$, $\eta_p^2 = .636$ | $F(1, 46) = 243.58$, $p < .001$, $\eta_p^2 = .841$ | $F(1, 44) = 45.03$, $p < .001$, $\eta_p^2 = .506$ |
| Go/Congruent | 98.9 (0.3) | 92.4 (1.1) | 90.2 (1.9) |
| No-Go/Incongruent | 79.0 (2.3) | 44.7 (3.3) | 78.0 (2.5) |
| **RT (ms)** | | | |
| *Trial type* | - | - | $F(1, 44) = 131.67$, $p < .001$, $\eta_p^2 = .750$ |
| Go/Congruent | 383 (5) | 439 (12) | 701 (13) |
| No-Go/Incongruent | - | - | 773 (13) |

Descriptive statistics for all other measures, including executive functions and physics tests, are reported in **Table 4.2**.

**Table 4.2.** Mean, standard deviation and range in VS-WM, attention in lessons, average correct in other physics test questions, average school physics score and end of year physics exam score.

| **Measure** | ***M (SD)*** | ***Range*** |
|---|---|---|
| VS-WM total | 10 (3) | 0-16 |
| Attention in lessons percentage on-task | 76 (25) | 0-100 |
| Physics test average correct | 2.3 (1) | 0-5 |
| Average school physics score | 81 (6) | 68-95 |
| End of year physics exam score | 67 (12) | 39-88 |

### 4.3.2.3 Association between executive functions and physics

Point-biserial partial correlations between misconception presence at each time point and inhibitory control (simple No-Go accuracy, complex No-Go accuracy, Stroop accuracy cost, Stroop RT cost), controlling for age, showed no significant associations, $p$'s > .1. There was also no association between pressure misconception presence at post-test and follow up when looking only at those who held the misconception at pre-test, $p$'s > .3.

Partial correlations between executive function (simple No-Go accuracy, complex No-Go accuracy, Stroop accuracy cost, Stroop RT cost, VS-WM, percentage on-task) and general physics performance (average physics test scores excluding the two misconceptions, average physics score at school, end of year physics exam), controlling for age, showed a few trend associations. Complex No-Go accuracy was negatively associated with physics school test average, $r(29) = -.31$, $p = .083$. Stroop RT cost was negatively associated with both physics school test average, $r(29) = -.33$, $p = .073$ and physics end of year exam, $r(29) = -.33$, $p = .073$.

### 4.3.3 Study A Discussion

This study attempted to determine the extent to which inhibitory control associates with the learning of new counterintuitive physics concepts in the classroom. There was no association between inhibitory control and misconception presence before or after teaching of counterintuitive concepts relating to gravity and pressure.

An unexpected finding was that after being taught the gravity concept, all participants answered correctly, preventing analysis into individual differences at post-test or follow up for this concept. This concept had been chosen in consultation with the physics teacher, in part because it was anticipated that students would continue to find this concept challenging following the lesson. This was based on the teacher's own experience teaching this concept, and was in line with findings from the adult neuroimaging study in which adults continued to get this concept wrong when they were not science experts (Masson et al., 2014). Ceiling performance following teaching may be due to the high ability of the sample included in this study, or alternatively due to the nature of the questions, which changed very little between tests. Perhaps changing the questions more would have reduced familiarity, or including a more complex question that required written explanation would have elicited latent misconceptions. It is also possible that extra attention was given to learning the concepts because the participants were aware that they were taking part in a research study, and so the new concepts were potentially more salient than would normally be the case. Finally, this unexpected result might have arisen because the time between teaching and testing was relatively short. The adults tested by Masson and colleagues (2014) would have been taught the concept many years previously, rather than just a few weeks before testing.

Performance improved, and variation remained for the pressure concept, but no inhibitory control variables associated with misconception presence at any time point. This lack of association may indicate that inhibitory control is not an important factor in the learning of this concept, which would replicate the previous findings in science learning (Rhodes et al., 2014, 2016). An alternative explanation is that the sample size was too low to detect what may be a small effect in relation to counterintuitive concept learning specifically. The sample size reduced throughout the course of the study due to participant absences from lessons. For a full data set, a participant had to be present for a total of 8 sessions, and in a busy school where many students have extracurricular lessons in the day this was a challenge. Nonetheless, generally there were only a few participants missing for each individual analysis.

Finally, the exploratory investigations into executive function measures and more general measures of physics performance did not show any significant associations. This time the finding did not replicate the previous findings in science learning which found an association between working memory and science learning (Rhodes et al., 2014, 2016). It is possible that had the working memory measure been administered closer to the time of the science lesson, a link would have been found. It was assumed that individual differences would be stable within the 7 month period, but this was not explicitly tested. Nonetheless, a replication of the two science learning studies by Rhodes and colleagues (2014, 2016) found that in physics there was no significant association between learning and executive function (S. M. Rhodes, J. N. Booth, L. E. Palmer, R. A. Blythe, M. Delibegovic, & N. J. Wheate, personal communication, November 22, 2018). There were some trend associations which can be tentatively discussed. There was a negative association between response inhibition in the context of a working memory load (complex No-Go accuracy) and physics school test average. This is the opposite of what might be expected, as the behavioural study in **Chapter 3** reported that this measure was positively associated with misconception RT, which was taken to reflect increased time reasoning. More consistently with findings from **Chapter 3**, Stroop RT cost, a measure of semantic inhibition, was negatively associated with both school measures of physics. This means that those who were less affected by the conflict required to respond to incongruent trials performed better in school physics tests. This is some further evidence that semantic inhibition may play a role in general physics performance, although these relations were not significant at an alpha level of .05 so firm conclusions cannot be drawn.

Overall, **study A** showed no evidence for inhibitory control in learning new counterintuitive physics concepts in the classroom.

## 4.4 Study B: Maths

**Study B** aimed to investigate inhibitory control in the learning of a new counterintuitive maths concept. Parts of the method were tweaked to capitalize on what was learnt from **study A** and in response to teacher requirements. This time, the VS-WM task was completed in the same session as the inhibitory control tasks. Two teachers were involved in choosing the misconception for **study B** as no teacher taught more than one class in a year group. The teachers felt that only one misconception should be taught as this would fit better with their schedule. The counterintuitive concept chosen in consultation with the teachers was reverse percentages (**Figure 4.3**), which fulfilled the same criteria as in **study A**. Due to the nature of the concept, it was considered that the same concept could be addressed through numerous questions, since the numbers and words can be changed. This would allow for more variation in responses rather than capturing presence of a misconception with just one question at each time point. Participants were a year older in this study, as the teachers who were keen to take part did not teach any Year 8 classes. Finally, to increase the number of participants, two classes took part in the study one year, and another two classes took part the following year. A total of three teachers taught the classes, with one teaching over both years of the study.

| a. | b. |
|---|---|
| A sale in a bike shop has 20% off all items.<br>Sally bought a bike for £300 in the sale.<br>What was the price of the bike before the sale? | A painting originally bought for £700 is up for auction.<br>The winning bid was 20% higher.<br>How much was the winning bid? |
| A violin has increased in value by 75%.<br>The violin is now worth £700.<br>What was the violin's original value? | Socks are currently on sale with 25% off.<br>A pair of stripy socks was originally £4.<br>How much do they cost in the sale? |

**Figure 4.3.** Example (a) misconception and (b) control questions in the maths test. Misconception questions provide the final value following a percentage change, while control questions provide the original value before the percentage change.

## 4.4.1 Methods

### 4.4.1.1 Participants

Four Year 9 classes of participants at the same English fee-paying independent secondary school where students come from a variety of ethnic backgrounds and performance is above the national average, took part ($N$ = 69, 38 girls, 31 boys). The age range was 13.00 to 14.42 years ($M$ = 13.85, $SD$ = 0.36). The classes ranged in ability from top to bottom set. Letters were sent home to parents inviting their children to take part, accompanied by opt-out consent forms. Written consent was obtained from participants before taking part. There was no exclusion criterion, and the local ethics committee approved the study.

### 4.4.1.2 Maths lessons

One maths lesson addressed the counterintuitive concept, and each class was taught by their own teacher. The lessons were designed and carried out by the teachers who usually teach these classes. Each lesson lasted 45 min and included taught material and practice questions in the format of a normal school maths lesson.

### 4.4.1.3 Measures

#### 4.4.1.3.1 Maths tests

Maths tests were administered at three time points as in **study A**. The 10 min paper and pen maths tests each consisted of ten questions: five relating to the counterintuitive concept, and five control questions (**Figure 4.3**). The questions all required calculations and were open ended, calculators were allowed, and the questions were designed to mimic a typical class test. All questions were checked with the teachers to ensure they correctly represented the counterintuitive concept and that they were at the right level for the class. Each test contained different questions, and the misconception and control questions appeared in random order. The number of correct answers was calculated for misconception and control questions on each test.

An overall number of correct responses on the control questions was also calculated and the school provided average maths scores and end of year maths exam results for more general measures of maths performance. Note that 33 participants do not have data for the school maths performance measures as these participants took part in the study in the second year, and so they have not been finalised and provided by the school at the time of writing.

#### 4.4.1.3.2 Executive function tasks

Inhibitory control and working memory measures were administered as described in **study A**. In order to increase the number of observations per participant for the attention in lessons measure, this was carried out over three lessons for each class, rather than two. This was to get a more representative measure of their classroom behaviour since there would have only been one observation if only the counterintuitive concept lesson was observed.

### 4.4.1.4 Procedure

The inhibitory control and VS-WM assessment took place in the normal school maths room or in the computer room, as a class group, for a session lasting approximately 15 min. Participants first performed the simple Go/No-Go, then the complex Go/No-Go, then the numerical Stroop, and finally the VS-WM task. The maths pre-test took place in the normal maths room as part of a lesson, either the day before or the day after the inhibitory control tasks. The 45 min counterintuitive concept lesson took place four or five days after the pre-test, and was given by the normal maths teacher. The maths post-test occurred seven days later, and the follow up was approximately four weeks later.

### 4.4.1.5 Statistical analysis

#### 4.4.1.5.1 Maths tests

Participants who did not attend the counterintuitive concept lesson were excluded from the analysis (final $N = 57$). A two (Question type: control, misconception) by three (Time: pre-test, post-test, follow up) repeated measures ANOVA was run on accuracy. These were followed up with further repeated measures ANOVAs.

#### 4.4.1.5.2 Executive function tasks

The same cut-off of ±3.29 *SD*s from the group mean was applied to outliers. One participant was not present for the initial testing sessions, one participant was removed from all Go/No-Go analyses due to low Go accuracy, and two participants were removed from all Stroop analyses because of low accuracy (simple Go/No-Go final $N = 67$, complex Go/No-Go final $N = 67$, numerical Stroop final $N = 66$, and VS-WM final $N = 68$). Repeated measures ANOVAs were performed on accuracy in the simple and complex Go/No-Go with the within-subject factor Trial type (Go, No-Go). Two further repeated measures ANOVAs were performed on accuracy and RT in the numerical Stroop with the within-subject factor Trial type (congruent, incongruent).

Due to computer failure, one class's observation data for one lesson was lost. The mean number of observations per participant was 12 (*SD* = 9). The variation was large due to the different sizes of the classes which ranged from 7 to 26 pupils.

#### 4.4.1.5.3 Association between executive functions and maths

First, partial correlations controlling for age were run between misconception total correct and inhibitory control (simple No-Go accuracy, complex No-Go accuracy, Stroop accuracy cost, and Stroop RT cost) at each time point. Second, partial correlations controlling for age were run between misconception total correct and inhibitory control at post-test and follow up only in those who scored 0, 1, or 2 at pre-test in misconception questions (this threshold indicating that they held the misconception). Those who missed the misconception lesson were excluded from analyses relating to post-test and follow up.

Third, partial correlations controlling for age were run to examine the association between executive function (simple No-Go accuracy, complex No-Go accuracy, Stroop accuracy cost, Stroop RT cost, VS-WM, percentage on-task) and general maths performance (number of correct control answers across the maths tests, average maths score at school, end of year maths exam).

## 4.4.2 Results

### 4.4.2.1 Maths tests

The repeated measures ANOVA showed a main effect of Question type $F(1, 50) = 66.82$, $p < .001$, $\eta_p^2 = .572$, with more correct responses to control questions than to misconception questions (**Figure 4.4**). There was also a main effect of Time, $F(2, 100) = 3.96$, $p = .022$, $\eta_p^2 = .073$, and a significant interaction between Question type and Time, $F(2, 100) = 13.62$, $p < .001$, $\eta_p^2 = .214$.

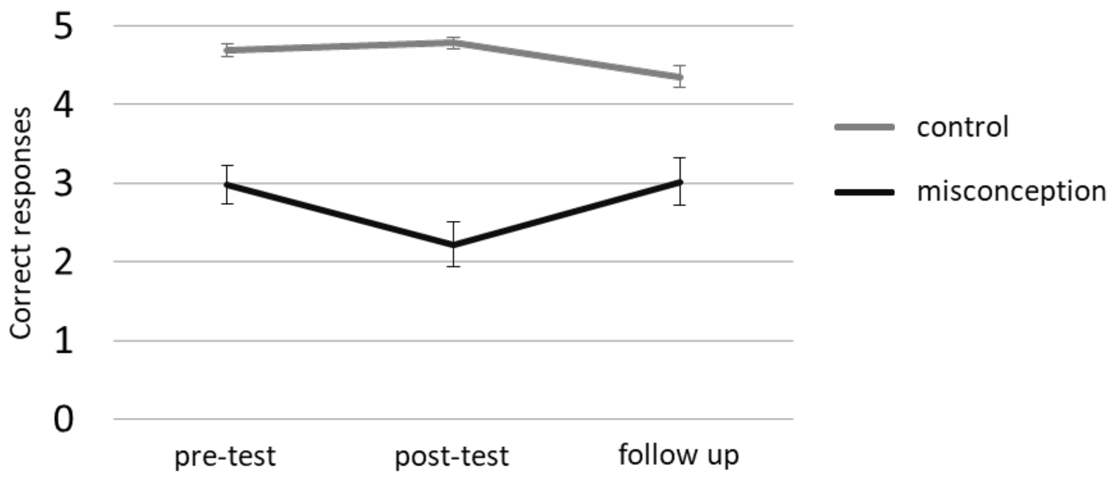



**Figure 4.4.** Estimated marginal mean number of correct responses (± SE) in control and misconception maths questions at each time point.

This interaction was first followed up with separate ANOVAs for accuracy in misconception questions. There was a main effect of Time, $F(2, 98) = 10.62$, $p < .001$, $\eta_p^2 = .178$. Follow up repeated measures ANOVAs showed a significant difference between pre- and post-test, $F(1, 51) = 11.98$, $p = .001$, $\eta_p^2 = .190$, post-test and follow up, $F(1, 51) = 24.68$, $p < .001$, $\eta_p^2 = .326$, but not between pre-test and follow up, $p = .855$.

A repeated measures ANOVA for accuracy in control questions showed a main effect of Time, $F(2, 100) = 5.496$, $p = .005$, $\eta_p^2 = .099$. Follow up repeated measures ANOVAs showed a significant decrease between post-test and follow up, $F(1, 51) = 7.65$, $p = .008$, $\eta_p^2 = .130$, but not from pre-test to post-test or follow up, $p$'s > .08.

Overall the maths tests did not show the anticipated pattern of sustained increased performance after the lesson. Performance in the misconception questions decreased after the lesson then returned to pre-test levels. Performance in the control questions increased slightly after the lesson and then decreased to pre-test levels.

### 4.4.2.2 Executive function tasks

Repeated measures ANOVAs on accuracy in the simple and complex Go/No-Go both showed a significant effect of Trial type, with higher accuracy in Go trials than No-Go trials (**Table 4.3**). Repeated measures ANOVAs on accuracy and RT in the Stroop showed significant effects of trial type with higher accuracy and faster RTs in congruent trials than incongruent trials (**Table 4.3**). The inhibitory control tasks showed the anticipated inhibition effects of lower accuracy and longer RTs in the trials requiring inhibitory control.

**Table 4.3.** Accuracy and RT estimated marginal means in the inhibitory control tasks.

| | **Simple Go/No-Go** | **Complex Go/No-Go** | **Numerical Stroop** |
|---|---|---|---|
| | ***M (SE)*** | ***M (SE)*** | ***M (SE)*** |
| **Accuracy (%)** | | | |
| *Trial type* | $F(1, 66) = 118.01$, $p < .001$, $\eta_p^2 = .641$ | $F(1, 66) = 316.14$, $p < .001$, $\eta_p^2 = .827$ | $F(1, 54) = 107.08$, $p < .001$, $\eta_p^2 = .622$ |
| Go/Congruent | 99.0 (0.2) | 92.9 (1.0) | 92.8 (0.9) |
| No-Go/Incongruent | 82.7 (1.6) | 48.4 (2.5) | 80.6 (1.4) |
| **RT (ms)** | | | |
| *Trial type* | n/a | n/a | $F(1, 65) = 148.38$, $p < .001$, $\eta_p^2 = .695$ |
| Go/Congruent | 365 (4) | 404 (6) | 706 (11) |
| No-Go/Incongruent | - | - | 779 (10) |

Descriptive statistics for the remaining executive function and maths measures are reported in **Table 4.4**.

**Table 4.4.** Mean, standard deviation and range in VS-WM, attention in lessons, overall correct in maths control questions, average maths test score and end of year maths exam score.

| **Measure** | ***M (SD)*** | ***Range*** |
|---|---|---|
| VS-WM total | 10 (3) | 0-16 |
| Attention in lessons percentage on-task | 90 (13) | 40-100 |
| Maths control correct | 14 (1) | 9-15 |
| Average school maths score | 59 (13) | 31-85 |
| End of year maths exam score | 44 (13) | 17-44 |

### 4.4.2.3 Association between executive functions and physics

Partial correlations between misconception total correct and inhibitory control (simple No-Go accuracy, complex No-Go accuracy, Stroop accuracy cost, and Stroop RT cost), controlling for age, showed only one marginal association. Simple No-Go accuracy was associated with misconception total correct at pre-test, $r(63) = .22$, $p = .075$, all other *p*'s > .1. The same partial correlations were run at post-test and follow up only in those who had low misconception accuracy at pre-test, and none of these associations were significant, *p*'s > .2.

Partial correlations, controlling for age, were run between all executive function measures (simple No-Go accuracy, complex No-Go accuracy, Stroop accuracy cost, Stroop RT cost, VS-WM, percentage on-task) and general maths performance (number of correct control answers across the maths tests, average maths score at school, end of year maths exam). Simple No-Go accuracy positively correlated with total correct responses in maths control questions, $r(57) = .29$, $p = .025$. The percentage on-task measure positively correlated with maths end of year exam, $r(31) = .55$, $p = .001$, and marginally correlated with maths school test average, $r(31) = .33$, $p = .065$. No other associations were significant, *p*'s > .1.

### 4.4.3 Study B Discussion

This study attempted to determine the extent to which inhibitory control associates with the learning of a new counterintuitive maths concept in the classroom. There was no convincing association between inhibitory control and misconception performance, but there were associations between both response inhibition and attention in lessons with general maths performance.

Contrary to the hypotheses, performance in misconception questions dropped between pre-test and post-test. This is contrast to **study A** which found an increase in performance after the counterintuitive concept lessons. It was anticipated that pre-test performance would be very low, and that performance would increase after the lesson. One explanation is that at post-test, participants thought they knew the procedure having received the lesson, but as it is a difficult concept were more likely to get it wrong through reduced effort associated with believing the procedure is well understood. Performance then increased to pre-test levels, perhaps because the time lag meant participants had to think harder about how to do the problems. An explanation given by a teacher for the persisting low misconception performance following the lesson is that the calculation procedure is very hard, and will take a long time for the students to learn. They have been doing the control questions for a long time, and the new method required for misconception questions is difficult to get used to.

Performance in control trials decreased at follow up, which was also surprising, although performance was high at all three time points. One explanation for the lower control performance at follow up is that those who had learned the correct strategy for the new counterintuitive concept were now using the misconception strategy for control questions, which is incorrect in this context. The finding might instead relate to perceived ability, whereby participants feel they have learnt the concept and put less effort in at post-test. In a future study, it would be worth asking students to write down the individual steps taken in coming to an answer, in order to test this directly. Since calculators were used, many students simply wrote down their answer without showing their working. Alternatively, time of day of testing may have influenced performance since the tests were carried out at different times.

As in **study A**, the results provide minimal evidence for the role of inhibitory control in learning new counterintuitive concepts, although response inhibition (simple No-Go accuracy) was marginally related to misconception performance at pre-test. It is noteworthy that response inhibition was also related to control performance, which perhaps indicates that response inhibition is important in solving maths problems of this kind. While the attention in lessons measure derived from observations did not relate to the maths misconception or control problems, it did relate to maths performance in the school tests. This may indicate a more long-term consequence of how much attention is paid in lessons.

Overall the findings from **study B** did not implicate inhibitory control in learning new counterintuitive maths concepts in the classroom, in concert with findings from **study A** and similar studies in science learning (Rhodes et al., 2014, 2016).

## 4.5 General Discussion

The results of **studies A and B** did not confirm the hypotheses that those with better inhibitory control would perform better on misconception questions following a lesson on the counterintuitive concept. Following a lesson on a counterintuitive gravity concept, participants were all able to answer the question correctly, precluding examination of individual differences at post-test and follow up. This may be due to the inclusion of a school where performance is above the national average for this study. There was variation in responses after a lesson on a counterintuitive pressure concept, but inhibitory control was not related to performance. Following a lesson on a counterintuitive maths concept relating to percentage change, misconception performance dipped and then returned to pre-test levels. Inhibitory control did not relate to misconception performance. It is possible that the studies would have shown a modulating role of inhibitory control in classes that were generally more disruptive and less focussed, where inhibitory control requirements would have been greater to enable concentration on the task at hand. While the teachers ensured that students were taught material that had not previously been covered at school, it is also possible that participants spent some time outside of the classroom learning more about the concepts that were addressed.

The studies were a first attempt to bring the research on inhibitory control and counterintuitive reasoning into the classroom to assess the relevance of the research to normal classroom learning. The process of working with teachers to decide on the best concepts to address was useful in ensuring that the concepts were relevant to school learning. It also meant that the teachers were enthusiastic, interested, and invested in the results. However, the sample sizes were determined by the number of teachers who volunteered to take part in the studies, and thus were not based on formal power analyses. Small sample size is one explanation for the lack of effects observed, given the results from **Chapter 3** that showed inhibitory control to relate to misconception problems. In addition, the age of the participants was determined by the classes that the teachers taught, rather than for theoretical reasons, and therefore changed between studies.

In order to make the physics and maths tests as close to normal classroom tests as possible, pen and paper tests were given. This meant that the experience for participants was as close to normal learning and testing as possible, but had the additional consequence that there was no RT data for those tests. It is possible that RT measures of performance may have been related to inhibitory control, but that these associations were undetectable in the data collected. It has been suggested that inhibitory control is involved in moment-by-moment decision making processes in science (Tolmie, Ghazali, & Morris, 2016) that would not be well captured in the measures used here. This example highlights a key struggle in educational neuroscience research: the trade-off between a design that is ecologically valid (in this case, the inclusion of a paper and pen test) and a design that can reveal more about underlying cognitive processes (such as a computerised test that gathers RTs). Neither type of study is able to tell us everything, which is why a mixture of designs is the best approach to understand more about underlying processes and their relevance to education. Nonetheless, while RT was important in the **Chapter 3**, accuracy was too, and the lack of association in these studies indicates that inhibitory control does not appear to relate to misconception performance.

Overall, the studies described here did not convincingly implicate inhibitory control in performance on new counterintuitive classroom concepts. While these findings were contrary to the hypotheses and previous research into behavioural associations (e.g. **Chapter 3**; Gilmore, Keeble, Richardson, & Cragg, 2015; Khng & Lee, 2009; Zaitchik, Iqbal, & Carey, 2014) they do fit with the studies into learning of new science concepts outside the context of misconceptions (Rhodes et al., 2014, 2016). It is possible that inhibitory control is important when reasoning about counterintuitive concepts that have already been learnt, but of less importance when learning new counterintuitive concepts. This would be contrary to the suggestion that old theories must be inhibited for new information to be learnt (Mareschal, 2016). The studies took a first step in bringing the research question to the classroom, in an initial exploratory manner. To improve on the studies reported here, future research could attempt larger sample sizes, and inclusion of a broader range of misconceptions to uncover more about the role of different executive functions.

Having used a behavioural study and two classroom studies to explore the role of inhibitory control in science and maths, the thesis next uses fMRI to examine the same associations. Specifically, **Chapter 5** will explore the neural correlates of counterintuitive science and maths reasoning, and the extent to which these brain regions overlap with those recruited during inhibitory control.

# Chapter 5 FMRI analysis 1: Neural correlates of inhibitory control and science and maths reasoning

## 5.1 Overview

Behavioural research has indicated that both response and semantic inhibition contribute to counterintuitive reasoning in adolescence. Adult research has shown recruitment of PFC regions during counterintuitive science reasoning. This has been interpreted as reflecting inhibition processes, but as yet there is no evidence for overlapping activation of counterintuitive reasoning and inhibitory control within the same participants. In the current fMRI study, 34 adolescents (aged 11-15 years) answered science and maths problems and completed response (simple and complex Go/No-Go) and semantic (numerical Stroop) inhibition tasks. Increased BOLD signal was observed in parietal (BA 40) and prefrontal (BA 8, 45) regions in misconception problems compared to control problems, where no counterintuitive reasoning was required. Both response and semantic inhibition BOLD signal overlapped with misconception-specific activation in the IPL and frontal gyri, but there was also activation unique to counterintuitive reasoning. These novel results provide neural evidence that inhibitory control is one of the mechanisms supporting counterintuitive science and maths reasoning in adolescence.

## 5.2 Introduction

The behavioural results from **Chapter 3** showed that both response and semantic inhibition were associated with counterintuitive science and maths reasoning in adolescence. Inhibitory control predicted unique variance in performance on misconception problems, indicating that it may be implicated in counterintuitive reasoning specifically rather than in general reasoning in science and maths. This finding provided further support for the theory that inhibitory control enables the suppression of prior knowledge, misleading perceptual cues, and naïve theories (Houdé, 2000; Mareschal, 2016). The neural correlates of reasoning about misconceptions have not previously been studied in adolescents, but there is neuroimaging research with adults. Two studies examined the neural underpinnings of reasoning about two scientific misconceptions and found that experts compared to novices recruited brain regions associated with inhibitory control: dorsolateral and ventrolateral PFC (BA 9/10/45/46/47) and ACC (BA 32) (Brault Foisy et al., 2015; Masson et al., 2014). Similarly, in maths, adults showed greater PFC (BA 47) activation when reasoning about counterintuitive stimuli relating to perimeter compared to when reasoning about non-counterintuitive stimuli (Stavy & Babai, 2010). In both cases, the authors argued that the recruitment of these brain areas indicates that participants were using inhibitory control in order to inhibit misconceptions.

However, this interpretation relies on reverse inference since inhibitory control brain activation was not captured in these studies. It is possible that within the set of participants tested, inhibitory control would not recruit those specific brain regions, and the activations reflect other processes. In addition, these tasks showed repeated stimuli referring to the same misconceptions throughout, which means that the results may not accurately demonstrate the brain regions involved in science and maths reasoning that occurs in real life, where different problems are encountered. Responses may have become automatized in those studies, with little reasoning occurring during observation of the stimuli, particularly after viewing many trials. Further, from these results it is not possible to determine whether these activations reflect response or semantic inhibition, if indeed they do show inhibition. The current study therefore sought to examine counterintuitive science and maths reasoning about a range of misconceptions relevant to school learning, and to directly compare the neural activation with response and semantic inhibitory control activation.

Adolescents completed versions of the science and maths misconceptions task and inhibitory control tasks that were adapted from **Chapter 3**, inside an fMRI scanner. The first aim was to establish the BOLD signal associated with reasoning about science and maths misconceptions. The second aim was to assess the brain regions associated with response and semantic inhibition. Previous research has shown response and semantic inhibition to recruit regions including the DLPFC (BA 44), the IFG (BA 48), and the ACC (BA 24/32) (e.g. Adleman et al., 2002; Tamm, Menon, & Reiss, 2002). The third aim was to establish the extent to which brain regions associated with counterintuitive reasoning overlapped with those associated with inhibitory control. Thus, the study was novel in examining adolescents and comparing inhibition and counterintuitive reasoning activation within participants. It was hypothesised that neural overlap would be observed in regions commonly associated with inhibitory control, in particular the lateral PFC (9/45/46) and the ACC (BA 32).

## 5.3 Methods

### 5.3.1 Participants

Thirty-eight pupils (20 girls, 18 boys) aged 11 to 15 years, with no known neurological or developmental disorders, from a range of schools in different demographic areas, took part. Flyers were sent from schools to parents of pupils in Years 7 to 10, inviting their children to take part. Written informed parental and participant consent was obtained in accordance with the guidelines of the local ethics committee, which approved the study. Participants were given pictures of their brain and £20 for participation, and travel expenses for participants and their parents were reimbursed.

To control for general cognitive ability in the behavioural analyses, the WASI Vocabulary and Matrix Reasoning subtests were administered as in **Chapter 3**. One participant was excluded due to low accuracy in the science and maths misconceptions task (15y girl). Three participants were excluded due to movement: one in both the science and maths misconceptions task and the Go/No-Go (12y girl), one in the Go/No-Go (15y girl), and one in the Stroop (12y boy). Two participants had just one run excluded due to movement in the science and maths misconceptions task (12y girl, 15y boy) so were kept in the analysis, discarding the concerning run. Exclusionary criteria relating to movement are described in detail below. The final sample consisted of 34 participants, of which 17 were girls and 17 were boys (**Table 5.1**).

**Table 5.1.** Participant characteristics of final sample, $N = 34$. There was no correlation between age in months and standardised WASI scores (Wechsler, 2011), $p$'s > .2, and no gender difference in mean age, $p = .8$.

| **Variable** | ***Range*** | ***M (SD)*** |
|---|---|---|
| Age in months | 137-185 | 161 (16) |
| WASI Vocabulary raw score | 31-43 | 38 (3) |
| WASI Matrix Reasoning raw score | 15-26 | 22 (3) |

### 5.3.2 Tasks

#### 5.3.2.1 Science and maths misconceptions

The science and maths misconceptions task was adapted from **Chapter 3** to be suitable for the scanner (see Method section 2.1 for details). Participants were presented with science and maths statements, and judged whether they were true or false. Stimuli appeared in four alternating blocks of separate science and maths trials. Between trials, participants were presented either with a fixation cross, or a simple arrows task that acted as an active baseline. A practice outside the scanner included three science and three maths control problems that were not repeated in the main trial, with no performance threshold.

#### 5.3.2.2 Inhibitory control

The Go/No-Go and numerical Stroop were adapted from **Chapter 3** for the scanner (see Method section 2.2 for details). Both tasks had a block design. In the Go/No-Go, there were blocks of Go trials only, blocks of simple Go/No-Go trials, and blocks of complex Go/No-Go trials. In the Stroop there were congruent blocks and mixed blocks. Practices were completed outside the scanner. Go/No-Go practice blocks of 10 trials were repeated if more than one No-Go error was made in the Simple and Complex blocks, and if more than one Go error was made in Go blocks. A Stroop familiarisation phase of eight trials of asterisks instead of digits was repeated if more than one error was made. A Congruent practice of eight trials was repeated if more than one error was made. Finally, a practice Mixed block of eight trials was repeated if more than two errors were made.

### 5.3.3 Procedure

The testing session lasted a total of 2 hours. Participants first practiced each of the three tasks outside the scanner to ensure the instructions were understood. Participants then completed the science and maths misconceptions task inside the scanner, followed by a structural scan for 5.5 min, then the Go/No-Go task and finally the numerical Stroop task. Overall, participants spent approximately 50 min inside the scanner. Additional behavioural tasks were administered outside the scanner and will be elaborated on in **Chapter 6**.

### 5.3.4 Behavioural data analysis

For the science and maths misconceptions task mean RTs were calculated from all trials. Two (Discipline: science, maths) x two (Trial type: control, misconception) repeated measures ANCOVAs covarying for z-scored age in months were run on accuracy and RT. For the inhibitory control tasks mean RTs were calculated from correct trials only. Repeated measures ANCOVAs investigated the difference between Trial types in the Go/No-Go task (Go in Go blocks, Go in Simple blocks, Go in Complex blocks, No-Go in Simple blocks, No-Go in Complex blocks) and the numerical Stroop task (congruent in Congruent blocks, congruent in Mixed blocks, incongruent in Mixed blocks) for accuracy and RT (for Go trials only in Go/No-Go task), covarying for z-scored Age in months. Main effects were followed up with Bonferroni-corrected pairwise comparisons.

Hierarchical multiple regressions investigated the extent to which inhibitory control variables account for individual differences in science and maths misconception accuracy and RT in a replication of **Chapter 3**, albeit with a much smaller sample. As in **Chapter 3**, raw scores are used for the WASI tasks so that age is controlled for only once (through entering age separately), and so that changes with age in the number of correct responses can be examined. Finally, standardised scores were not of particular interest in this study where participants had no known developmental disorders. Block 1 variables were inserted using the enter method: age in months, WASI Vocabulary and Matrix Reasoning raw scores, and science and maths control performance (accuracy or RT). Block 2 contained Go/No-Go and Stroop variables entered stepwise: Simple Go RT cost (Simple Go RT – RT in Go blocks), Complex Go RT cost (Complex Go RT – RT in Go blocks), Stroop accuracy cost (congruent – incongruent) and Stroop RT cost (incongruent – congruent), where congruent trials included those in both congruent and mixed blocks. RT cost measures were used in the Go/No-Go since high accuracy was anticipated in this version of the Go/No-Go where there are now 50% No-Go trials. This would be in line with previous research showing no age differences in accuracy where 50% No-Go trials were used in a study of 8- to 20-year-olds (Tamm et al., 2002). It was considered that RT cost would therefore better reflect response inhibition.

### 5.3.5 MRI data acquisition and preprocessing

Brain imaging data were acquired on a 1.5 Telsa Siemens Avanto MRI scanner with a 30-channel head coil. Structural data were acquired with a T1-weighted magnetization-prepared rapid gradient-echo (MPRAGE) with 2x generalised autocalibrating partial parallel acquisition (GRAPPA) acceleration, lasting 5.5 min. Functional data were acquired in six sessions using the Centre for Magnetic Resonance Research (CMRR) multiband echo-planar imaging (EPI) sequence (Xu et al., 2013) 4x acceleration, leak block on (Cauley et al., 2014) repetition time (TR) = 1 s, echo time (TE) = 45 ms, comprising 44 slices covering most of the cerebrum, with a resolution of 3 x 3 x 3 $mm^3$. There were six functional runs, with the structural MPRAGE typically acquired between the science and maths misconceptions task and the inhibitory control tasks.

MRI data were preprocessed and analysed using SPM12 (www.fil.ion.ucl.ac.uk/ spm/software/spm12/). Functional images were realigned to the mean images after the first realignment in a two pass procedure using a second-degree B-spline interpolation to correct for movement during the session. The bias-field-corrected structural image was co-registered to the mean realigned functional image, and segmented on the basis of MNI registered International Consortium for Brain Mapping tissue probability maps. Resulting spatial normalisation parameters were applied to the realigned images to obtain normalised functional images with a voxel size of 3 x 3 x 3 mm, which were smoothed with an 8 mm full-width at the half-maximum Gaussian kernel. FD was calculated for each volume as a scalar measure of head motion across the six realignment estimates (Siegel et al., 2014). Volumes with an FD greater than 0.9 mm were censored and excluded from the GLM estimation by including a regressor of no interest for each censored volume. Scanning runs with more than 15% of volumes censored or a root mean square movement greater than 1.5 mm were excluded from the analysis.

### 5.3.6 FMRI data analysis

Scanning runs were treated as separate time series, each of which was modelled by a set of regressors in the GLM. All regressors were convolved with a canonical haemodynamic response function (HRF) and, together with the separate regressors representing each censored volume and the session mean, comprised the full model for each session. All coordinates are given in MNI space, region labelling was completed with AAL (Tzourio-Mazoyer et al., 2002), and BA labelling was completed with MRIcron (Rorden & Brett, 2000).

Science and maths misconceptions task data in each of the four runs were modelled by box-car regressors separately representing Misconception and Control trials using each trial's RT as the duration, and arrows task blocks, using 16 s minus each preceding Misconception or Control trial's RT as the duration. A first-level contrast of the difference between Misconception and Control trials was calculated. Science and maths conditions were collapsed as there was no difference between science and maths in the Misconception > Control contrast.

In the Go/No-Go task, box-car regressors modelled Go blocks, Simple Go/No-Go blocks, Complex Go/No-Go blocks (22 s duration), and fixation blocks (10 or 15 s duration) separately. First-level contrasts between Simple Go/No-Go and Go blocks, and Complex Go/No-Go and Go blocks were calculated. In the numerical Stroop task, box-car regressors modelled Congruent blocks and Mixed blocks (21 s duration) and fixation blocks (10 s or 15 s duration) separately. First-level contrasts between Mixed and Congruent blocks were calculated. For both tasks there was an additional single event-related regressor of duration zero for all errors.

First-level contrasts were entered into one sample *t*-tests to create SPM maps which were thresholded at $p < .001$ uncorrected at the voxel level and at FWE-corrected $p < .05$ at the cluster level. Voxels surviving a voxel level FWE-corrected $p < .05$ are also reported.

In order to identify overlapping activations inclusive masking was used to identify brain areas of increased BOLD signal in the Misconception > Control contrast and in either of the Simple Go/No-Go > Go, Complex Go/No-Go > Go, and Mixed > Congruent numerical Stroop contrasts, using the same statistical threshold of $p < .001$ at the voxel level and FWE-corrected $p < .05$ at the cluster level. MNI coordinates and cluster size of overlapping regions were obtained using MarsBaR (Brett et al., 2002).

In exploratory correlation analyses, age effects were examined in each contrast through one-sample *t*-tests entering age in months as a covariate. Correlations of individual differences in BOLD signal were run between inhibitory control activations (Complex Go/No-Go > Go, Stroop Mixed > Congruent) and science and maths activation (Misconception > Control) in each overlapping region identified, averaging the data over each cluster using MarsBaR. Associations between BOLD signal and behavioural measures were investigated through a Misconception > Control one sample *t*-test, entering performance measures from the Go/No-Go (Simple Go RT cost, Complex Go RT cost), numerical Stroop (RT cost, accuracy cost), and science and maths misconceptions accuracy cost (control – misconception) and RT cost (misconception – control) as covariates.

# 5.4 Results

## 5.4.1 Behavioural results

### 5.4.1.1 Science and maths misconceptions

The repeated measures ANCOVA on accuracy (**Table 5.2**) showed significant main effects of Trial type and Discipline, with higher accuracy in control trials compared to misconception trials and in science compared to maths. There was an interaction between Trial type and Discipline, whereby the difference between science and maths was significant for misconception trials, $F(1, 32) = 13.20$, $p = .001$, $\eta_p^2 = .292$, but not for control trials, $p = .5$.

**Table 5.2.** Accuracy and RT estimated marginal means in the science and maths misconceptions task.

| | | Accuracy (%) | RT (ms) |
|---|---|---|---|
| | | ***M (SE)*** | ***M (SE)*** |
| **Main effects** | | | |
| *Trial type* | | $F(1, 32) = 324.30$, $p < .001$, $\eta_p^2 = .910$ | $F(1, 32) = 228.80$, $p < .001$, $\eta_p^2 = .877$ |
| Control | | 85.8 (1.2) | 5448 (130) |
| Misconception | | 59.9 (1.6) | 6515 (147) |
| *Discipline* | | $F(1, 32) = 11.25$, $p = .002$, $\eta_p^2 = .260$ | $F(1, 32) = 24.60$, $p < .001$, $\eta_p^2 = .435$ |
| Science | | 75.1 (1.4) | 5726 (143) |
| Maths | | 70.6 (1.4) | 6237 (144) |
| **Interaction effects** | | | |
| *Trial type* | *Discipline* | $F(1, 32) = 4.22$, $p = .048$, $\eta_p^2 = .116$ | n.s., $p = .465$ |
| Control | Science | 84.8 (0.9) | 5176 (145) |
| | Maths | 79.6 (1.1) | 5721 (133) |
| Misconception | Science | 60.2 (1.1) | 6276 (153) |
| | Maths | 49.3 (1.4) | 6754 (166) |

There was a main effect of Age, $F(1, 32) = 8.06$, $p = .008$, $\eta_p^2 = .201$, which was modulated by a significant interaction between Trial type and Age, $F(1, 32) = 4.19$, $p = .049$, $\eta_p^2 = .116$. Follow-up Pearson correlations indicated that there was a significant positive correlation between Age and accuracy in misconception trials, $r = .47$, $p = .005$, while the correlation between Age and accuracy in control trials was not significant, $p = .095$ (**Figure 5.1**).

The repeated measures ANCOVA on RT (**Table 5.2**) showed significant main effects of Trial type, and Discipline, with faster RTs in control trials compared to misconceptions trials and in science compared to maths.

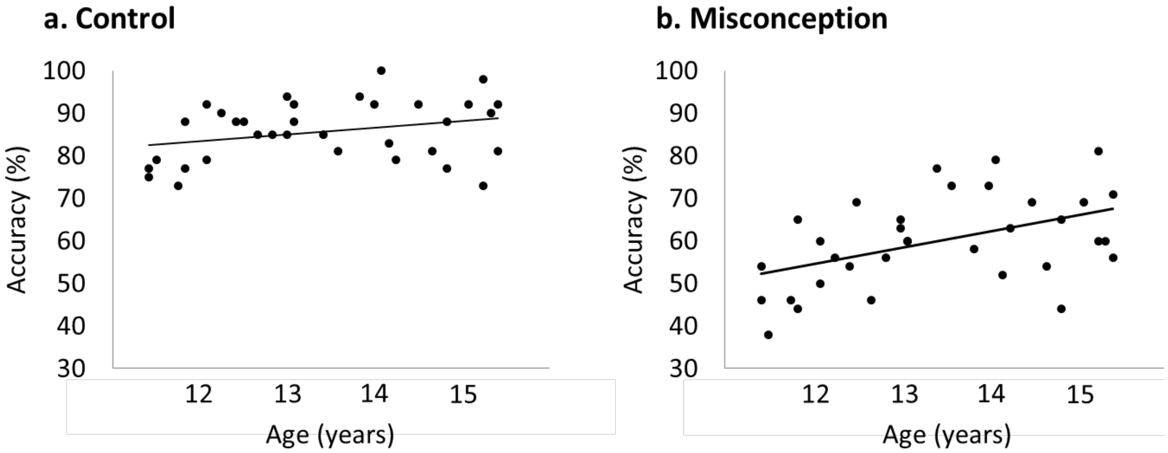



**Figure 5.1.** Accuracy as a function of age in the science and maths misconceptions task. Scatterplots displaying correlation between age and percentage accuracy in (a) control trials and (b) misconception trials.

There was no main effect of age, $p = .7$, but there was a significant interaction between Trial type, Discipline, and Age, $F(1, 32) = 4.92$, $p = .034$, $\eta_p^2 = .133$, which was followed up with ANCOVAs and Pearson correlations. The interaction between Discipline and Age was significant in control trials, $F(1, 32) = 6.481$, $p = .016$, $\eta_p^2 = .168$, with no association between Age and science control trials, $r = .07$, $p = .7$ and a non-significant negative correlation between Age and maths control trials, $r = -.25$, $p = .2$. The interaction between Discipline and Age was not significant in misconception trials, $p = .7$.

### 5.4.1.2 Inhibitory control

The repeated measures ANCOVA on accuracy in the Go/No-Go showed a significant effect of Trial type (**Table 5.3**). Bonferroni-corrected pairwise comparisons showed significant differences ($p$'s < .028) between all trial types apart from between Complex Go and Complex No-Go trials, $p = .5$. Accuracy was highest in Simple No-Go trials, followed by Go trials in Go blocks, then Simple Go trials, Complex No-Go trials, and finally Complex Go trials. There was a main effect of Age, $F(1, 32) = 5.97$, $p = .020$, $\eta_p^2 = .157$, where accuracy improved with age, and there was no interaction between Trial type and Age, $p = .87$. For RT, there was a significant effect of Trial type (**Table 5.3**) with significant differences between all trial types, $p$'s < .001. RTs were fastest in Go blocks, then Simple Go trials, and slowest in Complex Go trials. There was no significant effect of Age and no interaction, $p$'s > .6.

**Table 5.3.** Accuracy and RT estimated marginal means in the inhibitory control tasks.

| | **Accuracy (%)** | **RT (ms)** |
|---|---|---|
| | ***M (SE)*** | ***M (SE)*** |
| **Go/No-Go** | $F(4, 128) = 20.90$, $p < .001$, $\eta_p^2 = .395$ | $F(2, 64) = 130.62$, $p < .001$, $\eta_p^2 = .803$ |
| *Go trials* | | |
| Go blocks | 93.1 (0.8) | 429 (6) |
| Simple blocks | 92.8 (1.2) | 470 (6) |
| Complex blocks | 84.7 (1.5) | 507 (7) |
| *No-Go trials* | | |
| Simple blocks | 95.5 (0.7) | - |
| Complex blocks | 87.9 (1.4) | - |
| **Numerical Stroop** | $F(2, 64) = 74.41$, $p < .001$, $\eta_p^2 = .699$ | $F(2, 64) = 301.11$, $p < .001$, $\eta_p^2 = .904$ |
| *Congruent trials* | | |
| Congruent blocks | 89.6 (0.9) | 634 (11) |
| Mixed blocks | 93.4 (1.2) | 693 (11) |
| *Incongruent trials* | | |
| Mixed blocks | 76.0 (1. 9) | 787 (11) |

The repeated measures ANCOVA on accuracy (**Table 5.3**) in the numerical Stroop showed a significant effect of Trial type, with significant differences between all trial types, *p*'s < .01. The highest accuracy was in Congruent trials in Mixed blocks, followed by Congruent trials in Congruent blocks, and finally Incongruent trials. There was no main effect of Age, $p = .060$. For RT there was a significant effect of Trial type (**Table 5.3**) with significant differences between all trial types, *p*'s < .001. RTs were fastest in Congruent trials within Congruent blocks, then Congruent trials in Mixed blocks, then Incongruent trials. There was no significant effect of Age and no interaction, *p*'s > .5.

### 5.4.1.3 Regression analyses

The regression model for accuracy was significant, explaining 50% of the variance in science and maths misconception accuracy (**Table 5.4**). Age, WASI Vocabulary, WASI Matrix Reasoning, and science and maths control accuracy were entered into the model but none were significant predictors of performance on their own. The model for RT was significant, explaining 77% of the variance, and only RT for control trials was a significant predictor of performance (**Table 5.4**).

**Table 5.4.** Regression models for science and maths misconception accuracy and RT. Significant predictors (*p* < .05) are highlighted in bold. β = standardised coefficients.

| Dependent variables | Independent variables | β | *t* | *p* |
|---|---|---|---|---|
| Science and maths misconception accuracy | | | | |
| Model 1<br>*F*(4, 29) = 7.12,<br>*p* < .001, $R^2$ = 50% | **Constant** | | **-2.48** | **.019** |
| | Age (months) | .26 | 1.69 | .101 |
| | WASI Vocabulary raw | .30 | 1.78 | .086 |
| | WASI Matrix Reasoning raw | .10 | 0.68 | .502 |
| | Science and maths control accuracy | .31 | 1.84 | .077 |
| Science and maths misconception RT | | | | |
| Model 2<br>*F*(4, 29) = 24.87,<br>*p* < .001, $R^2$ = 77% | Constant | | 1.06 | .297 |
| | Age (months) | .05 | 0.50 | .622 |
| | WASI Vocabulary raw | -.06 | -0.62 | .543 |
| | WASI Matrix Reasoning raw | -.02 | -0.19 | .851 |
| | **Science and maths control RT** | **.86** | **8.93** | **< .001** |

### 5.4.1.4 Summary of behavioural results

In summary, behavioural results showed that, as expected, performance was worse in misconception trials than control trials, worse in Complex Go/No-Go blocks than in Simple Go/No-Go blocks and in these than in Go blocks – although accuracy was high in No-Go trials themselves – and worse in incongruent trials and mixed numerical Stroop blocks than in congruent blocks. The only observed associations with age were an increase in misconception accuracy and an increase in overall Go/No-Go task accuracy. The regression analyses did not demonstrate the anticipated associations between inhibitory control and science and maths misconception performance.

## 5.4.2 Neuroimaging results

### 5.4.2.1 Science and maths misconceptions

The contrast Misconception > Control (**Table 5.5, Figure 5.2**) showed increased BOLD signal in bilateral supramarginal gyrus (BA 40), extending into the IPL and AG, as well as in superior and middle frontal gyri (predominantly BA 8, extending into BA 9) and middle and inferior frontal gyri (BA 45, 47, 11). The right hemisphere activation extended along both dorsolateral and ventrolateral PFC along BA 45 (**Figure 5.2**). There was an additional small cluster in the left lingual gyrus.

**Table 5.5.** Brain activation in science and maths misconception trials compared to control trials. [a,b] indicate $p_{FWE} < .05$ at the voxel level, and $p_{FWE} < .05$ at the cluster level (cluster defining threshold: $p_{uncorr} < .001$), respectively; L, R indicate left and right hemispheres respectively.

| **Brain region** | **L/R** | **BA** | **MNI** | | | ***Z*-score** | **Cluster size** |
|---|---|---|---|---|---|---|---|
| | | | ***x*** | ***y*** | ***z*** | | |
| Supramarginal gyrus | R | 40 | 60 | -31 | 50 | 4.94[a] | 527[b] |
| Lingual gyrus | L | 37 | -27 | -49 | -7 | 4.67[a] | 26 |
| Superior frontal gyrus | R | 8 | 18 | 20 | 62 | 4.61[a] | 220[b] |
| Middle frontal gyrus | R | 9 | 36 | 17 | 59 | 4.13 | |
| Supramarginal gyrus | L | 40 | -63 | -28 | 44 | 4.58[a] | 163[b] |
| Inferior parietal lobule | L | 40 | -57 | -37 | 47 | 3.99 | |
| Middle frontal gyrus | L | 8 | -36 | 14 | 62 | 4.56[a] | 176[b] |
| Superior frontal gyrus | L | 8 | -21 | 26 | 62 | 3.73 | |
| Middle frontal gyrus | R | 45 | 45 | 44 | 11 | 4.49 | 337[b] |
| Inferior frontal gyrus (orbital) | R | 47 | 42 | 47 | -10 | 4.32 | |
| Middle frontal gyrus (orbital) | R | 11 | 27 | 41 | -19 | 4.18 | |
| Inferior frontal gyrus (orbital) | L | 47 | -42 | 47 | -10 | 3.68 | 134[b] |
| Inferior frontal gyrus | L | 45 | -45 | 41 | 14 | 3.48 | |
| Inferior frontal gyrus | L | 46 | -45 | 50 | 5 | 3.39 | |

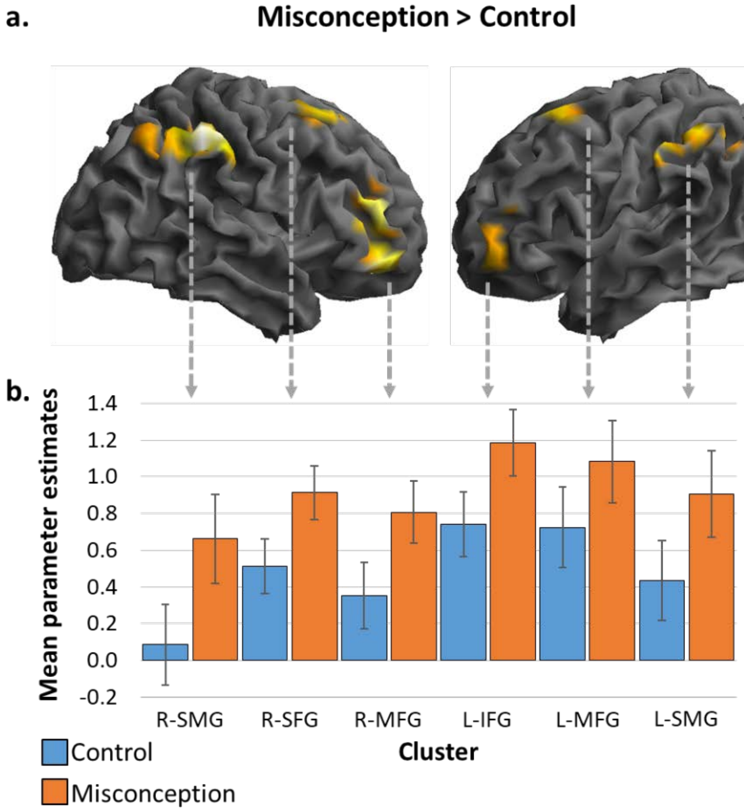



**Figure 5.2.** (a) Regions of increased BOLD signal in the right and left hemispheres in the Misconception > Control contrast of the science and maths task. (b) Mean parameter estimates (± *SE*) in misconception and control trials for the six main clusters (see **Table 5.5**). Zero represents the implicit baseline of the model, which includes the fixation phases. R: right; L: left; SMG: supramarginal gyrus; SFG: superior frontal gyrus; MFG: middle frontal gyrus; IFG: inferior frontal gyrus; contrasts $p_{uncorr} < .001$ at the voxel level and $p_{FWE} < .05$ at the cluster level.

### 5.4.2.2 Inhibitory control

The Simple Go/No-Go > Go contrast showed no region of increased BOLD signal. The Complex Go/No-Go > Go contrast (**Table 5.6**, **Figure 5.3**) showed increased BOLD signal in a large bilateral parietal cluster covering the IPL, superior parietal gyri and precuneus, in a large bilateral frontal cluster covering the middle frontal gyri, precentral gyri, SMA and extending into the right anterior insula, as well as in smaller clusters in the left insula and caudate nucleus, right middle orbital frontal gyrus, and in the cerebellum.

**Table 5.6.** Brain activation in the complex Go/No-Go. [a,b] indicate $p_{FWE} < .05$ at the voxel level, and $p_{FWE} < .05$ at the cluster level (cluster defining threshold: $p_{uncorr} < .001$), respectively; L, R indicate left and right hemispheres respectively.

| Brain region | L/R | BA | MNI | | | *Z*-score | Cluster size |
|---|---|---|---|---|---|---|---|
| | | | ***x*** | ***y*** | ***z*** | | |
| Inferior parietal lobule | R | 40 | 39 | -49 | 50 | 7.20[a] | 2555[b] |
| Inferior parietal lobule | L | 40 | -42 | -46 | 50 | 6.67[a] | |
| Precuneus | R | 7 | 9 | -67 | 56 | 6.58[a] | |
| Precuneus | L | 7 | -6 | -70 | 53 | 6.43[a] | |
| Middle frontal gyrus | R | 45 | 45 | 35 | 35 | 6.66[a] | 3898[b] |
| Supplementary motor area | L | 32 | -3 | 14 | 50 | 6.51[a] | |
| Superior frontal gyrus | R | 6/8 | 33 | 8 | 65 | 6.42[a] | |
| Anterior insula | R | | 33 | 23 | -4 | 6.14[a] | |
| Middle frontal gyrus | R | 46 | 36 | 50 | 20 | 6.14[a] | |
| Middle frontal gyrus | L | 6 | -30 | 2 | 56 | 5.98[a] | |
| Inferior frontal gyrus | R | 44 | 45 | 8 | 26 | 5.39[a] | |
| Precentral gyrus | L | 6 | -42 | 2 | 35 | 5.35[a] | |
| Middle frontal gyrus | L | 46 | -33 | 56 | 17 | 5.06[a] | |
| Insula | L | 48 | -30 | 20 | 5 | 6.17[a] | 191[b] |
| Caudate nucleus | L | | -18 | 2 | 20 | 3.65 | |
| Crus I of cerebellar hemisphere | L | | -33 | -58 | -34 | 6.10[a] | 590[b] |
| Crus II of cerebellar | L | | -6 | -79 | -28 | 5.43[a] | |
| Crus I of cerebellar hemisphere | R | | 36 | -61 | -31 | 6.00[a] | 186[b] |
| Lobule III of vermis | R | | 3 | -43 | -19 | 5.72[a] | 228[b] |
| Middle orbital frontal gyrus | R | 11 | 27 | 44 | -19 | 5.47[a] | 70 |

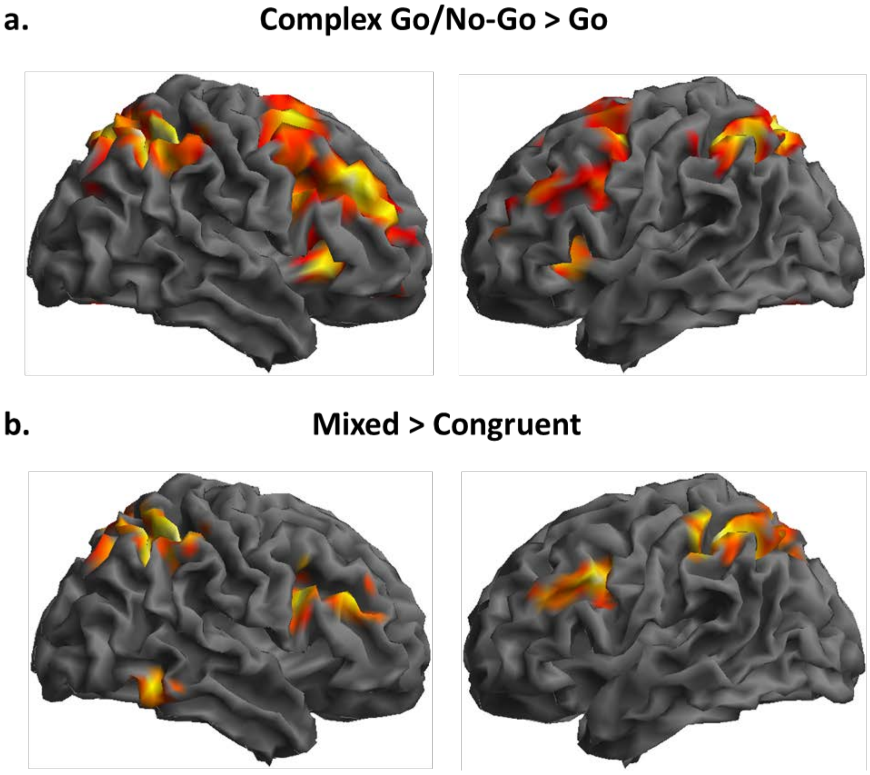



**Figure 5.3.** Regions of increased BOLD signal in the right and left hemispheres in (a) the complex Go/No-Go blocks > Go blocks contrast of the Go/No-Go task and (b) the Mixed blocks > Congruent blocks contrast of the numerical Stroop task. For both contrasts $p_{\text{uncorr}} < .001$ at the voxel level and $p_{\text{FWE}} < .05$ at the cluster level.

The Mixed > Congruent numerical Stroop contrast (**Table 5.7, Figure 5.3**) similarly showed increased BOLD signal in parietal clusters covering the IPL, superior parietal gyri and precuneus, frontal clusters located in the middle/inferior frontal gyri, extending into the precentral gyri, and a cluster in the right inferior temporal gyrus.

**Table 5.7.** Brain activation in numerical Stroop. [a,b] indicate $p_{\text{FWE}} < .05$ at the voxel level, and $p_{\text{FWE}} < .05$ at the cluster level (cluster defining threshold: $p_{\text{uncorr}} < .001$), respectively; L, R indicate left and right hemispheres respectively.

| **Brain region** | **L/R** | **BA** | **MNI** | | | ***Z*-score** | **Cluster size** |
|---|---|---|---|---|---|---|---|
| | | | ***x*** | ***y*** | ***z*** | | |
| Precentral gyrus | L | 44 | -42 | 8 | 32 | 5.89[a] | 517[b] |
| Inferior frontal gyrus | L | 45 | -51 | 29 | 32 | 5.10[a] | |
| Middle frontal gyrus | L | 46 | -48 | 50 | 14 | 4.62[a] | |
| Inferior parietal lobule | L | 40 | -45 | -43 | 56 | 5.88[a] | 1007[b] |
| Inferior parietal lobule | L | 40 | -48 | -37 | 44 | 5.59[a] | |
| Precuneus | L | 7 | -6 | -64 | 50 | 4.81[a] | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Superior parietal gyrus | L | 7 | -27 | -61 | 53 | 4.74[a] | |
| Inferior parietal lobule | R | 2/40 | 48 | -37 | 53 | 5.82[a] | 888[b] |
| Angular gyrus | R | 40/7 | 33 | -52 | 44 | 5.73[a] | |
| Inferior parietal lobule | R | 40 | 39 | -43 | 44 | 5.68[a] | |
| Superior occipital gyrus | R | 7 | 30 | -64 | 41 | 5.30[a] | |
| Inferior frontal gyrus | R | 44 | 45 | 8 | 23 | 5.46[a] | 564[b] |
| Middle frontal gyrus | R | 45 | 51 | 35 | 23 | 4.97[a] | |
| Inferior temporal gyrus | R | 20 | 51 | -52 | -13 | 4.96[a] | 143[b] |

### 5.4.2.3 Overlapping activation

Of the six clusters observed in the Misconception > Control contrast, all except the left IFG showed partial overlap with the Complex Go/No-Go > Go contrast, and all except the left middle and right superior frontal gyri also showed partial overlap with the numerical Stroop Mixed > Congruent contrast (**Table 5.8, Figure 5.4**). However, this overlap was not complete. The network of brain regions showing increased BOLD signal in the inhibitory control tasks was broader (**Figure 5.4**) and part of the increased BOLD signal in the six science and maths clusters was unique to the Misconception > Control contrast (**Figure 5.4**).

**Table 5.8.** Overlapping activation between the inhibitory control tasks contrasts and the science and maths task Misconception > Control contrast. Coordinates are the centre of mass of each cluster as calculated by MarsBaR. L, R, indicate left and right hemispheres respectively.

| **Brain region** | | | **MNI** | | | |
|---|---|---|---|---|---|---|
| | **L/R** | **BA** | ***x*** | ***y*** | ***z*** | **Cluster size** |
| *Overlap between Misconception > Control and Complex Go/No-Go > Go* | | | | | | |
| Superior frontal gyrus | R | 8 | 25 | 15 | 61 | 127 |
| Inferior frontal gyrus | R | 11 | 28 | 44 | -18 | 8 |
| Middle frontal gyrus | R | 45 | 45 | 43 | 21 | 69 |
| Inferior parietal lobule | R | 40 | 49 | -43 | 47 | 333 |
| Middle frontal gyrus | L | 8 | -23 | 13 | 63 | 20 |
| Inferior parietal lobule | L | 40 | -52 | -44 | 49 | 66 |
| *Overlap between Misconception > Control and Mixed > Congruent Numerical Stroop* | | | | | | |
| Middle frontal gyrus | R | 45 | 46 | 41 | 21 | 67 |
| Inferior parietal lobule | R | 40 | 47 | -41 | 48 | 188 |
| Inferior frontal gyrus | L | 45 | -47 | 47 | 9 | 9 |
| Inferior parietal lobule | L | 40 | -53 | -40 | 49 | 58 |

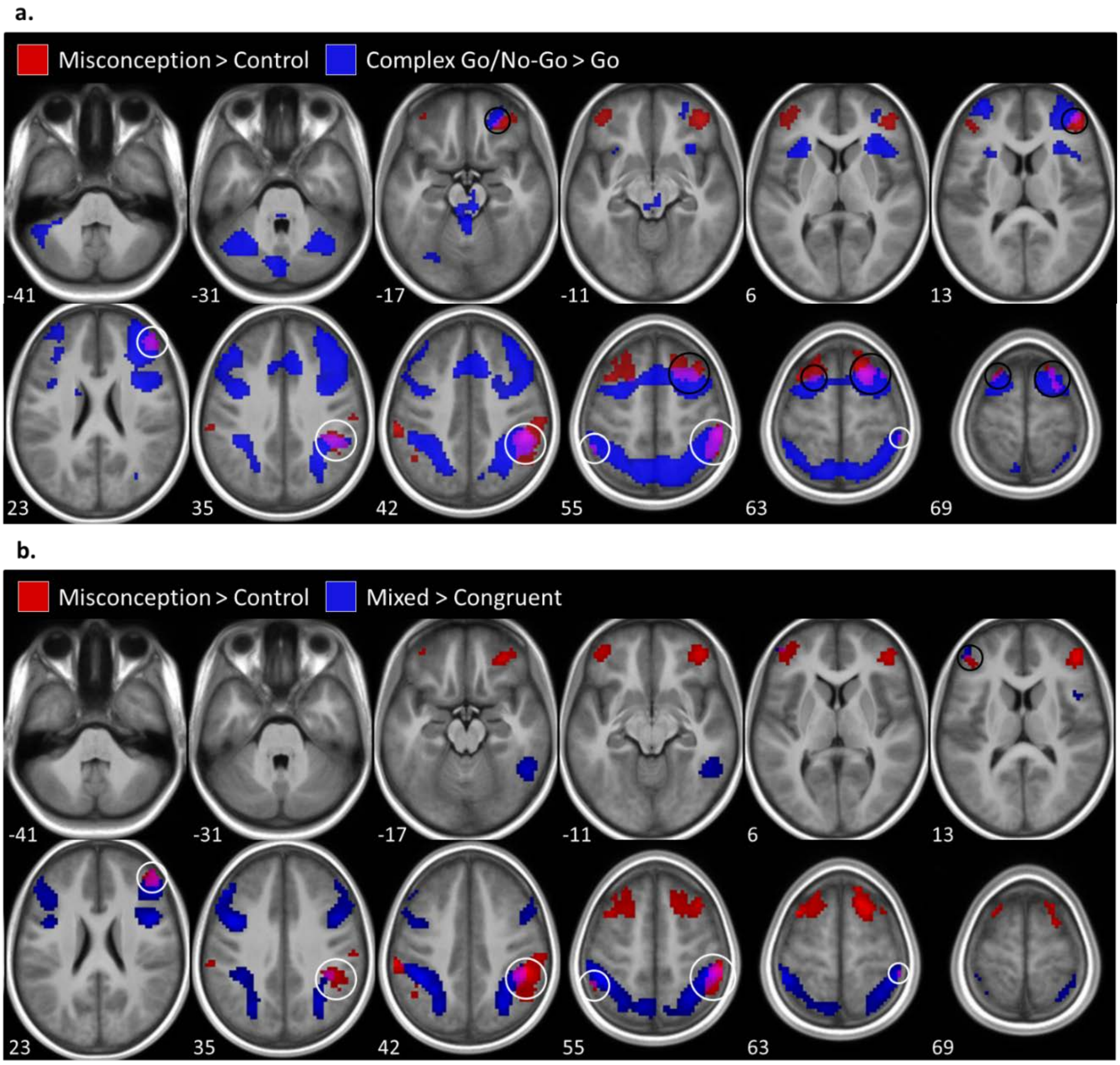



**Figure 5.4.** Overlapping activation between the science and maths and inhibitory control tasks contrasts. BOLD signal in Misconception > Control and (a) Complex Go/No-Go > Go and (b) Mixed > Congruent in the numerical Stroop. Inhibitory control task contrasts are shown in blue, while the science and maths Misconception > Control contrast is shown in red, and regions of overlap are shown in purple. Note that the slices shown are the same in (a) and (b) to enable comparison between images, and the z-coordinate is indicated in the bottom left corner. Black circles highlight regions of overlap between the two contrasts, while white circles highlight regions common to all three contrasts. Contrasts are overlaid using MRIcron onto an image of the mean normalised structural brain image of the 34 participants created using ImCalc in SPM.

### 5.4.2.4 Exploratory correlations

Age in months was not a significant covariate for any contrast. Correlations of individual differences in mean BOLD signal between inhibitory control activations (Complex Go/No-Go > Go, Stroop Mixed > Congruent) and science and maths activation (Misconception > Control) in each overlapping region identified were not significant. No performance measures on any of the tasks were significant covariates for the Misconception > Control contrast (performance measures were Simple Go RT cost, Complex Go RT cost, Stroop RT cost, Stroop accuracy cost, science and maths misconceptions accuracy cost, and science and maths misconceptions RT cost).

## 5.5 Discussion

The current study investigated the neural mechanisms of counterintuitive reasoning in adolescence and the overlap with inhibitory control. It was hypothesised that there would be overlap in the lateral PFC (9/45/46) and the ACC (BA 32). This was the first study to provide a direct comparison of the neural correlates of counterintuitive reasoning and inhibitory control within the same participants, and is unique in its inclusion of a broad range of science and maths topics. It was also the first study to address this research question with adolescents, rather than using an adult sample. The results showed some overlap but also considerable specificity of the neural correlates of counterintuitive reasoning and inhibitory control.

Analysis of the behavioural data showed better performance in control trials than misconception trials, Complex than Simple Go/No-Go trials, and incongruent than congruent Stroop trials, as anticipated. Accuracy improved in misconception but not control trials with age, and in the Go/No-Go but not the numerical Stroop with age. The regression analyses did not replicate the results from **Chapter 3** which may be due to the smaller sample size in the present study.

The comparison of activation in trials involving counterintuitive reasoning to those with no counterintuitive reasoning showed increased BOLD signal in parietal and frontal regions, which overlapped to some extent with BOLD signal in both response inhibition in the context of a small working memory load (complex Go/No-Go) and semantic inhibition (numerical Stroop). Parietal overlap may be related to common requirements for attentive control and responding to alerting stimuli (Singh-Curry & Husain, 2009), which may be greater in counterintuitive problems and not related to inhibition per se. Prefrontal overlap suggests common executive processes in inhibitory control and counterintuitive reasoning in science and maths, supporting adult neuroimaging findings of increased activation in inhibitory control brain regions during counterintuitive science reasoning (Brault Foisy et al., 2015; Masson et al., 2014). For the main contrast of interest (Misconception > Control), there was no difference between science and maths, indicating that counterintuitive reasoning requires domain-general processes that are not specific to reasoning or knowledge within the discipline.

Despite some overlap, the activation for science and maths counterintuitive reasoning was not fully aligned with inhibitory control activation. In part this was due to the broad network of inhibitory control regions, which formed some large clusters that extend away from counterintuitive reasoning areas. However, there was also activity specific to counterintuitive reasoning that was not present during the performance of response or semantic inhibition tasks. Although previous neuroimaging research has predominantly focussed on the involvement of inhibitory control in counterintuitive science and maths reasoning, these results highlight the need to explore other factors, beyond inhibitory control, which may play a role, including relational reasoning (see **Chapters 6 and 7**). This novel finding also highlights the importance of directly examining the brain activation associated with the cognitive function assumed to explain activity in a different task, rather than to rely on reverse inference. It is also worth considering the potential role of the multiple-demand (MD) network, a system which reflects cognitive challenge (Duncan, 2010). It is possible that these overlapping regions are indicative of increased challenge as opposed to common inhibition demands.

The Simple Go/No-Go blocks, when contrasted with Go blocks, showed no increased BOLD signal. While 50% No-Go trials are common in fMRI studies of inhibitory control to increase the proportion of trials requiring inhibitory control per task block (e.g. Tamm et al., 2002), this high percentage may have meant that the task became too simple to demonstrate strong response inhibition effects. It is therefore not clear that it is only response inhibition in the context of cognitive load that overlaps with counterintuitive reasoning, since the purer response inhibition measure appears to have been watered down by the task design. However, the possible interpretation that response inhibition combined with cognitive load contributes to counterintuitive reasoning is consistent with the behavioural study reported in **Chapter 3** using 25% No-Go trials, whereby No-Go accuracy in the complex Go/No-Go task, but not in the simple Go/No-Go task, was associated with counterintuitive reasoning performance.

In contrast to the adult research and the hypothesis, increased BOLD signal was not observed in the ACC on any task. ACC activation is thought to reflect error-processing and underlie the transition from adolescent to adult-level inhibitory control performance (Ordaz et al., 2013). This may also be attributable to the block design of the inhibitory control tasks, leading to a reduction in ACC BOLD signal that is specific to trial by trial conflict and error monitoring. Nonetheless, ACC activation was not observed in the counterintuitive science and maths reasoning contrast, which did not follow a block design. There were also no age effects in the neuroimaging data for any task, where previous research has shown decreased recruitment of PFC with age in inhibitory control (Luna, Marek, Larsen, Tervo-Clemmens, & Chahal, 2015).

While a range of schools were approached for recruitment, with participants from both state and fee-paying independent schools, it is likely that the sample is not entirely representative of the adolescent population. Those who opted to take part may have been those with parents who have a particular interest in science and maths, or who would encourage their children to take part in a research study outside of school. It is possible that such parents would be from a higher socioeconomic bracket. It is also likely that the adolescents who volunteered perform relatively well at school science and maths; those who strongly dislike or perform poorly in science and maths at school may have decided not to do extra science and maths in their spare time. Overall, this may have meant that the results are less generalizable to the adolescent population as a whole, potentially masking different effects that would arise in a sample with wider variation in results.

In sum, the neuroimaging results provide some further support for the theory that inhibitory control enables suppression of naïve theories and distracting information, supporting counterintuitive reasoning across a wide range of science and maths topics in adolescence. The existence of overlapping regions in part supports previous neuroimaging studies of reasoning about misconceptions (Brault Foisy et al., 2015; Masson et al., 2014), but there are also distinct prefrontal and parietal patterns of activation during counterintuitive reasoning.

The thesis next considers a second candidate cognitive process that may support science and maths problem-solving: relational reasoning. **Chapter 6** will examine the extent to which individual differences in behaviour and brain activation during science and maths associates with analogical and non-verbal relational reasoning. In particular, the analysis is concerned with examining these associations over and above the possible role of executive functions.

# Chapter 6 FMRI analysis 2: Relational reasoning and the neural correlates of science and maths reasoning

## 6.1 Overview

Relational reasoning is a domain-general skill that matures through adolescence and is thought to be involved in science and maths reasoning. The unique contributions of verbal analogical reasoning and non-verbal matrix reasoning to science and maths performance are not well understood. In the current fMRI analysis, 11- to 15-year-olds (*N* = 34) answered science and maths problems inside the scanner, and took tests of relational reasoning outside the scanner. Better verbal analogical reasoning was associated with better science and maths accuracy and faster RTs, while non-verbal matrix reasoning was associated only with faster science RTs. More fine-grained analysis showed that analogical reasoning predicted science but not maths performance when other contributing factors were controlled for. Science and maths reasoning led to broad activation of bilateral networks in regions including the occipital lobe (BA 17, 18, 19), the inferior temporal gyrus (BA 37), hippocampus (BA 27), IPL (BA 7, 40, 2) and inferior frontal gyri (BA 44, 45, 48). Maths reasoning showed additional activation in the midcingulate area (BA 24) and the bilateral postcentral gyrus (BA 3). Increased BOLD signal in the parahippocampal gyrus, paracentral lobule and cerebellum during science reasoning was associated with better performance on the non-verbal matrix reasoning task. Increased BOLD signal in the left temporal lobe during maths reasoning was associated with better performance on the verbal analogical reasoning task. There were no associations between misconception performance specifically and relational reasoning when relevant executive function and vocabulary measures were controlled for. Overall, relational reasoning showed association with general science and maths reasoning.

## 6.2 Introduction

Relational reasoning is the ability to detect patterns between different representations (Crone et al., 2009). Analogical reasoning, the ability to observe similarities between items based on common associations (Whitaker et al., 2017), is the most studied type of relational reasoning, and is thought to be particularly important in learning new information. Other posited types of relational reasoning include anomalous, antinomous, and antithetical reasoning (Alexander et al., 2016), each with different developmental trajectories (Jablansky et al., 2015). Relational reasoning ability continues to develop through adolescence (Richland et al., 2006; Whitaker et al., 2017), and there is evidence that as age increases, so does relational reasoning-related activation in the RLPFC (BA 10/47) (Dumontheil et al., 2010), VLPFC (BA 45) and SPL (BA 7) (Wright et al., 2008).

Behavioural studies have linked relational reasoning to executive functions. One study found that early semantic inhibition and general executive function was associated with later performance on a verbal analogy task (Richland & Burchinal, 2013). Similarly, VS-WM was related to all four types of non-verbal relational reasoning measured by the TORR (Grossnickle et al., 2016). Behavioural studies have also shown significant associations between relational reasoning and science and maths. The research in science is limited however, with studies typically using analogy as a successful method of teaching new scientific information (Jee et al., 2013; Matlen et al., 2009), as opposed to investigating individual differences. The role of analogical reasoning in moment-to-moment processing is therefore less clear. The literature is somewhat more developed in maths, with individual differences in maths shown to associate with individual differences in relational reasoning in 4- to 5-year-olds (White et al., 1998), 7- to 8-year-olds (Farrington-Flint et al., 2007), and adults (Alexander et al., 2016).

The current study therefore aimed to expand the research on relational reasoning in three ways. First, the study aimed to look at the association between relational reasoning and science and maths while controlling for possible shared associations with executive function measures. It is currently unclear the extent to which associations between relational reasoning and science and maths may be driven by shared reliance on executive functions. Second, the study aimed to additionally study associations between relational reasoning and science and maths counterintuitive reasoning, again controlling for possible shared associations with executive functions. Third, the study aimed to determine whether individual differences in relational reasoning relate to individual differences in brain activation during science and maths reasoning.

The science and maths data described in **Chapter 5** were re-analysed to assess possible unique associations with two types of relational reasoning. Verbal analogical reasoning and non-verbal matrix reasoning tasks were administered. It was first hypothesised that both types of relational reasoning would improve with age, as shown in previous behavioural studies (Jablansky et al., 2015; Richland et al., 2006; Whitaker et al., 2017). It was further hypothesised that better relational reasoning ability would be associated with better science and maths performance, in line with behavioural research (e.g. White, Alexander, & Daugherty, 1998). In terms of brain data, it was hypothesised that better relational reasoning would be associated with greater recruitment of brain regions associated with science and maths. In particular, greater activation was anticipated in regions involved in relational reasoning: the RLPFC (BA 10/46) (Dumontheil et al., 2010), the VLPFC (BA 45), and the SPL (BA 7) (Wright et al., 2008).

It was hypothesised that verbal analogical reasoning would be more important in science, since the language requirements are greater in science trials than maths trials, some of which present equations. It is thought that verbal encoding of associations is a key skill in the learning of science (Tolmie et al., 2016), indicating possible shared verbal requirements of both science and verbal analogical reasoning. It was also hypothesised that non-verbal matrix reasoning would be more important in maths, which requires less language and more visuospatial reasoning. This would fit with behavioural research that has shown a greater link between maths and non-verbal reasoning than verbal analogical reasoning (van der Sluis, de Jong, & van der Leij, 2007). Finally, it was anticipated that associations between relational reasoning and science and maths (in terms of behaviour and brain activation) would be most important in the context of overall performance, rather than specifically for misconceptions.

## 6.3 Methods

### 6.3.1 Participants

The sample consisted of the participants from fMRI analysis 1 in **Chapter 5**, i.e. 34 adolescents aged 11 to 15 years old (17 girls, 17 boys) (see section 5.3.1 for details). None of the participants were outliers on the relational reasoning measures using a threshold of ±3.29 *SD*s from the group mean, therefore there was no further participant exclusion.

### 6.3.2 Tasks

#### 6.3.2.1 Science and maths misconceptions

The behavioural and brain data from the science and maths misconceptions task in **Chapter 5** (see section 5.3.2.1 for details) was used for this analysis. Here, rather than focusing only on contrasting activation between misconception and control trials, analyses also considered activation collapsing across these two trial types, but separately for science and maths.

#### 6.3.2.2 Inhibitory control

The behavioural inhibitory control data from both the Go/No-Go and numerical Stroop in **Chapter 5** (see section 5.3.2.2 for details) were used for this analysis. Simple and complex Go RT cost were the response inhibition measures used. Accuracy and RT costs in the numerical Stroop were the semantic inhibition measures used.

#### 6.3.2.3 Analogical reasoning

A verbal analogical reasoning task (**Table 6.1**) adapted from Leech, Mareschal, and Cooper (2007) was administered on a laptop using an online Google Form in a quiet room outside the scanner. Four practice items were administered with discussion between the participant and experimenter, to ensure that the task was understood. There were 24 questions given in the format A:B::C:?, with a choice of four options. No time limit was given; participants were left to complete the task in their own time and could scroll up and down through different questions as all were present on the screen at once. The total number of correct responses was recorded (range = 15-24, *M* = 20, *SD* = 3).

#### 6.3.2.4 WASI

The Matrix Reasoning subtest of the WASI-II (Wechsler, 2011) was carried out using the manual, as described in **Chapter 3** (see section 3.3.2.1 for details). This provided a measure of non-verbal relational reasoning ability. The raw score was used in the analyses, as opposed to the standardised score, to enable meaningful examination of changes with age (range = 15-26, *M* = 22, *SD* = 3).

The Vocabulary subtest of the WASI-II (Wechsler, 2011) was carried out using the manual, as described in **Chapter 3** (see section 3.3.2.1 for details), providing a control measure of verbal ability. The raw score was used in the analyses (range = 31-43, *M* = 38, *SD* = 3).

**Table 6.1.** Questions in the verbal analogical reasoning task. Response options appeared in fixed order. P = practice item. Correct responses are in bold.

| No. | Question | Response 1 | Response 2 | Response 3 | Response 4 |
|---|---|---|---|---|---|
| P 1 | Nose is to Smelling as Eye is to | Stink | Glasses | **Seeing** | Listening |
| P 2 | Plane is to Air as Submarine is to | **Water** | Torpedo | Flight | Navy |
| P 3 | Right is to Wrong as Good is to | Mistaken | Morality | Decent | **Bad** |
| P 4 | Man is to Child as Oak is to | Tree | **Acorn** | Teenager | Leaf |
| 1 | Bear is to Cave as Dog is to | **Dog House** | Bulldog | Leash | Cat |
| 2 | Ocean is to Fountain as Desert is to | **Sandbox** | Water | Oasis | Dry |
| 3 | Barber is to Scissor as Fireman is to | **Hose** | Policeman | Fire Station | Postman |
| 4 | Asleep is to Wide Awake as Calm is to | Tranquiliser | Napping | Cycling | **Frantic** |
| 5 | Bright is to Dark as Midday is to | Sun | **Midnight** | Lunchtime | Shiny |
| 6 | Bird is to Wings as Fish is to | Shrimp | **Fins** | Shark | Water |
| 7 | Unjust is to Fair as Obsessed is to | Fanatic | Biased | Passionate | **Indifferent** |
| 8 | Necklace is to Neck as Watch is to | Watch Case | Bracelet | Ring | **Wrist** |
| 9 | Noisy is to Deafening as Quiet is to | Ear-Splitting | Peaceful | **Silent** | Headphones |
| 10 | Tractor is to Aeroplane as Walk is to | Trainers | Bicycle | **Sprint** | Sit |
| 11 | Turtle is to Shell as Bird is to | Feathers | **Nest** | Parrot | Pigeon |
| 12 | Guzzle is to Sip as Cascade is to | Drink | Rocks | **Stream** | Chemical Reaction |
| 13 | Boiling is to Warm as Chilly is to | **Freezing** | Pepper | Winter | Scorching |
| 14 | Sock is to Foot as Hat is to | Woolly Hat | **Head** | Cap | Hat Rack |
| 15 | Snack is to Feast as Peckish is to | Food | Thirsty | **Starving** | Bite |
| 16 | Valley is to Mountain as Mine is to | Coal | Canyon | **Tower** | Bury |

| | | | | | |
|---|---|---|---|---|---|
| 17 | Fish is to Water as Squirrel is to | Mouse | Nuts | **Tree** | Hamster |
| 18 | House is to City as Star is to | **Galaxy** | Sun | Dwelling | Celebrity |
| 19 | Doctor is to Hospital as Policeman is to | Detective | Police Car | **Police Station** | Security Man |
| 20 | Damp is to Drenched as Drizzle is to | **Downpour** | Storm | Moist | Umbrella |
| 21 | Cat is to Milk as Dog is to | Wolf | **Bone** | Fox | Ball |
| 22 | Genius is to Stupid as Crazy is to | **Sane** | Straightjacket | Mastermind | Bad |
| 23 | Radio is to Ear as TV is to | CD Player | Microwave | **Eye** | Table |
| 24 | Sparkly is to Dull as Happy is to | Smile | **Miserable** | Enthusiastic | Bored |

#### 6.3.2.5 Working memory

VS-WM was assessed using the dot matrix as in the classroom studies in **Chapter 4** (see section 4.3.1.3.2 for details), but this version was programmed in MATLAB (The MathWorks, Inc., Natick, MA) and administered one-to-one, rather than administered online in a group testing session. Verbal working memory (VWM) was assessed using a verbal backwards digit span task. The experimenter read out a series of numbers and the participant verbally repeated the numbers in reverse order straight away. A practice of two items was given, and repeated with additional explanation if the participant did not get the answer right. Both tasks started with a load of three items, and the load increased in sets of four trials until two incorrect answers were given at any one load. The number of correct trials was recorded for each task (VS-WM score range = 2-16, $M$ = 9, $SD$ = 4; VWM score range = 2-18, $M$ = 10, $SD$ = 4).

### 6.3.3 Procedure

Participants were tested on these behavioural measures in a quiet room either before or after taking part in the fMRI procedure described in **Chapter 5** (see section 5.3.3 for details), in a session lasting approximately 30 min. The experimenter explained each task to ensure that participants understood what was required.

### 6.3.4 Statistical analysis

Analyses first focussed on the behavioural and brain data collapsing across misconception and control trials in science and maths. In a second stage, analyses contrasted misconception and control trials, as in **Chapter 5**.

#### 6.3.4.1 Overall science and maths reasoning

##### 6.3.4.1.1 Behavioural analysis

Correlations were run between the key variables to examine associations and test for multicollinearity. Hierarchical multiple regressions investigated the extent to which relational reasoning variables could account for individual differences in overall science and maths accuracy and RT by discipline. Block 1 variables included the control variables, which were inserted stepwise so that only the significant predictors were kept in the model: age in months, WASI Vocabulary raw score, VS-WM total correct, VWM total correct, simple Go RT cost, complex Go RT cost, Stroop accuracy cost, Stroop RT cost. Block 2 contained the relational reasoning measures entered stepwise: analogical reasoning score and WASI Matrix Reasoning raw score. Relational reasoning measures were entered last to establish the unique variance they could explain, over and above age, vocabulary, and executive functions.

#### 6.3.4.1.2 FMRI analysis

FMRI data were preprocessed and modelled as described in **Chapter 5** (see section 5.3.5 for details). First-level contrasts of the difference between science trials (misconception and control trials combined) and the arrows task (Science > Arrows) and between maths trials (misconception and control trials combined) and the arrows task (Maths > Arrows) were calculated. Contrasts were entered into one sample *t*-tests to create SPM maps which were thresholded at $p < .001$ uncorrected at the voxel level and at FWE-corrected $p < .05$ at the cluster level. Peak voxels significant at FWE-corrected $p < .05$ at the voxel level are also indicated. Associations between BOLD signal and age were investigated through entering age in months as a covariate to the one sample *t*-tests.

Associations between BOLD signal and relational reasoning were investigated through entering analogical reasoning scores and WASI Matrix Reasoning raw scores as covariates. Where there were significant regions of activation, these were followed up by averaging the data for the Science > Arrows or Maths > Arrows contrast over each identified cluster using MarsBaR (Brett et al., 2002). Regressions were first run on mean activation in each cluster, entering science or maths accuracy and RT in block 1 using the enter method, followed by the relevant relational reasoning measure stepwise in block 2. Finally, further regressions were run on mean activation in each cluster, entering the same variables as in the behavioural analyses stepwise, to examine whether relational reasoning associations remained when controlling for the effect of any other significant factors (age, vocabulary, and executive function).

### 6.3.4.2 Counterintuitive science and maths reasoning

To examine misconception-specific associations, further correlations were run on science and maths accuracy and RT by trial type (misconception and control). Science and maths were combined for these analyses because there were no discipline effects for the Misconception > Control contrast. Further hierarchical multiple regressions investigated the extent to which the relational reasoning variables could account for individual differences in science and maths misconception accuracy and RT specifically. Block 1 variables included the control variables, which were inserted stepwise: age in months, WASI Vocabulary raw score, VS-WM total correct, VWM total correct, and control performance (accuracy or RT). Block 2 contained the relational reasoning measures entered stepwise: analogical reasoning score and WASI Matrix Reasoning raw score.

In order to investigate associations specific to misconceptions, relational reasoning measures were entered as covariates into the Misconception > Control contrast described in **Chapter 5** (see section 5.3.6 for details).

## 6.4 Results

### 6.4.1 Overall science and maths reasoning

#### 6.4.1.1 Behavioural results

Correlations between the variables of interest (**Table 6.2**) showed neither relational reasoning measure to correlate with age. The two relational reasoning measures were positively correlated with one another. In terms of associations with science and maths, WASI Matrix Reasoning was significantly associated only with science RT, such that those with better Matrix Reasoning scores were faster to answer science problems. Analogical reasoning, on the other hand, was associated with all science and maths measures; those with better analogical reasoning were more accurate and faster on both science and maths problems.

In terms of associations with executive function, again different patterns were observed for the two relational reasoning measures. WASI Matrix Reasoning was positively associated with both measures of working memory, but none of the inhibitory control measures. In contrast analogical reasoning was positively associated with VS-WM but not VWM, and positively associated with Simple Go RT cost. Those who had better analogical reasoning had a larger RT cost, so were more affected by No-Go trials, i.e. they slowed down more when No-Go trials were present.

There were also links between science and maths performance and executive function. Simple Go RT cost was positively associated with science accuracy, showing that those who performed better on the science task had a larger RT cost. VS-WM was significantly related to all science and maths measures, such that those with better VS-WM scores were faster and more accurate in science and maths. Finally, VWM was positively associated with science accuracy.

**Table 6.2.** Pearson correlation coefficients of regression variables for science and maths by discipline. Statistically significant (two-tailed) correlations are highlighted in bold, [a,b,c] indicate $p < .05$, $p < .01$ and $p < .001$ respectively.

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Science accuracy | | | | | | | | | | | | | |
| 2. Science RT | -.19 | | | | | | | | | | | | |
| 3. Maths accuracy | **.61**[c] | -.26 | | | | | | | | | | | |
| 4. Maths RT | -.16 | **.73**[c] | -.28 | | | | | | | | | | |
| 5. Age (months) | **.40**[a] | .02 | **.42**[a] | -.16 | | | | | | | | | |
| 6. WASI Vocabulary | **.63**[c] | -.32 | **.51**[b] | -.23 | **.45**[b] | | | | | | | | |
| 7. Simple Go RT cost | **.41**[a] | -.33 | .03 | -.11 | .00 | .33 | | | | | | | |
| 8. Complex Go RT cost | .06 | -.28 | -.01 | -.05 | .05 | .23 | **.59**[c] | | | | | | |
| 9. Stroop accuracy cost | .04 | .10 | .15 | .01 | .21 | .08 | .00 | -.07 | | | | | |
| 10. Stroop RT cost | -.14 | -.06 | .06 | .15 | -.09 | -.22 | .00 | .14 | -.14 | | | | |
| 11. WASI Matrix Reasoning | .33 | **-.37**[a] | .28 | -.24 | -.06 | .20 | .28 | .09 | .06 | .24 | | | |
| 12. Analogical reasoning | **.67**[c] | **-.48**[b] | **.48**[b] | **-.42**[a] | .04 | **.48**[b] | **.38**[a] | .11 | .11 | -.09 | **.54**[b] | | |
| 13. VS-WM | **.36**[a] | **-.36**[a] | **.45**[b] | **-.35**[a] | .15 | .14 | .09 | .06 | .06 | .21 | **.44**[a] | **.45**[b] | |
| 14. VWM | **.40**[a] | -.32 | .28 | -.26 | -.06 | .14 | .26 | .00 | -.09 | .21 | **.39**[a] | .24 | **.43**[a] |

The first hierarchical multiple regression investigated whether relational reasoning variables could account for individual differences in science accuracy when relevant age, vocabulary, and executive function differences were taken into account (**Table 6.3**). First, executive function measures, vocabulary, and age were entered stepwise, so that any significant variables could be taken account of before adding relational reasoning. In model 1a, WASI Vocabulary was a significant predictor, and 40% of the variance in science accuracy was explained. In model 1b, VWM total was added, explaining an extra 11% of the variance. Finally, analogical reasoning was selected by model 1c, explaining an additional 12% of the variance in science accuracy.

**Table 6.3.** Regression models for overall science and maths accuracy by discipline. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science accuracy | | | | |
| Model 1a<br>$F(1, 31) = 20.56$,<br>$p < .001$, $R^2 = 40\%$ | Constant | | 0.44 | .667 |
| | **WASI Vocabulary raw** | **.63** | **4.54** | **< .001** |
| Model 1b<br>$F(2, 30) = 15.27$,<br>$p < .001$, $R^2 = 50\%$,<br>$\Delta R^2 = 10.6\%$ | Constant | | 0.30 | .765 |
| | **WASI Vocabulary raw** | **.59** | **4.51** | **< .001** |
| | **VWM total** | **.33** | **2.53** | **.017** |
| Model 1c<br>$F(3, 29) = 16.31$,<br>$p < .001$, $R^2 = 63\%$,<br>$\Delta R^2 = 12.3\%$ | Constant | | .01 | .995 |
| | **WASI Vocabulary raw** | **.39** | **3.02** | **.005** |
| | **VWM total** | **.24** | **2.07** | **.047** |
| | **Analogical reasoning score** | **.42** | **3.10** | **.004** |
| Maths accuracy | | | | |
| Model 2a<br>$F(1, 31) = 10.85$,<br>$p = .002$, $R^2 = 26\%$ | Constant | | 1.07 | .295 |
| | **WASI Vocabulary raw** | **.51** | **3.29** | **.002** |
| Model 2b<br>$F(2, 30) = 10.14$,<br>$p < .001$, $R^2 = 40\%$,<br>$\Delta R^2 = 14.4\%$ | Constant | | 0.98 | .336 |
| | **WASI Vocabulary raw** | **.46** | **3.20** | **.003** |
| | **VWM total** | **.38** | **2.69** | **.012** |

The second regression looked at maths accuracy (**Table 6.3**) and similarly found that WASI Vocabulary was a significant predictor in model 2a, explaining 26% of the variance, and VWM total was a significant predictor in model 2b, explaining 14% more of the variance in maths accuracy. However, no relational reasoning measures were selected in the next step.

The next hierarchical multiple regressions were performed on RT (**Table 6.4**). VS-WM total was a significant predictor in model 3a for science RT, explaining 13% of the variance. Model 3b selected analogical reasoning score, explaining an additional 13% of the variance, and within this model VS-WM, was no longer significant. In maths, model 4 also selected VS-WM as a significant predictor, explaining 12% of the variance. No further variables were selected for maths RT.

**Table 6.4.** Regression models for overall science and maths RTs by discipline. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science RT | | | | |
| Model 3a | **Constant** | | **17.25** | **< .001** |
| $F(1, 31) = 4.56$, $p = .041$, $R^2 = 13\%$ | **VS-WM total** | **-.36** | **-2.14** | **.041** |
| Model 3b | **Constant** | | **8.97** | **< .001** |
| $F(2, 30) = 5.29$, | VS-WM total | -.18 | -1.01 | .323 |
| $p = .011$, $R^2 = 26\%$, $\Delta R^2 = 13.2\%$ | **Analogical reasoning score** | **-.41** | **-2.32** | **.028** |
| Maths RT | | | | |
| Model 4 | **Constant** | | **18.33** | **< .001** |
| $F(1, 31) = 4.22$, $p = .048$, $R^2 = 12\%$ | **VS-WM total** | **-.35** | **-2.06** | **.048** |

Overall, the behavioural results showed that analogical reasoning was associated with science and maths accuracy and RT. Associations with science, but not maths, remained significant when VS-WM (for RT) or vocabulary and VS-WM (for accuracy) were taken into account. Better analogical reasoning was therefore uniquely associated with higher accuracy and faster RTs in science only. WASI Matrix Reasoning was associated with faster science RTs, but this association did not hold when other measures were controlled for.

#### 6.4.1.2 FMRI results

The contrast of [(Maths > Arrows) > (Science > Arrows)] was significant (although not the inverse) so science and maths were kept separate. The Science > Arrows contrast showed increased BOLD signal in a broad bilateral network of regions, with greater activation in the left hemisphere (**Figure 6.1**). A large bilateral cluster was observed in the occipital lobe (BA 17, BA 18, BA 19) extending to the inferior temporal gyrus (BA 37) and hippocampus (BA 27), and dorsally to the IPL (BA 7, 40, 2). Within the PFC, bilateral activation was seen in the inferior frontal gyri (BA 44, BA 45, BA 48). Further bilateral activation was observed in the superior temporal pole (BA 38) and insula (BA 47). Age was not a significant covariate for this contrast.

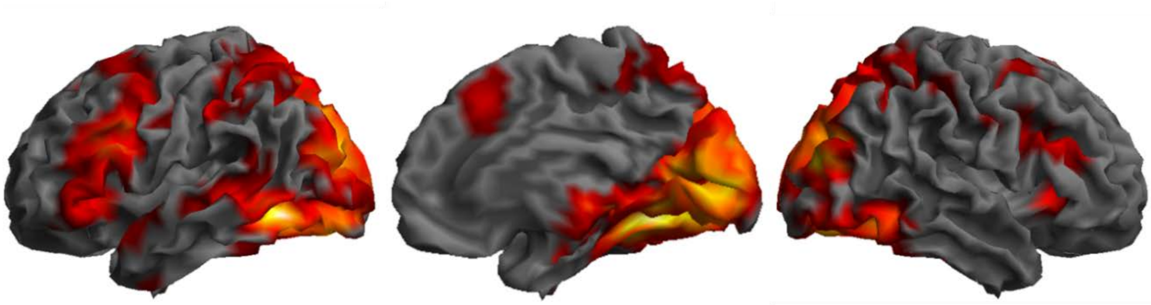


**Figure 6.1.** Regions of increased BOLD signal in the Science > Arrows contrast from the one sample *t*-test with no covariates added. Contrasts $p_{uncorr} < .001$ at the voxel level and $p_{FWE} < .05$ at the cluster level. Images are rendered on the canonical brain in SPM, showing from left to right: the lateral view of the left hemisphere, and medial and lateral views of the right hemisphere.

The Maths > Arrows contrast showed increased BOLD signal in similar regions observed in the Science > Arrows contrast (**Figure 6.2**). A cluster was observed in the SMA (BA 8), extending to the midcingulate area (BA 24). There was also activation in the thalamus, and the right IFG (BA 45). Finally, this contrast showed increased BOLD signal in the bilateral postcentral gyrus (BA 3). Age was not a significant covariate for the contrast.

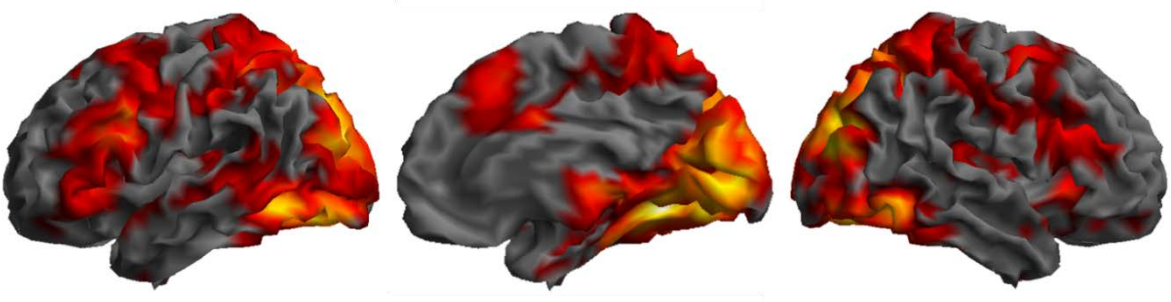


**Figure 6.2.** Regions of increased BOLD signal in the Maths > Arrows contrast from the one sample *t*-test with no covariates added. Contrasts $p_{uncorr} < .001$ at the voxel level and $p_{FWE} < .05$ at the cluster level. Images are rendered on the canonical brain in SPM, showing from left to right: the lateral view of the left hemisphere, and medial and lateral views of the right hemisphere.

Further analyses assessed whether brain activation in either of these contrasts was associated with relational reasoning behavioural measures. In the Science > Arrows contrast, analogical reasoning was not a significant covariate but there were three clusters significantly positively correlated with WASI Matrix Reasoning. The first had a peak in the parahippocampal gyrus (BA 37), the second was more dorsal with a peak in the paracentral lobule (BA 4), and the third was in the cerebellum (BA 18) (**Table 6.5**, **Figure 6.3**).

**Table 6.5.** Regions where BOLD signal during science or maths reasoning versus the arrows task positively correlates with relational reasoning, $^{a}p_{FWE} < .05$ at the voxel level, $^{b}p_{FWE} < .05$ at the cluster level (cluster defining threshold: $p_{uncorr} < .001$).

| **Brain region** | **L/R** | **BA** | **MNI** | | | ***Z*-score** | **Cluster size** |
|---|---|---|---|---|---|---|---|
| | | | ***x*** | ***y*** | ***z*** | | |
| *Science > Arrows and WASI Matrix Reasoning* | | | | | | | |
| Parahippocampal gyrus | R | 37 | 33 | -34 | -10 | 4.72[a] | 1396[b] |
| Cuneus | R | 18 | 9 | -76 | 17 | 4.67[a] | |
| Lobule III of vermis | R | 27 | 6 | -46 | -22 | 4.69 | |
| Paracentral lobule | L | 4 | -3 | -34 | 56 | 4.42 | 228[b] |
| Precuneus | L | 5 | -9 | -46 | 68 | 4.15 | |
| Midcingulate area | L | 23 | -3 | -43 | 44 | 3.35 | |
| Crus I of cerebellar | R | | 12 | -76 | -28 | 3.99 | 91[b] |
| Lobule VI of vermis | R | 18 | 6 | -82 | -19 | 3.72 | |
| *Maths > Arrows and analogical reasoning* | | | | | | | |
| Middle temporal gyrus | L | 21 | -57 | 2 | -16 | 4.64 | 264[b] |
| Middle temporal pole | L | 20 | -39 | 8 | -31 | 4.06 | |
| Superior temporal pole | L | 38 | -42 | 14 | -22 | 3.99 | |

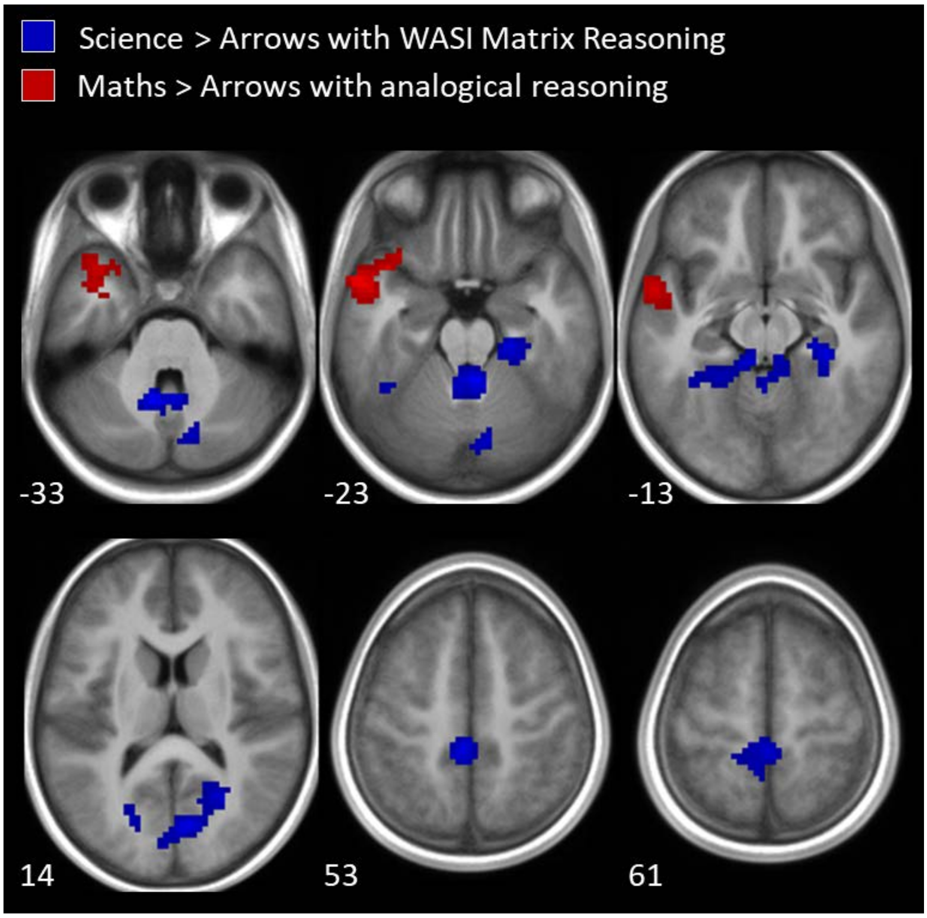



**Figure 6.3.** Brain regions where BOLD signal during science or maths reasoning positively correlated with behavioural relational reasoning performance, *z*-coordinates are indicated in the bottom left corner. Clusters are plotted on six horizontal slices of the average normalised structural image of the 34 participants.

In the Maths > Arrows contrast a 264 voxel cluster in the left middle temporal gyrus (BA 21) extending to the middle (BA 20) and superior temporal pole (BA 38) was a significant positive covariate with analogical reasoning (**Table 6.5**, **Figure 6.3**). WASI Matrix Reasoning did not correlate with the maths contrast.

Average parameter estimates within each cluster were calculated and plotted against accuracy in the relevant relational reasoning task (**Figure 6.4**), indicating that these associations were not due to outliers but general trends across participants. Follow-up analyses were performed in SPSS. The association between WASI Matrix Reasoning and Science > Arrows BOLD signal in all three clusters remained significant when controlling for science accuracy and RT, $p$'s < .001.

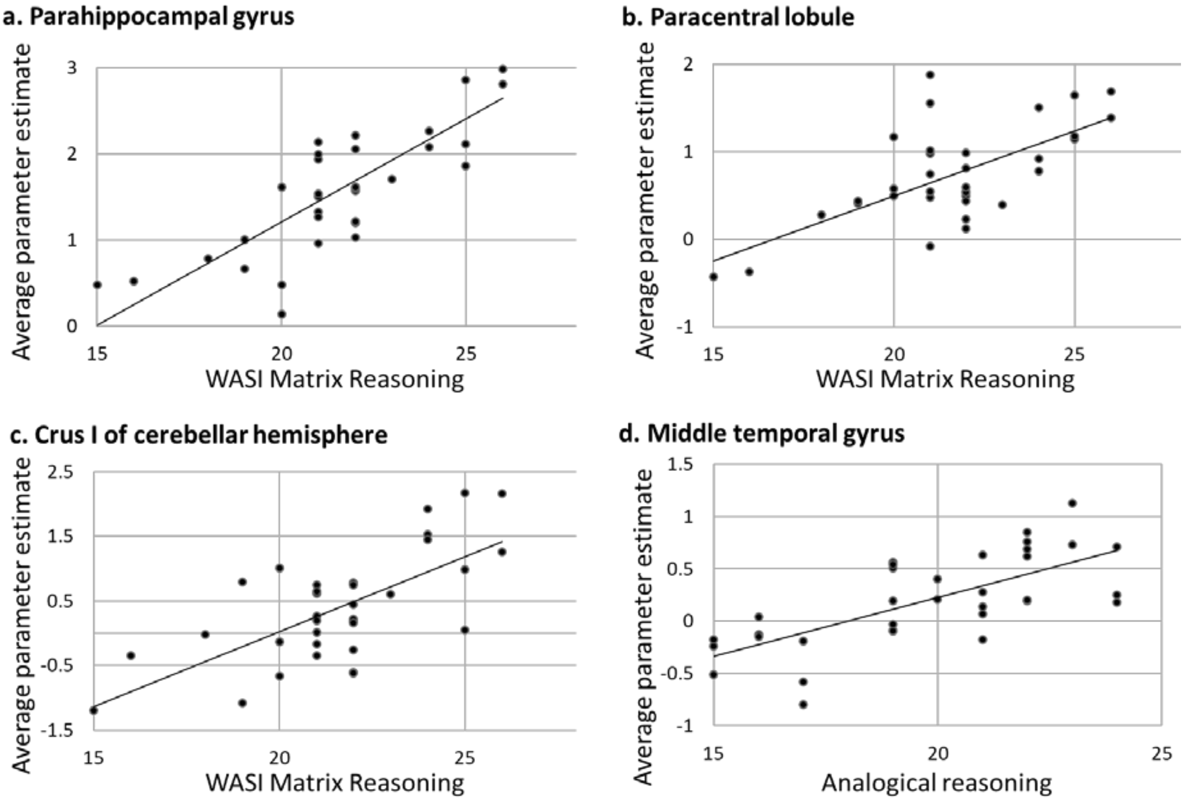



**Figure 6.4.** Average parameter estimates plotted against relational reasoning accuracy for the four significant clusters. Science > Arrows covaried with WASI Matrix Reasoning in clusters that have their peak in (a) the parahippocampal gyrus, (b) the paracentral lobule, and (c) crus I of cerebellar hemisphere; (d) Maths > Arrows covaried with analogical reasoning in the cluster that has its peak in the middle temporal gyrus.

Regressions were run on each of the clusters, entering the control variables (age, WASI Vocabulary, and executive functions) stepwise in block 1, followed by WASI Matrix Reasoning stepwise in block 2. Full regression tables are reported in **Appendix 2**, and for all three clusters, WASI Matrix Reasoning score remained a significant predictor of Science > Arrows BOLD signal, $p$'s < .001. In addition, the model for activation in the parahippocampal gyrus cluster selected VWM and Simple Go RT cost, although neither were significant after WASI Matrix Reasoning was included in the model. The model for activation in the paracentral lobule selected Complex Go RT cost and VWM, although VWM was no longer significant when WASI Matrix Reasoning was added. Finally, the model for crus I of the cerebellum selected VS-WM, and this did not remain significant when WASI Matrix Reasoning was added. In all cases, the beta values were positive, such that better working memory and higher cost, i.e. more slowing down in the presence of No-Go trials, were associated with larger parameter estimates (i.e. greater activation).

The association between analogical reasoning and Maths > Arrows BOLD signal in the temporal lobe cluster remained significant when controlling for maths accuracy and RT, $p < .001$. In addition, the regression (see full table in **Appendix 2**) showed that analogical reasoning remained a significant predictor of Maths > Arrows BOLD signal when the control variables were entered stepwise in a first block, $p < .001$. The model selected Simple Go RT cost as a significant positive predictor, but this was no longer significant when analogical reasoning was selected by the model.

The neuroimaging results revealed that a broad network of bilateral brain regions showed increased BOLD signal during science and maths reasoning, with maths activation spreading out further than science. Increased BOLD signal across the parahippocampal gyrus, paracentral lobule and cerebellum during science reasoning was associated with higher WASI Matrix Reasoning raw scores. Increased BOLD signal in the left temporal lobe during maths reasoning was associated with better performance on the analogical reasoning task. These associations held when controlling for the effects of working memory and response inhibition.

## 6.4.2 Counterintuitive science and maths reasoning

### 6.4.2.1 Behavioural results

For these analyses, data from the science and maths disciplines were combined as no difference in brain activation between misconception and control trials were observed between the disciplines in **Chapter 5**. Correlations split by misconception and control trials (**Table 6.6**) showed that WASI Matrix Reasoning was positively associated with control accuracy, while analogical reasoning was associated with higher accuracy and faster RTs across both trial types. Better VS-WM was associated with higher accuracy and faster RTs on control trials. VWM total was positively associated with control accuracy.

**Table 6.6.** Pearson correlation coefficients of regression variables for science and maths misconceptions. Statistically significant (two-tailed) correlations are highlighted in bold, [a,b,c] indicate $p$ <.05, $p$ < .01 and $p$ < .001 respectively.

| Variable | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Misconception accuracy | | | | |
| 2. Misconception RT | -.08 | | | |
| 3. Control accuracy | **.57[c]** | -.34 | | |
| 4. Control RT | -.20 | **.88[c]** | **-.44[b]** | |
| 5. Age (months) | **.47[b]** | -.06 | .29 | -.10 |
| 6. WASI Vocabulary | **.59[c]** | -.29 | **.53[b]** | -.29 |
| 7. Simple Go RT cost | .25 | -.21 | .17 | -.25 |
| 8. Complex Go RT cost | .00 | -.15 | .09 | -.19 |
| 9. Stroop accuracy cost | .14 | .07 | .03 | .04 |
| 10. Stroop RT cost | -.13 | .09 | .10 | .00 |
| 11. WASI Matrix Reasoning | .26 | -.31 | **.38[a]** | -.33 |
| 12. Analogical reasoning | **.55[b]** | **-.44[b]** | **.63[c]** | **-.49[b]** |
| 13. VS-WM | .30 | -.32 | **.54[b]** | **-.42[a]** |
| 14. VWM | .28 | -.28 | **.42[a]** | -.33 |

Hierarchical multiple regressions examined age, vocabulary and executive function as predictors of science and maths misconception performance and tested whether relational reasoning measures accounted for unique variance in performance. For misconception accuracy, Model 5a was significant, selecting WASI Vocabulary as a significant positive predictor, explaining 35% of the variance, $F(1, 31) = 16.63$, $p < .001$. Model 5b was significant, adding control accuracy as a positive predictor, which explained an additional 10% of the variance, $F(2, 30) = 11.97$, $p < .001$. No further variables were selected. The regression on misconception RT showed that only control RT was a significant positive predictor, explaining 77% of the variance $F(1, 31) = 103.55$, $p < .001$.

Neither relational reasoning measure was found to associate with brain activation in the Misconception > Control contrast.

Overall, the results showed that analogical reasoning was associated with all trial types, while WASI Matrix Reasoning was associated only with control accuracy. When accounting for control performance, no relational reasoning measures were predictors of misconception performance. There were no associations between science and maths misconception brain activation and relational reasoning performance.

## 6.5 Discussion

This study aimed to investigate the role of relational reasoning in science and maths. In particular, both verbal analogical reasoning and non-verbal matrix reasoning were examined in the context of general science and maths, and in the context of misconceptions. Contrary to the hypothesis, there was no association between relational reasoning measures and age in this sample, where previous research has found relational reasoning to develop in this age range (Richland et al., 2006; Whitaker et al., 2017). This may be due to the insufficient power of this study for behavioural analyses or the relatively narrow age range of the participants, who were selected to be in or having just finished Key Stage 3. The study by Richland and colleagues found age improvements in a sample of 68 3- to 14-year-olds, while the study by Whitaker and colleagues observed age improvements in a sample of 95 6- to 18-year-olds. It is also possible that the tasks used here were not sensitive to age changes.

In terms of general science and maths, when controlling for other factors, better analogical reasoning was associated with higher accuracy and faster responses in science only. While the previous evidence linking analogical reasoning with science has typically focussed on learning (Jee et al., 2013; Matlen et al., 2009), and analogy is specifically thought to be key in learning (Whitaker et al., 2017), this new evidence shows that analogical reasoning is also associated with in-the-moment reasoning. Correlations indicated an association between maths and analogical reasoning but when other factors were taken into account this effect was no longer present. Previous evidence linking analogical reasoning to maths (Alexander et al., 2016; White et al., 1998), may be in part attributable to visuospatial working memory (VS-WM) or vocabulary, which were not controlled for in those studies.

Even though non-verbal matrix reasoning was not associated with science behaviourally when controlling for other factors, it was positively associated with increased BOLD signal during science reasoning. Increased activation in the parahippocampal gyrus was associated with better performance in matrix reasoning. Luo and colleagues (2003) found activation in this region for a similar analogical reasoning task, and suggested that the parahippocampal gyrus had a role in maintaining individual mental representations during analogical reasoning. Increased activation in the paracentral lobule, encompassing primary motor and sensory areas, during science reasoning was also associated with better non-verbal matrix reasoning performance, with less clear links to shared functions. Since there were no behavioural associations, it is possible that those who were better at matrix reasoning engaged those brain networks more during science reasoning, but they did not necessarily hold the knowledge necessary to get the answers correct. This would explain the lack of association between neural activity and accuracy.

Working memory and response inhibition were associated with brain activation in science regions that covaried with matrix reasoning, although when matrix reasoning was selected by the model, these effects typically reduced or disappeared. This suggests that the association between relational reasoning and science is not driven solely by executive functions.

During maths reasoning, increased BOLD signal in the temporal lobe was associated with better analogical reasoning. The middle temporal gyrus is thought to be crucial for the semantic processing of multimodal information (Visser, Jefferies, Embleton, & Ralph, 2012). The requirement for processing of multimodal stimuli is clear in the maths task, which presented both words and pictures, and sometimes equations. While the analogical reasoning task only presented words, it is possible that the words presented conjured mental pictures, or that the spatial processing inherent in the task (i.e. A is to B as C is to …) required integration of the two modes of processing. Response inhibition was associated with activation in the temporal lobe, with the effect disappearing when analogical reasoning was selected by the model. As with science, this indicates that the association between analogical reasoning and maths is not driven by executive functions. The activation of this region may reflect the involvement of cognitive processes that are involved in both inhibitory control and analogical reasoning, but more strongly involved in analogical reasoning.

It was surprising that maths did not relate to non-verbal matrix reasoning, since previous evidence has shown matrix-type reasoning measures to relate with maths in particular (Farrington-Flint et al., 2007; Green et al., 2017). The greater link between maths and verbal reasoning compared to non-verbal reasoning is in contrast to other research (van der Sluis et al., 2007) and may highlight the relatively high language requirements of the current maths task. It is worth noting that there are examples of other studies that have shown behavioural and neuroimaging data not to map directly onto each other. One such study attributed these differences to the varying sensitivity of the different methods, since there are likely factors that influence behavioural data which might not be reflected in imaging data (Dumontheil, Wolf, & Blakemore, 2016).

The two relational reasoning measures were correlated with each other, indicating that high performance on one task is related to high performance on the other. However, the patterns of behavioural association were distinct, indicating that relational reasoning is an umbrella term to describe different types of ability. Aspects of verbal ability were controlled for in terms of expressive vocabulary and in the backwards digit span (measuring VWM), indicating that it was likely not just the verbal component of the analogical reasoning task that was driving the association. Rather, it seems that the ability to reason analogically may have specific importance in science reasoning.

This study was a first attempt to investigate different types of relational reasoning in science and maths within behavioural and neuroimaging data. The small sample size was a limitation of the study, particularly in looking at the behavioural data. A sufficiently powered behavioural study (see **Chapter 7**) is needed to draw firmer conclusions about the role of relational reasoning in science and maths. Overall the data so far suggest that analogical reasoning predicts unique variance in science performance, with more limited behavioural, but some neural associations in maths. Matrix reasoning showed minimal behavioural associations, but was related to neural activation in science. These effects are observed over and above individual differences in relevant executive functions, age and vocabulary.

Having established possible links between relational reasoning and science and maths, the thesis will next probe this association with a sufficiently powered behavioural study. **Chapter 7**, the final experimental chapter of the thesis, will further consider the extent to which individual differences in analogical and non-verbal relational reasoning associate with science and maths performance, again while examining a possible interplay with executive functions, including inhibitory control.

# Chapter 7 Behavioural study 2: Relational reasoning and inhibitory control in science and maths

## 7.1 Overview

Inhibitory control and relational reasoning are two domain-general skills that continue to mature through adolescence. Both of these abilities are thought to be involved in science and maths reasoning, but their relative importance is unknown. In the current study, 11- to 15-year-olds ($N$ = 120) completed response and semantic inhibition tasks (complex Go/No-Go, numerical Stroop, Hayling sentence completion), verbal analogical and non-verbal relational reasoning tasks (analogical reasoning and Cattell Culture Fair, CCF), and a counterintuitive science and maths reasoning task. Both verbal and non-verbal reasoning was associated with better performance in science and maths in general and specifically for misconception problems. These associations held when vocabulary and executive function were taken into account. Inhibitory control as measured by the Hayling sentence completion test was associated with science and maths performance generally, but not in the context of misconceptions. Overall the results indicate that relational reasoning is associated with general science and maths reasoning, with a particular link to reasoning about misconceptions.

## 7.2 Introduction

This study aimed to investigate further the interplay of inhibitory control, relational reasoning, and science and maths problem-solving. Results from **Chapter 6** indicated that both verbal analogical and non-verbal matrix reasoning were related to science and maths reasoning in terms of both behavioural and brain data, with executive functions and vocabulary possibly playing a mediating role. The results also indicated that relational reasoning was important for overall science and maths performance, and possibly in the context of misconceptions specifically. In **Chapter 5**, behavioural measures of inhibitory control were not significant predictors of science and maths performance in the regression models, and it was considered that a larger behavioural study would enable a better examination of the associations between inhibitory control, relational reasoning, and science and maths while controlling for vocabulary and working memory. Therefore, this study aimed to establish the unique contributions of inhibitory control and relational reasoning to science and maths performance.

In this study, it was important that the tasks fit into one school lesson of 45 min, as one of the schools requested that the testing session last only one lesson per participant to minimise school disruption. This requirement constrained some of the task decisions. The simple Go/No-Go was not included as **Chapters 3 and 5** found that it was only the complex Go/No-Go that related to science and maths performance. The Stroop task was administered, but the number of trials was halved compared to **Chapters 3 and 4**. An additional measure of semantic inhibition was added to the testing session, in an attempt to include a task that more closely reflects the kind of inhibitory control that might be used while reasoning in science and maths. The measure selected was the Hayling sentence completion test (Burgess & Shallice, 1997), which requires participants to listen to a sentence and inhibit the word that fits at the end of the sentence. Since the other inhibitory control tasks have no verbal component and do not rely on prior knowledge, it was considered that an inhibitory control task that recruits language and prior knowledge might show a unique association with science and maths reasoning (see full details in the Methods section below). Previous research has suggested that the Stroop and Hayling sentence completion measure different aspects of inhibition, as in one study performance on the two was not associated when controlling for fluid intelligence, and the neural correlates were dissociable (Cipolotti et al., 2016). Despite both aiming to measure semantic inhibition, the Hayling seems to measure inhibition in the context of implementing a strategy, while the Stroop measures inhibition in the context of suppressing the salient information. An exploration of the role of the two tests in science and maths reasoning is therefore of interest.

In **Chapters 3, 5, and 6**, the WASI Matrix Reasoning task was used as a measure of non-verbal matrix reasoning. In order to capture a greater range of non-verbal relational reasoning skills, the CCF test (Cattell & Cattell, 1959) was used in this study. This test was chosen because it combines four types of relational reasoning, and was thus considered likely to provide a more holistic estimate of relational reasoning. The CCF also has a time limit per scale and it was therefore possible to better estimate the length of the testing session with this measure, in contrast to the WASI Matrix Reasoning task and other relational reasoning tasks that have no time limit or a very generous time allowance. As in **Chapters 4 and 6**, working memory was considered a potentially important contributing factor. To control for the role of working memory, the backwards digit span was administered as in **Chapter 6** to measure VWM. This task was chosen as it was a significant predictor of science and maths accuracy (arguably the more important indicator of performance as compared to RT) in **Chapter 6**.

It was hypothesised that inhibitory control would specifically predict misconception performance, in replication of **Chapter 3**. It was also anticipated that the Hayling sentence completion test might be a better predictor of science and maths misconception performance than the other inhibitory control variables, due to the greater shared task requirements. It was further hypothesised that relational reasoning would predict science and maths performance generally, with verbal analogical reasoning being a more important predictor than non-verbal matrix reasoning, as in **Chapter 6**.

# 7.3 Methods

## 7.3.1 Participants

One hundred and twenty pupils participated in this study (**Table 7.1**). An a priori power analysis conducted using G*Power showed that for a medium effect size with 90% power and 5 tested predictors, 116 would be a sufficient sample size. Letters were sent to parents of 11- to 15-year-olds (in Years 7 to 10), inviting their children to take part. Written informed parental consent confirmed that participating children had no known neurological or developmental disorders. Participants aged 11 or 12 years verbally consented, while 13- to 15-year-olds provided written consent, in accordance with the guidelines of the local ethics committee, which approved the study. In contrast to the previous studies, the sample is predominantly boys (**Table 7.1**).

**Table 7.1.** Participant characteristics. There was no gender difference in mean age, $p = .644$.

| | | **Age (years)** | | **Girls : Boys** |
|---|---|---|---|---|
| **Age group** | ***n*** | ***M (SD)*** | ***Range*** | ***n*** |
| 12y | 30 | 12.05 (0.34) | 11.50-12.67 | 7:23 |
| 13y | 30 | 13.12 (0.36) | 12.33-13.67 | 7:23 |
| 14y | 30 | 14.08 (0.39) | 13.41-14.75 | 7:23 |
| 15y | 30 | 15.15 (0.33) | 14.42-15.67 | 7:23 |

Two different schools were recruited for the study, the first of which was an English state secondary school where most students are from minority ethnic heritages, and the proportion of free school meals (determined by parental income-related benefits) is well above average, with 7 girls and 8 boys recruited per year group, while the second was an English fee-paying independent secondary boys' school where performance is above the national average, with 15 boys per year group taking part. It was initially intended that each year group would have half girls and half boys, but recruitment for this study proved challenging and the boys' school was particularly keen to be involved.

## 7.3.2 Tasks

### 7.3.2.1 Science and maths misconceptions

The science and maths misconceptions task was administered in a similar manner to **Chapter 3** (see section 2.1 for details). Science and maths problems were intermixed, as in the first behavioural study, with a 12 s limit and a 9 s warning as in the fMRI study, to ensure the maximum time spent on this task was under 20 min.

### 7.3.2.2 Inhibitory control

The complex Go/No-Go measured response inhibition and was administered as in **Chapter 3** (see section 2.2.1 for details). The numerical Stroop measured semantic inhibition, and had half as many trials as in **Chapter 3** (see section 2.2.2 for details), to reduce the time spent on this task, and trials were presented in two blocks. The Hayling sentence completion test was administered as in the manual, although it was programmed in MATLAB (The MathWorks, Inc., Natick, MA) so that the researcher could type answers in and track RTs with the laptop rather than a stopwatch. In section 1 of the test, 15 sentences missing their last word are read by the experimenter, and the participant must, as quickly as possible, say a word that completes the sentences (the *sensible completion* condition, **Table 7.2**). In section 2 of the test, again 15 sentences missing their last word are read by the experimenter and this time the participant must say a word as quickly as possible that does not fit at the end of the sentences (the *unconnected* condition, **Table 7.2**). Two practice items are given before the main test begins, to give the researcher an opportunity to explain the requirements of the task if they are not understood. The task measures semantic inhibition in the context of verbal and prior knowledge requirements. The Hayling overall scaled score was calculated as per the manual, by adding together an RT score from section 1, an RT score from section 2, and a score from section 2 based on number and type of errors given (i.e. if the word is connected or somewhat connected).

**Table 7.2.** Example sentences missing their last word in the Hayling sentence completion test, taken from Burgess and Shallice (1996).

| **Unconnected sentences** |
|---|
| The captain wanted to stay with the sinking |
| He bought them in the sweet |
| They went as far as they |
| The whole town came to hear the mayor |
| Most sharks attack very close to |
| She called the husband at his |
| None of the books made any |
| Most cats see very well at |
| The dog chased our cat up the |
| The dough was put in the hot |
| Jean was glad the affair was |

### 7.3.2.3 Relational reasoning

The verbal analogical reasoning task was administered on a laptop and adapted from **Chapter 6** (see section 6.3.2.3 for details). This time the task was programmed in MATLAB (The MathWorks, Inc., Natick, MA), rather than as an online form. This enabled the task to be used on a laptop in school where there was no internet, and meant that only one question was presented at a time. There was no time limit for each question, as in the fMRI study. A few items were altered to make them gender neutral (e.g. fireman to firefighter) or to use British English (leash to lead). A self-timed break screen was included halfway through. The total number of correct responses out of 24 trials was recorded.

Scale 3 Form A of the CCF was administered as per the instructions, using the stimulus booklet and a response sheet whereby the participant marked their answer with a pencil. The booklet contained four non-verbal relational reasoning tests, measuring different aspects of relational reasoning: series, classifications, matrices, and conditions (topology). Instructions were read from the booklet verbatim. Each test in the booklet had a fixed time limit, and the task lasted a total of 12.5 minutes (not including instructions). The total number of correct responses was recorded for each scale, and the overall number of correct responses from all scales was calculated.

#### 7.3.2.4 Working memory

The backwards digit span measured VWM and was administered as in **Chapter 6** (see section 6.3.2.5 for details).

### 7.3.3 Procedure

Participants were tested in a quiet space in school for approximately 45 min during the school day. The experimenter described each computerised task, emphasising that responses should be given as quickly and accurately as possible. The tasks were performed in the following order: analogical reasoning, science and maths misconceptions, complex Go/No-Go, numerical Stroop, CCF. Due to constraints at the schools, it was not always possible to complete the full testing session. The Hayling sentence completion and backwards digit span were therefore completed at the end where time permitted, as these were considered to be less central to the main research questions. Participants were given no reward for taking part and were not provided with their results on any task, and it was explained that their responses would remain anonymous and independent of school assessments.

### 7.3.4 Statistical analysis

Exclusionary criteria were put in place before analysis commenced. Participants whose mean accuracy or RT was further than ±3.29 *SD*s away from the group mean were excluded from analyses of the task on which they were an outlier. Main effects of Age group were followed up with three planned tests assessing differences between 12y and 15y, 13y and 15y, and 14y and 15y, since the greatest differences were anticipated in comparison to the oldest group.

#### 7.3.4.1 Science and maths misconceptions

There were no outliers on this task (final $N = 120$). Two (Trial type: control, misconception) x two (Discipline: science, maths) x two (Statement type: true, false) x four (Age group: 12y, 13y, 14y, 15y) mixed model repeated measures ANOVAs were performed on accuracy and RT in the science and maths misconceptions task.

#### 7.3.4.2 Inhibitory control

Due to low accuracy, three participants were excluded from the complex Go/No-Go task (two 12y, one 15y, final $N = 117$), and two participants were excluded from the numerical Stroop (two 13y, final *N* = 118). Two (Trial type: Go, No-Go or congruent, incongruent) by four (Age group: 12y, 13y, 14y, 15y) mixed model repeated measures ANOVAs were performed on accuracy in the complex Go/No-Go and accuracy and RT in the numerical Stroop. A one-way (Age group: 12y, 13y, 14y, 15y) ANOVA was performed on RT in the complex Go/No-Go.

There were no outliers on the Hayling sentence completion test, although one participant was not included due to experimenter error in test administration (final *n*s: 12y: *n* = 16, 13y: *n* = 17, 14y: *n* = 23, 15y: *n* = 20, total *N* = 76 since not all participants completed the task). A one-way (Age group: 12y, 13y, 14y, 15y) ANOVA was performed on the Hayling overall scaled score, which does not take account of age.

### 7.3.4.3 Relational reasoning

There were no outliers on either relational reasoning test (final *N* = 120). The four CCF subscales correlated highly with each other, *r*'s between .429 and .482, and all scales were combined for analyses. A one-way (Age group: 12y, 13y, 14y, 15y) ANOVA was performed on the total number of correct responses in the analogical reasoning task and the CCF score.

### 7.3.4.4 Working memory

There were no outliers on the backwards digit span measuring VWM (final *N* = 75 since not all participants completed the task, range of scores = 1-18 *M* = 8, *SD* = 4).

### 7.3.4.5 Regression analyses

#### 7.3.4.5.1 Overall science and maths reasoning

Participants excluded from the Go/No-Go or Stroop analyses were also excluded from the regression analyses (final *n*s: 12y: *n* = 28, 13y: *n* = 28, 14y: *n* = 30, 15y: *n* = 29), leaving a final total *N* = 115. Correlations were run between the variables of interest to examine collinearity and assess associations between measures on all tasks. Hierarchical multiple regressions investigated whether inhibitory control and relational reasoning variables could account for individual differences in science and maths accuracy and RT.

Regression models added age in months using the enter function in block 1. Inhibitory control variables were entered stepwise in block 2: complex No-Go accuracy, Stroop accuracy cost, and Stroop RT cost. Relational reasoning variables were entered stepwise in block 3: CCF score, and analogical reasoning score. Variables were entered in this order, as the results from **Chapters 3 and 6** indicated that inhibitory control was likely to be a significant but small predictor, with relational reasoning a larger predictor when it has an effect. Entering inhibition variables first meant that it would be possible to examine inhibitory control effects on their own first. This also follows the logic of **Chapter 6** where executive function measures were entered first, and the analysis then considered the extent to which relational reasoning could explain variance over and above executive function.

#### 7.3.4.5.2 Counterintuitive science and maths reasoning

Correlations were run between misconception and control accuracy and RT with the other variables of interest to examine collinearity and assess associations. Hierarchical multiple regressions investigated whether inhibitory control and relational reasoning variables could account for individual differences in science and maths misconception accuracy and RT, when accounting for control performance.

Regression models added age in months and control accuracy or RT using the enter function in block 1. Inhibitory control variables were entered stepwise in block 2: complex No-Go accuracy, Stroop accuracy cost, and Stroop RT cost. Relational reasoning variables were entered stepwise in block 3: CCF score, and analogical reasoning score.

#### 7.3.4.5.3 Exploratory regressions

Exploratory regression analyses were run on the sample who completed the Hayling sentence completion and VWM tasks (final *n*s: 12y: $n$ = 14, 13y: $n$ = 17, 14y: $n$ = 23, 15y: $n$ = 19, leaving a final total $N$ = 73). The same regression models were run as above, with VWM total entered as a control variable in block 1 and Hayling sentence completion overall scaled score entered as an inhibitory control variable in block 2.

Where a regression was significant for science and maths combined, a follow up exploratory regression was run on science and maths separately, to examine possible discipline-specific effects and to explore whether directions of association were consistent between disciplines. The follow up model included the control variables and the reasoning and inhibitory control variables identified in the science and maths combined regression, using the enter method.

# 7.4 Results

## 7.4.1 Science and maths misconceptions

A two (Trial type: control, misconception) x two (Discipline: science, maths) x two (Statement type: true, false) x four (Age group: 12y, 13y, 14y, 15y) mixed model repeated measures ANOVA performed on accuracy showed main effects of Trial type, Discipline and Statement type, with greater accuracy in control compared to misconception trials, science compared to maths trials, and true compared to false statements (**Table 7.3**).

**Table 7.3.** Accuracy and RT estimated marginal means in the science and maths misconceptions task.

| | | **Accuracy (%)** | **RT (ms)** |
|---|---|---|---|
| | | ***M (SE)*** | ***M (SE)*** |
| **Main effects** | | | |
| *Trial type* | | $F(1, 116) = 632.57$, $p < .001$, $\eta_p^2 = .845$ | $F(1, 116) = 352.55$, $p < .001$, $\eta_p^2 = .752$ |
| Control | | 82.3 (0.9) | 5601 (77) |
| Misconception | | 59.2 (1.1) | 6425 (79) |
| *Discipline* | | $F(1, 116) = 11.74$, $p < .001$, $\eta_p^2 = .092$ | $F(1, 116) = 24.11$, $p < .001$, $\eta_p^2 = .172$ |
| Science | | 72.1 (0.9) | 5874 (87) |
| Maths | | 69.4 (1.0) | 6153 (72) |
| *Statement type* | | $F(1, 116) = 18.98$, $p < .001$, $\eta_p^2 = .141$ | n.s., $p = .732$ |
| True | | 73.3 (1.0) | 6005 (83) |
| False | | 68.2 (1.1) | 6022 (74) |
| *Age group* | | $F(3, 116) = 4.14$, $p = .008$, $\eta_p^2 = .097$ | n.s., $p = .367$ |
| 12y | | 65.5 (1.8) | 6130 (149) |
| 13y | | 71.4 (1.8) | 6142 (149) |
| 14y | | 72.1 (1.8) | 5965 (149) |
| 15y | | 73.9 (1.8) | 5817 (149) |
| **Interaction effects** | | | |
| *Trial type* | *Discipline* | $F(1,116) = 17.89$, $p = .001$, $\eta_p^2 = .134$ | n.s., $p = .614$ |
| Control | Science | 81.9 (1.0) | 5472 (89) |
| | Maths | 82.6 (1.1) | 5731 (77) |

| | | | |
|---|---|---|---|
| Misconception | Science | 62.3 (1.2) | 6276 (92) |
| | Maths | 56.1 (1.4) | 6576 (80) |
| *Trial type* | *Statement type* | $F$(1, 116) = 6.34, $p$ = .013, $\eta_p^2$ = .052 | n.s., $p$ = .060 |
| Control | True | 83.7 (1.1) | 5543 (89) |
| | False | 80.8 (1.1) | 5660 (78) |
| Misconception | True | 62.9 (1.3) | 6467 (91) |
| | False | 55.6 (1.5) | 6384 (85) |
| *Discipline* | *Statement type* | $F$(1, 116) = 32.96, $p$ < .001, $\eta_p^2$ = .221 | n.s., $p$ = .384 |
| Science | True | 77.3 (1.0) | 5885 (97) |
| | False | 66.9 (1.4) | 5862 (88) |
| Maths | True | 69.3 (1.3) | 6124 (85) |
| | False | 69.5 (1.2) | 6182 (77) |

These main effects were modulated by interactions that were followed up with repeated measures ANOVAs. The interaction between Trial type and Discipline was attributable to a difference between science and maths for misconception trials, $F(1, 116) = 27.27$, $p < .001$, $\eta_p^2 = .190$, and not control trials, $p = .534$. The interaction between Trial type and Statement type was explained by a greater difference in accuracy between true and false statements in misconception trials, $F(1,116) = 20.52$, $p = .001$, $\eta_p^2 = .150$, compared to control trials, $F(1,116) = 5.48$, $p = .021$, $\eta_p^2 = .045$. Finally, the interaction between Discipline and Statement type was due to a significant difference between true and false trials for science, $F(1,116) = 40.81$, $p < .001$, $\eta_p^2 = .260$, and not maths, $p = .923$. Overall accuracy increased with age and there was no interaction between age and any other variable. Follow-up planned comparisons revealed a significant difference only between 12y and 15y, $p = .001$.

A two (Trial type: control, misconception) x two (Discipline: science, maths) x two (Statement type: true, false) x four (Age group: 12y, 13y, 14y, 15y) mixed model repeated measures ANOVA performed on RT showed main effects of Trial type and Discipline, with faster RTs in control compared to misconception trials and in science compared to maths (**Table 7.3**). There was no main effect of Statement type or Age group. There was a significant three-way interaction between Trial type, Discipline, and Statement type, $F(1, 116) = 6.20$, $p = .014$, $\eta_p^2 = .051$. This was followed up with further repeated measures ANOVAs, demonstrating that the difference between true and false was significant only for control maths trials, $F(1,116) = 10.65$, $p = .001$, $\eta_p^2 = .084$, with faster RTs for true ($M$ = 5602 ms) compared to false trials ($M$ = 5860 ms).

Overall the science and maths misconceptions task showed the main anticipated effect of lower accuracy and longer RTs in misconception trials compared to control trials, and an improvement in accuracy with age.

### 7.4.2 Inhibitory control

A two (Trial type: Go, No-Go) x four (Age group: 12y, 13y, 14y, 15y) mixed model repeated measures ANOVA on accuracy in the complex Go/No-Go showed a significant effect of Trial type, with higher accuracy in Go trials than No-Go trials (**Table 7.4**). There was no significant effect of Age group, and no interaction between Trial type and Age group for accuracy, $p = .774$. A one-way ANOVA (Age group: 12y, 13y, 14y, 15y) on RT showed no significant effect of age (**Table 7.4**).

**Table 7.4.** Accuracy and RT estimated marginal means in the complex Go/No-Go and numerical Stroop.

| | **Complex Go/No-Go** | **Numerical Stroop** |
|---|---|---|
| | ***M (SE)*** | ***M (SE)*** |
| **Accuracy (%)** | | |
| *Trial type* | $F(1, 113) = 275.87$, $p < .001$, $\eta_p^2 = .709$ | $F(1, 114) = 178.02$, $p < .001$, $\eta_p^2 = .610$ |
| Go/Congruent | 88.9 (0.9) | 96.3 (0.4) |
| No-Go/Incongruent | 62.4 (1.6) | 84.0 (1.0) |
| *Age group* | n.s., $p = .483$ | n.s., $p = .126$ |
| 12y | 75.6 (2.0) | 88.3 (1.2) |
| 13y | 76.1 (2.0) | 90.7 (1.3) |
| 14y | 77.7 (2.0) | 89.4 (1.2) |
| 15y | 73.3 (2.0) | 92.2 (1.2) |
| **RT (ms)** | | |
| *Trial type* | - | $F(1, 114) = 261.94$, $p < .001$, $\eta_p^2 = .967$ |
| Go/Congruent | 406 (5) | 714 (10) |
| No-Go/Incongruent | - | 800 (10) |
| *Age group* | n.s., $p = .135$ | $F(3, 114) = 2.82$, $p = .042$, $\eta_p^2 = .069$ |
| 12y | 422 (10) | 805 (20) |
| 13y | 405 (9) | 751 (20) |
| 14y | 407 (9) | 741 (20) |
| 15y | 389 (9) | 731 (20) |

A two (Trial type: congruent, incongruent) x four (Age group: 12y, 13y, 14y, 15y) mixed model repeated measures ANOVA on accuracy in the numerical Stroop showed a significant effect of Trial type, with higher accuracy in congruent trials than incongruent trials (**Table 7.4**). There was no effect of Age group on accuracy, and no interaction between Trial type and Age group, $p$ = .430. A repeated measures ANOVA on RT showed a significant effect of Trial type, with faster RTs in congruent trials than incongruent trials (**Table 7.4**). There was a main effect of Age group on RT. Planned comparisons showed that 12y was the only Age group significantly different to 15y, $p$ = .009.

A one-way (Age group: 12y, 13y, 14y, 15y) ANOVA showed no effect of Age group in the Hayling sentence completion, $p$ = .182, overall $M$ = 5.5, $SD$ = 1.1.

Overall the Go/No-Go and Stroop tasks showed the anticipated inhibition effects of lower accuracy in trials requiring inhibition, and longer RTs in trials requiring inhibition in the numerical Stroop. Age effects were observed in numerical Stroop RTs where the oldest participants were faster than the youngest.

## 7.4.3 Relational reasoning

### 7.4.3.1 Overall science and maths reasoning

Descriptive statistics for the CCF, analogical reasoning, and VWM are reported in **Table 7.5**. One-way (Age group: 12y, 13y, 14y, 15y) ANOVAs showed no effect of Age group in the CCF score, $p$ = .744, or analogical reasoning, $p$ = .617.

**Table 7.5.** Estimated marginal means and range in the CCF, analogical reasoning, and VWM by age. [a] indicates the subsample size that refers only to VWM.

| | | **Age Group** | | | |
|---|---|---|---|---|---|
| | | **12y** | **13y** | **14y** | **15y** |
| **CCF score** | ***M*** | 21.8 (1.1) | 22.5 (1.1) | 22.5 (1.1) | 23.5 (1.1) |
| | ***Range*** | 11-32 | 13-33 | 9-36 | 10-35 |
| **Analogical Reasoning score** | ***M*** | 16.7 (0.6) | 17.5 (0.6) | 16.9 (0.6) | 17.8 (0.6) |
| | ***Range*** | 9-23 | 10-23 | 10-22 | 10-23 |
| **VWM total** | ***M*** | 7.3 (0.9) | 8.5 (0.9) | 7.5 (0.7) | 8.6 (0.8) |
| | ***Range*** | 4-12 | 2-12 | 1-17 | 2-18 |
| | ***n*[a]** | 15 | 17 | 23 | 20 |

### 7.4.4 Regression analyses

#### 7.4.4.1 Overall science and maths reasoning

Correlations between the variables of interest were examined and assumptions regarding multicollinearity were met (**Table 7.6**). Note that there was a small positive correlation between age in months and performance on the Hayling sentence completion, and between age in months and CCF score, where there were no age effects in terms of age group.

**Table 7.6.** Pearson correlation coefficients of regression variables for science and maths by discipline. Statistically significant (two-tailed) correlations are highlighted in bold, [a,b,c] indicate $p$ <.05, $p$ < .01 and $p$ < .001 respectively, [d,e] denote subsamples of participants where $n$ = 76 and $n$ = 75 respectively.

| **Variable** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Science accuracy | | | | | | | | | | | |
| 2. Science RT | **-.22**[a] | | | | | | | | | | |
| 3. Maths accuracy | **.69**[c] | **-.26**[b] | | | | | | | | | |
| 4. Maths RT | -.02 | **.73**[c] | -.11 | | | | | | | | |
| 5. Age (months) | **.36**[c] | **-.26**[b] | **.29**[b] | -.18 | | | | | | | |
| 6. Complex No-Go accuracy | .09 | -.09 | .11 | -.13 | -.04 | | | | | | |
| 7. Stroop accuracy cost | .09 | -.01 | .01 | .07 | -.11 | .18 | | | | | |
| 8. Stroop RT cost | .13 | -.11 | .00 | .03 | -.17 | .04 | **.36**[a] | | | | |
| 9. CCF score | **.64**[c] | -.07 | **.71**[c] | -.07 | **.19**[a] | .15 | **.19**[a] | -.01 | | | |
| 10. Analogical reasoning score | **.54**[c] | -.01 | **.59**[c] | .07 | .14 | .00 | -.04 | .06 | **.51**[c] | | |
| 11. Hayling scaled score[d] | .21 | **-.34**[b] | **.23**[a] | **-.27**[a] | **.23**[a] | .22 | .14 | .04 | **.27**[a] | .10 | |
| 12. VWM total[e] | **.30**[a] | .07 | **.31**[b] | .08 | .08 | .10 | -.15 | -.10 | **.28**[b] | **.29**[a] | -.01 |

The first multiple regression investigated whether inhibitory control and relational reasoning could explain unique variation in science and maths accuracy (**Table 7.7**). Model 1a only included age and was significant, explaining 12% of the variance in accuracy. No inhibitory control variables were selected. Model 1b selected CCF score, explaining an additional 47% of the variance, and finally model 1c selected analogical reasoning score, explaining an extra 7% of the variance. Higher science and maths accuracy was associated with increasing age and better performance on the CCF and analogical reasoning tests. The second hierarchical multiple regression investigated which measures could explain variance in science and maths RT (**Table 7.7**). Model 2a, with age as the predictor, was significant, explaining just 6% of the variance. Older participants were faster on the task. No inhibitory control or relational reasoning measures were selected in further steps.

**Table 7.7.** Regression models for science and maths accuracy and RT. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science and maths accuracy | | | | |
| **Model 1a** | **Constant** | | **2.97** | **.004** |
| $F(1, 113) = 16.00$, $p < .001$, $R^2 = 12\%$ | **Age (months)** | **.35** | **4.00** | **< .001** |
| **Model 1b** | **Constant** | | **2.69** | **.008** |
| $F(2, 112) = 80.65$, | **Age (months)** | **.22** | **3.57** | **.001** |
| $p < .001$, $R^2 = 59\%$, $\Delta R^2 = 46.6\%$ | **CCF score** | **.70** | **11.29** | **< .001** |
| **Model 1c** | Constant | | 1.65 | .101 |
| $F(3, 111) = 71.61$, | **Age (months)** | **.21** | **3.64** | **< .001** |
| $p < .001$, $R^2 = 66\%$ | **CCF score** | **.54** | **8.31** | **< .001** |
| $\Delta R^2 = 6.9\%$ | **Analogical reasoning score** | **.31** | **4.75** | **< .001** |
| Science and maths RT | | | | |
| **Model 2a** | **Constant** | | **10.01** | **< .001** |
| $F(1, 113) = 6.90$, $p = .010$, $R^2 = 6\%$ | **Age (months)** | **-.24** | **-2.63** | **.010** |

The accuracy regressions were followed up by running them again for each discipline separately (**Table 7.8**), with the significant variables from the combined regressions inserted using the enter method. In both cases (models 3 and 4), all variables remained significant predictors with 52% of the variance explained in science and 60% of the variance explained in maths.

**Table 7.8.** Regression models for science and maths accuracy by discipline. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science accuracy | | | | |
| **Model 3** | Constant | | 1.38 | .170 |
| $F(3, 111) = 40.41$, | **Age (months)** | **.23** | **3.48** | **.001** |
| $p < .001$, $R^2 = 52\%$ | **CCF score** | **.46** | **5.97** | **< .001** |
| | **Analogical reasoning score** | **.27** | **3.49** | **.001** |
| Maths accuracy | | | | |
| **Model 4** | Constant | | 1.30 | .197 |
| $F(3, 111) = 54.71$, | **Age (months)** | **.15** | **2.41** | **.018** |
| $p < .001$, $R^2 = 60\%$ | **CCF score** | **.53** | **7.51** | **< .001** |
| | **Analogical reasoning score** | **.30** | **4.21** | **< .001** |

### 7.4.4.2 Counterintuitive science and maths reasoning

Next, analyses concerned misconception performance. Correlations (**Table 7.9**) showed that both analogical reasoning and the CCF were positively associated with accuracy in both trial types, with no significant associations with RT. The Hayling was associated with faster RTs in both trial types, and higher accuracy in control trials only, although also in the positive direction for misconception accuracy. There were no significant associations with the other inhibitory control variables. Finally, VWM was positively associated with accuracy in both trial types.

**Table 7.9.** Pearson correlation coefficients of regression variables for science and maths split by misconception and control trials. Statistically significant (two-tailed) correlations are highlighted in bold, [a,b,c] indicate $p$ <.05, $p$ < .01 and $p$ < .001 respectively, [d,e] denote subsamples of participants where $n$ = 76 and $n$ = 75 respectively.

| Variable | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Misconception accuracy | | | | |
| 2. Misconception RT | -.11 | | | |
| 3. Control accuracy | **.62**[c] | .02 | | |
| 4. Control RT | **-.30**[b] | **.82**[c] | **-.24**[a] | |
| 5. Age (months) | **.36**[c] | **-.23**[a] | **.27**[b] | **-.23**[a] |
| 6. Complex No-Go accuracy | .13 | -.11 | .06 | -.11 |
| 7. Stroop accuracy cost | .03 | .03 | .06 | .03 |
| 8. Stroop RT cost | -.01 | -.04 | .14 | -.06 |
| 9. CCF score | **.67**[c] | .03 | **.66**[c] | -.17 |
| 10. Analogical reasoning score | **.55**[c] | .12 | **.56**[c] | -.07 |
| 11. Hayling scaled score[d] | .19 | **-.30**[b] | **.27**[a] | **-.33**[b] |
| 12. VWM total[e] | **.29**[a] | .08 | **.32**[b] | .07 |

Next, regressions were run on science and maths misconception performance while controlling for performance in control trials (**Table 7.10**). Model 5a was significant, explaining 42% of the variance in misconception accuracy, with age and control performance both significant predictors. No inhibitory control variables were selected. CCF score was added in model 5b which explained an additional 12% of the variance. Analogical reasoning score was added in model 5c and explained an extra 3% of the variance in misconception accuracy. The misconception RT regression was significant, with model 6a explaining 67% of the variance. Age was not a significant predictor, but control RT was. No inhibitory control variables were selected. Analogical reasoning score was entered in model 6b and explained an additional 3% of the variance, with better analogical reasoning associated with longer RTs.

**Table 7.10.** Regression models for science and maths misconception accuracy and RT. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science and maths misconception accuracy | | | | |
| **Model 5a** | **Constant** | | **-2.44** | **.016** |
| $F(2, 112) = 41.25$, | **Age (months)** | **.20** | **2.69** | **.008** |
| $p < .001$, $R^2 = 42\%$ | **Science and maths control accuracy** | **.57** | **7.62** | **< .001** |
| **Model 5b** | Constant | | -1.69 | .094 |
| $F(3, 111) = 43.70$, | **Age (months)** | **.20** | **2.92** | **.004** |
| $p < .001$, $R^2 = 54\%$, | **Science and maths control accuracy** | **.27** | **3.09** | **.003** |
| $\Delta R^2 = 11.7\%$ | **CCF score** | **.46** | **5.33** | **< .001** |
| **Model 5c** | Constant | | -1.98 | .050 |
| $F(4, 110) = 36.62$, | **Age (months)** | **.20** | **3.05** | **.003** |
| $p < .001$, $R^2 = 57\%$ | **Science and maths control accuracy** | **.19** | **2.07** | **.041** |
| $\Delta R^2 = 3.0\%$ | **CCF score** | **.40** | **4.69** | **< .001** |
| | **Analogical reasoning score** | **.21** | **2.76** | **.007** |
| Science and maths misconception RT | | | | |
| **Model 6a** | **Constant** | | **3.55** | **.001** |
| $F(2, 112) = 115.06$, | Age (months) | -.05 | -0.90 | .369 |
| $p < .001$, $R^2 = 67\%$ | **Science and maths control RT** | **.81** | **14.55** | **< .001** |
| **Model 6b** | **Constant** | | **2.76** | **.007** |
| $F(3, 111) = 88.43$, | Age (months) | -.07 | -1.39 | .167 |
| $p < .001$, $R^2 = 71\%$, | **Science and maths control RT** | **.81** | **15.38** | **< .001** |
| $\Delta R^2 = 3.2\%$ | **Analogical reasoning score** | **.18** | **3.49** | **.001** |

Misconception regressions were followed up by discipline. For accuracy in each of science and maths, the CCF score and analogical reasoning score both remained significant predictors of performance (**Table 7.11**). In science, model 7 explained 46% of the variance, with age and control accuracy both significant predictors in addition to the relational reasoning measures. In maths, model 8 explained 41% of the variance, with age a marginal predictor of performance, and control accuracy not a predictor.

**Table 7.11.** Regression models for science and maths misconception accuracy by discipline. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science misconception accuracy | | | | |
| **Model 7** $F(4, 110) = 23.57$, $p < .001$, $R^2 = 46\%$ | **Constant** | | **-2.04** | **.044** |
| | **Age (months)** | **.23** | **3.24** | **.002** |
| | **Science control accuracy** | **.19** | **2.21** | **.029** |
| | **CCF score** | **.29** | **3.16** | **.002** |
| | **Analogical reasoning score** | **.24** | **2.83** | **.006** |
| Maths misconception accuracy | | | | |
| **Model 8** $F(4, 110) = 18.70$, $p < .001$, $R^2 = 41\%$ | Constant | | -0.60 | .547 |
| | Age (months) | .13 | 1.77 | .080 |
| | Maths control accuracy | -.01 | -0.06 | .951 |
| | **CCF score** | **.48** | **4.93** | **< .001** |
| | **Analogical reasoning score** | **.20** | **2.15** | **.034** |

Analogical reasoning remained a significant predictor of both science and maths misconception RT (**Table 7.12**). In science, model 9 explained 66% of the variance, with control RT but not age a significant predictor. In maths, model 10 explained 46% of the variance, with age and control RT both significant predictors of misconception RT.

**Table 7.12.** Regression models for science and maths misconception RT by discipline. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science misconception RT | | | | |
| **Model 9** $F(3, 111) = 72.25$, $p < .001$, $R^2 = 66\%$ | Constant | | 1.32 | .190 |
| | Age (months) | .00 | -0.07 | .947 |
| | **Science control RT** | **.81** | **14.08** | **< .001** |
| | **Analogical reasoning score** | **.16** | **2.86** | **.005** |
| Maths misconception RT | | | | |
| **Model 10** $F(3, 111) = 31.11$, $p < .001$, $R^2 = 46\%$ | **Constant** | | **4.28** | **< .001** |
| | **Age (months)** | **-.16** | **-2.23** | **.028** |
| | **Maths control RT** | **.62** | **8.88** | **< .001** |
| | **Analogical reasoning score** | **.17** | **2.39** | **.019** |

#### 7.4.4.3 Exploratory regressions

Hierarchical multiple regressions were then run on the subsample who had Hayling sentence completion and VWM scores. For brevity, a summary is provided here and the regressions are reported fully in **Appendix 3**. For overall science and maths accuracy, VWM and Hayling scores were significant positive predictors, but when the relational reasoning measures were included in the model, they were no longer significant. This pattern held for science and maths accuracy separately. For science and maths overall RT, VWM total was not a significant predictor, while the Hayling was a predictor, explaining 11% of the variance in RT. Faster RTs were associated with better scores on the Hayling sentence completion test. Again, the pattern remained for both science and maths when they were considered separately. For science and maths misconception accuracy and RT, neither measure was a significant predictor of performance.

In summary, the regression analyses showed that both relational reasoning measures predicted accuracy in science and maths, with the CCF explaining more variance than analogical reasoning. Those who had better relational reasoning scores also had higher science and maths accuracy. The same pattern was seen for misconception accuracy, when controlling for accuracy in control trials. For overall science and maths RT, there were no significant predictors, but analogical reasoning was a significant positive predictor of misconception RT: those who were better at analogical reasoning took longer to answer misconception problems. The relational reasoning associations remained significant predictors when working memory was added to the models. No Go/No-Go or Stroop variables were selected by any model. However, when verbal semantic inhibition as measured by the Hayling sentence completion test was included, it was a significant predictor of RT in science and maths.

## 7.5 Discussion

The aim of this study was to determine the relative importance of inhibitory control and relational reasoning in science and maths problem-solving. There were three measures of inhibitory control, measuring response inhibition, semantic inhibition, and semantic inhibition in the context of increased verbal and prior knowledge requirements. Verbal analogical and non-verbal relational reasoning measures were used. Results showed that both relational reasoning measures were predictors of performance in science and maths general performance, and specifically in misconception performance. The only inhibitory control variable that was related to science and maths performance was the Hayling sentence completion test, which had greater reliance on verbal ability and prior knowledge.

### 7.5.1 Inhibitory control and science and maths

Contrary to the hypothesis and the results from **Chapter 3**, inhibitory control as measured by the complex Go/No-Go and numerical Stroop did not predict misconception performance. Inhibitory control measures were entered into the models before the relational reasoning measures, so it is not the case that relational reasoning is explaining the same variance as inhibitory control. In this study, the simple Go/No-Go was not included and the numerical Stroop was halved in order to reduce the length of the testing session. It is possible that the shorter time spent on the inhibitory control tasks led to a different pattern of performance, which was less predictive of misconceptions. Nonetheless, the anticipated inhibitory control effects were observed: overall higher accuracy in Go and congruent trials compared to No-Go and incongruent trials, and longer RTs in incongruent trials.

The Hayling sentence completion test was included as a measure of verbal semantic inhibition that may be more similar to the type of inhibition required during science and maths reasoning. The test is typically used in clinical populations so linking it to science and maths was novel here. Stroop performance did not correlate with Hayling sentence completion performance, which corroborates previous indications that the two tasks capture unique aspects of inhibition. This may be because strategy implementation is necessary for the Hayling, whereas in the Stroop, there is a constrained response set, possibly with much less of a requirement to implement a strategy (Cipolotti et al., 2016). Correlations showed that those who were better on the Hayling sentence completion test also had higher accuracy and faster RTs in the science and maths task. In the regression analyses, controlling for age and working memory, the Hayling was a significant predictor of overall accuracy and RT, which held for both science and maths separately. Only the association with RT remained when relational reasoning measures were entered into the models. Scores on the Hayling sentence completion test did not associate with misconception performance when accounting for control performance. These findings combined may highlight the verbal and knowledge-based nature of the science and maths task, such that the ability to suppress semantic verbal information is important in all types of problem.

The results provide little evidence for a role of individual differences in inhibitory control in counterintuitive science and maths reasoning in adolescence, and also in science and maths more generally. Previous research relating inhibitory control to science has been inconsistent. One example showed that inhibitory control measured using a gift delay paradigm related to 3- and 4-year-olds' ability to predict the movement of an object down an opaque tube (Baker et al., 2011). Another example linked a general measure of executive function that included inhibition to understanding of biology in 5- to 7-year-olds (Zaitchik et al., 2014). These measures of inhibitory control were very different to those used in **Chapters 3 and 7**. It is possible that the aggregate executive function measure used by Zaitchik and colleagues (2014) did not reflect inhibition processes at all, since the different components were not separated. The findings here are more consistent with a study of 10-year-olds that found no link between semantic inhibition and scientific reasoning (Mayer et al., 2014), and learning studies that showed no association between response inhibition and learning biology (Rhodes et al., 2014) or chemistry (Rhodes et al., 2016) or physics (Rhodes, n.d.) in 12- to 13-year-olds. Given the ages tested in the studies described, it is possible that inhibitory control is indeed important in science, but that individual differences in inhibitory control does not relate to science performance in adolescents. It may be more important for younger participants who have not yet reached the baseline level of ability that enables them to suppress misleading information or old theories. This might explain why the studies with younger children show a link (Baker et al., 2011; Zaitchik et al., 2014) while those with older children (Mayer et al., 2014) and adolescents (Rhodes et al., 2014, 2016) do not.

This argument is less plausible in maths, where there have been links between individual differences in inhibitory control and maths in 3- to 6-year-olds (Merkley et al., 2016), as well as in 11- to 14-year-olds (Gilmore et al., 2015; Khng & Lee, 2009). In addition, a negative priming study showed both 11- and 14-year-olds to respond more slowly to a maths problem that followed a prime with an incongruent relational term (Lubin et al., 2013). While they did not investigate individual differences on an inhibitory control task, the findings suggest that participants needed to inhibit the incorrect strategy that was primed. Nonetheless, it has been suggested that inhibitory control might be important in maths at younger ages, when less sophisticated strategies for problem-solving need to be inhibited, but then again later when number facts need to be suppressed (Cragg & Gilmore, 2014). According to this argument, inhibitory control's role in maths takes an inverse U-shaped function, although the specific ages are unclear. The results of **Chapter 3 and 7** do not convincingly implicate individual differences in inhibitory control in science and maths, but it remains possible that once a certain level of competence in inhibition is reached, there may be no advantage of better inhibitory control in the context of these misconception and control problems.

This conclusion has important implications for individuals with developmental disorders where inhibitory control may be impaired, such as attention deficit hyperactivity disorder (ADHD) (Simmonds, Pekar, & Mostofsky, 2008). For these adolescents, poor inhibitory control may mean that they are less able to suppress their intuitive but incorrect response, in spite of knowing the correct answer. Those with inhibitory control deficits therefore may require extra support in the classroom, to help them reason more effectively about counterintuitive concepts.

### 7.5.2 Relational reasoning and science and maths

The CCF showed a weak positive correlation with age, while the analogical reasoning task showed a non-significant positive correlation with age. Previous studies have shown maturation in relational reasoning between 5 and 17 years in one study (Jablansky et al., 2015), between 3 and 14 years in another (Richland et al., 2006), and between 6 and 14 years in another (Whitaker et al., 2017), and as such, a greater age effect was anticipated here than was found. This finding is similar to that seen in **Chapter 6** where age did not relate to analogical reasoning or matrix reasoning, although that was thought to be related to the small sample size in part. Nonetheless, one study showed the largest improvement from 6 to 10 years of age, and ceiling level performance at age 14 (Whitaker et al., 2017), although this was likely due to the nature of the task which relied on pictures of familiar objects. Performance on the relational reasoning tasks in the current study was not at ceiling, and so the minimal age effects may simply reflect the relatively small age range tested here. When age was considered as a categorical factor according to school year group, age showed no association with relational reasoning, which highlights the difficulty in providing the best description of age effects. Continuous analyses better reflect the true values of individuals' ages, but group analyses are also of interest because these participants are grouped for their learning, and groups therefore reflect the time spent in formal learning environments. Considering both, as presented here, can therefore reveal more than just one approach alone, although more sophisticated curve fitting might be even more informative: other studies have found non-linear associations between relational tasks and age (Dumontheil et al., 2010; Magis-Weinberg, Blakemore, & Dumontheil, 2017), although these studies examined much larger age ranges.

Both relational reasoning measures predicted accuracy in science and maths generally, as anticipated based on previous research (Farrington-Flint et al., 2007; Jee et al., 2013; White et al., 1998) and the findings presented in **Chapter 6**. The CCF explained more of the variance in performance than analogical reasoning, in contrast to **Chapter 6**, where the non-verbal matrix reasoning task explained less variance than analogical reasoning. The CCF was selected for this study as a more comprehensive measure of relational reasoning, with four different subscales, so this finding suggests that the CCF is indeed capturing more than the WASI Matrix Reasoning measure. The inclusion of both relational reasoning measures in the models suggests that there is not just one type of relational reasoning ability important for counterintuitive reasoning, which fits with the idea that relational reasoning can be split into distinguishable categories (Alexander et al., 2016). It is particularly noteworthy that the relational reasoning measures remained strong predictors of misconception accuracy when controlling for problems that require no counterintuitive reasoning. This novel finding indicates that relational reasoning may have a particular role in counterintuitive reasoning, which was not found in **Chapter 6**, where the sample size was smaller. In addition, the correlations between reasoning and performance by trial type in **Chapter 6** indicated a slightly larger association with control trials than misconception trials. However, this difference was minimal, and as such, the results from this larger behavioural study are more convincing.

No relational reasoning measures predicted overall RT in science and maths (as in **Chapter 6**), but analogical reasoning predicted misconception RTs when controlling for control problem RTs (in contrast to **Chapter 6**). The association was positive, such that better analogical reasoning predicted longer RTs (relative to control RTs). This is especially interesting because the correlations between analogical reasoning and science and maths RT were not significant, and yet in **Chapter 6** they were. This novel finding suggests that analogical reasoning may have specific importance for counterintuitive problem-solving, in addition to all types of science and maths reasoning. While fast RTs may traditionally be seen as indicators of better performance, this is further evidence in support of the conclusion in **Chapter 3** that longer RTs may in fact reflect better performance; in particular, taking more time to reason about the problem presented. Comparing relations between different sets may be of importance when considering new problems, and the ability to do this when the problem is counterintuitive may be especially important, requiring a slowing down of responses to apply the analogy without error. For instance, when thinking about objects with different weights falling to the ground, it may be necessary to apply a previously learnt instance of objects falling to the current problem, with different objects relating analogously to one another. This reasoning would need to overcome the intuition that the heavier object falls faster. Overall, this finding suggests that encouraging students to slow down when answering problems in science and maths may help them to overcome misconceptions. Those with better analogical reasoning may recognise at some level that taking greater care in responding is helpful.

Overall, the study found little evidence for a role of inhibitory control in science and maths reasoning. However, both verbal analogical and non-verbal relational reasoning were predictors of science and maths performance generally, and in the context of misconceptions, when controlling for age, vocabulary, and working memory.

Having presented the results of the experimental chapters, the thesis will move on to the overall discussion. **Chapter 8** will situate the findings from **Chapters 3 to 7** in the wider literature, and consider the limitations of the studies presented throughout the thesis. It will then evaluate the educational neuroscience approach taken, and finally consider how the work could move forward in future studies, and the possible educational implications.

# Chapter 8 Discussion

Science and maths reasoning are essential skills that children and adolescents develop through the school years. The ability to reason scientifically and mathematically remains important throughout life, allowing individuals to make informed decisions and contribute to society (The Royal Society, 2014). An understanding of the underlying skills that enable success in these disciplines may lead to improvements in teaching strategies, learning practices, and academic outcomes. This PhD aimed to take an educational neuroscience approach to investigate the cognitive and neural bases of science and maths reasoning in adolescence.

In **Chapter 3**, I showed that both response and semantic inhibition were associated with counterintuitive reasoning in a behavioural study. In **Chapter 4**, I presented the results of two classroom studies that were designed with the help of teachers, and showed that inhibitory control did not associate with misconception presence before or after a lesson on the counterintuitive concept. In **Chapter 5**, I showed that brain activations associated with both response and semantic inhibition overlapped with brain activations observed when adolescents reasoned about science and maths misconceptions. In **Chapter 6**, I showed that verbal analogical reasoning predicted unique variance in science performance and neural activation in maths, and non-verbal matrix reasoning associated with neural activation in science. These findings were for overall science and maths performance and brain activation, with no specific links to misconception performance or brain activation. Finally, in **Chapter 7** I showed that verbal analogical reasoning and non-verbal relational reasoning were related to both general science and maths performance but also specifically to counterintuitive reasoning in a larger behavioural study. Overall, the results of the thesis show that inhibitory control (albeit weakly) and relational reasoning are two skills associated with success in school-related science and maths performance.

In this discussion section I will first provide a summary and discussion of the findings presented across the chapters. I will then provide a reflection on the educational neuroscience approach adopted throughout the PhD. I will then consider how future research might take these findings forward. Next I will consider the educational implications of the findings and finally, I will provide an overall conclusion, highlighting the novel contributions of this thesis.

## 8.1 Summary and discussion of findings

### 8.1.1 Inhibitory control and science and maths

Inhibitory control is thought to have a specific role in counterintuitive science and maths reasoning, in enabling the suppression of incorrect, intuitive responses (Mareschal, 2016). Intuitions may arise because of the way we experience the world. For instance, we see the Sun rise, move in the sky, and set, and it is therefore counterintuitive that the Earth goes around the Sun. Alternatively, perceptual cues or previously taught knowledge might interfere with our in-the-moment reasoning. Behavioural evidence has linked individual differences in inhibitory control to performance in both science (Zaitchik et al., 2014) and maths (Cragg, Keeble, Richardson, Roome, & Gilmore, 2017) in children, and neuroimaging evidence from adults showed inhibitory control brain regions to be involved in counterintuitive reasoning (Brault Foisy et al., 2015; Masson et al., 2014).

**Chapters 3 and 5** sought to extend the previous literature in three ways: first, by examining the unique contributions of response and semantic inhibition to counterintuitive reasoning; second, by using a broad range of problems related to the school curriculum, rather than using a repeated small set of problems; third, by comparing the brain regions recruited during inhibitory control to those specific to counterintuitive reasoning in adolescence. The studies had greater ecological validity than those reported previously, as they were the first to use problems that are related to the school curriculum. This is essential if links are to be made to learning and reasoning in real life. **Chapter 3** reported a behavioural study with 90 participants, and was the first study to use a novel science and maths misconceptions task designed to broadly cover misconceptions relevant to the Key Stage 3 curriculum for England. The results showed that the task worked as anticipated, with higher accuracy and shorter RTs in control trials compared to misconception trials. This indicated that the misconception trials, requiring counterintuitive reasoning, were indeed harder than control trials, even though the topic areas were matched, and in all cases these were problems covered in the school curriculum. The difference in performance between trial types was present even in the oldest participants, who had been taught all of these concepts in school. The key finding was that those with better inhibitory control were more accurate and took longer to respond on the misconception problems, when controlling for performance on the control problems, age, and general cognitive ability.

Specifically, better semantic inhibition, as measured by lower RT cost in the numerical Stroop task, was associated with higher accuracy on misconception trials. Response inhibition as measured by No-Go accuracy in the complex Go/No-Go task was associated with longer RTs on misconception trials. These novel findings indicated that both types of inhibitory control have a unique role in counterintuitive science and maths reasoning. Semantic inhibition may enable the suppression of intuitive theories, whereas response inhibition may allow for more time to reason about the problem.

**Chapter 5** reported a neuroimaging study of 34 participants, who completed adapted versions of the same tasks as in **Chapter 3**, while inside the fMRI scanner. The regression analyses on the behavioural data did not show the anticipated associations between inhibitory control and misconception performance, and therefore did not replicate the findings of **Chapter 3**, which is likely due to the reduced power in the neuroimaging study, since practical constraints led to a smaller sample size. **Chapter 3** used simple and complex No-Go accuracy as the key measures of response inhibition, while **Chapter 5** used simple and complex Go RT cost. This change was made because the Go/No-Go was adapted for use in the scanner through an increase in the number of No-Go trials, which led to higher accuracy in No-Go trials. Decreasing the frequency of the dominant response is thought to reduce how difficult it is to inhibit this response, but this change reduces the oddball effect, whereby neural responses to No-Go trials are different because of their lower frequency as opposed to their inhibition requirements (Simmonds, Pekar, & Mostofsky, 2008). The Go RT cost measures therefore aimed to capture response inhibition through the impact on changing RTs of the introduction of No-Go trials compared to blocks where there were only Go trials. The lack of association between response inhibition and misconception performance may therefore be because these are less sensitive measures of response inhibition, as the increased proportion of No-Go trials makes a non-response easier. Although not significant, the correlations (reported in **Chapter 6**) tended to show that those with a higher RT cost in both the simple and complex Go/No-Go were more accurate and slower in the science and maths misconception task. While a large RT cost is typically considered to reflect poorer performance, this would support the idea that slowing down may in fact indicate taking more time to answer the problem, and thus reflect better performance.

The semantic inhibition measures in **Chapter 5** were the same as in **Chapter 3**: Stroop RT cost and Stroop accuracy cost. However, the Stroop task now included blocks of congruent trials only, which again may have changed the pattern of performance, perhaps making it harder to focus on the number of items. Stroop accuracy cost was not a significant measure in **Chapter 3**, but Stroop RT cost was. In **Chapter 5**, the correlations between Stroop RT cost and performance were not significant, and the direction of association was such that a higher RT cost was associated with lower accuracy in misconception trials. This would fit with the findings in **Chapter 3**, but appears to contradict the argument above that a large RT cost reflects more careful and better performance. This may highlight the unique roles of semantic and response inhibition, which possibly behave differently in science and maths reasoning. It is possible that being slowed down in a response inhibition task by No-Go trials is linked to the ability to spend more time problem-solving before giving a response. Conversely, being slowed down in a semantic inhibition task by incongruent trials may be linked to being unable to suppress an incorrect theory, regardless of the time spent on the problem. Further research is needed to tease apart these possible associations.

The neuroimaging results in **Chapter 5** showed parietal (BA 40) and prefrontal (BA 8, 45) regions with increased BOLD signal specifically in response to misconception problems (compared to control problems). These regions showed overlap with those involved in both response and semantic inhibition, indicating that inhibitory control may indeed be a mechanism supporting performance. Given the potential issues with the behavioural measures, as described above (and the fact that these issues arose from optimising the tasks for the scanner), these neuroimaging results implicate inhibitory control more strongly. However, there was also activation unique to counterintuitive reasoning, indicating that other cognitive mechanisms are also likely at play. In addition, the investigations into individual differences in BOLD signal did not show any significant associations. There were no links between misconception-specific BOLD signal and activation in the inhibitory control tasks, nor between misconception-specific BOLD signal and behavioural measures of science and maths or inhibitory control. While the overlap is more convincing than previous studies that did not measure inhibitory control activation, the lack of correlations may be evidence that despite the overlap, these regions do not reflect the same processes. In addition, fMRI does not provide the granularity required to measure activity of specific neurons, and it is therefore possible that the overlapping regions may in fact not show overlapping neurons at all, with different neurons within a region responding to inhibitory control compared to those responding to misconceptions. It is also possible that some of the overlap between the Stroop and misconception BOLD signal is due to the numerical nature of the Stroop task. Nonetheless, congruent trials of the Stroop also present numbers, and it was therefore assumed that any number-specific processing arising from the Stroop would disappear through the subtraction of Congruent blocks.

Finally, with regards to the fMRI data, it is possible that the regions that showed increased BOLD signal across tasks reflect engagement of the MD network. The MD system refers to common recruitment of certain brain areas in response to many different types of cognitive challenge (Duncan, 2010). The system extends over regions of the prefrontal and parietal cortex, and incorporates the IPS, inferior frontal sulcus (IFS), anterior insula and frontal operculum (AI/FO), rostral PFC, pre-SMA and ACC (Duncan, 2010). From the regions that showed overlapping recruitment of misconception reasoning and inhibition, two were similarly located to those typically observed in the MD system: the IPS and the IFS. These regions showed activation across all three contrasts of interest (misconception, response inhibition, semantic inhibition), which further suggests that this increased BOLD signal may reflect the MD network and cognitive challenge, rather than showing inhibitory control processes. The regions that were specific to overlap in just one inhibitory control task did not match any areas of the MD system, which may indicate that these are more specific to response and semantic inhibition, while the IPS and IFS could be MD-related. Nonetheless, this argument rests on the assumption that misconception problems are more challenging than control problems, and this is only true where there is a conflict within a participant, and this is unlikely to be the case for all trials in all participants. For example, a participant might get a misconception problem wrong because they have no awareness of the correct concept, having not been taught it in school yet. For that participant, the problem is no more difficult that the accompanying control problem. Overall, it seems unlikely that the overlap is due to the MD network, and if the MD network is involved, this does not explain all of the overlap observed.

**Chapter 4** reported a study of 48 participants who were taught two new counterintuitive physics concepts by their class teacher, and a study of 69 participants who were taught a new counterintuitive maths concept by their class teacher, in a school where performance is above the national average. Inhibitory control did not associate with misconception performance at any time point, even when considering only those who held the misconception at pre-test. There are a number of possible reasons for this finding. The administration of the inhibitory control tasks as a whole class may have impacted performance. Participants were often sitting next to their friends, and were in some cases distracted, particularly the younger participants in the physics study who were a year younger than those in the maths study. Overall, it was a very different testing environment to the one-to-one testing session in **Chapter 3**. There sample size in both studies was based on the participants available in school rather than a power analysis, and this was compounded by the fact that many participants missed lessons due to extracurricular activities. It is also possible that this lack of association arose due to the increased salience of the newly learnt topics (since participants were aware they were part of a research project), particularly as ceiling performance was observed for one of the concepts following the lesson. There were just three misconceptions examined across these studies, and had a broader range of misconceptions been investigated, there may have been increased need for inhibitory control to answer the problems. In addition, the science study only looked at physics. Nonetheless, the results did align with other research showing no role for inhibitory control in the learning of new science concepts, albeit outside the context of misconceptions (Rhodes et al., 2014, 2016).

**Chapter 7** brought together inhibitory control and relational reasoning in science and maths in a behavioural study of 120 participants. This study did not replicate the findings from **Chapter 3** that those with better inhibitory control also performed better on the science and maths misconceptions task. Complex No-Go accuracy was used as the measure of response inhibition, and contrary to **Chapter 3**, No-Go accuracy was negatively correlated with RT in the science and maths misconceptions task (although not reaching significance). This finding may have been due to the changes made to the inhibitory control tasks, since they were shortened for **Chapter 7**, which may have led to a different pattern of performance, for example, requiring less attention in a shorter task. The science and maths misconception task was also reduced by a time limit. This would have led to a reduction in average science and maths RTs, and a smaller range of values, which may have reduced associations with this measure. Verbal semantic inhibitory control, as measured by the Hayling sentence completion test, was related to faster RTs and higher accuracy in science and maths, and not specifically in the context of misconceptions. This task was chosen as a measure of inhibitory control that might more closely reflect the type of inhibition that is required during counterintuitive reasoning, but the results did not support this hypothesis. The task required rapid responses to complete the sentence, and administration experience indicated that some participants may have used a strategy to aid performance that meant that the task did not necessarily reflect inhibition: it was observed that participants sometimes seemed to respond without listening to the sentence, having already decided upon the word they would say to complete the sentence. The observation that strategy-use played a role corroborates the explanation from Cipolotti and colleagues (2016) that the Hayling requires strategy implementation. This is in contrast to the Stroop, where there are a set number of responses, and the information which must be suppressed is visually available (Cipolotti et al., 2016). This difference likely led to the different associations between the two measures of semantic inhibition and science and maths.

One key possible reason for the inconsistencies in the results presented throughout the thesis is that changes were made to the tasks across studies. In addition to the changes already discussed above, participants received practice in the inhibitory control tasks before performing them in the scanner in **Chapter 5**. Although practices were given in all chapters, the practice was more comprehensive in **Chapter 5**, to ensure that participants knew what to expect inside the scanner. There was also a gap between the practice and performance of the task in **Chapter 5**, which may have allowed more time for consolidation of the task requirements, possibly with mental rehearsal. Overall, this would suggest that the inhibitory control tasks were easier in **Chapter 5** due to those amendments. In addition, the scanner environment may have led participants to perform the task in a different way, since it was unusual. The science and maths misconceptions task had no time limit in **Chapter 3**, which may have led to a different pattern of performance, although participants were told to answer as quickly as possible. In all other chapters using this task, a 12 s time limit was given, with a red border prompt appearing at 9 s, which some participants may have found somewhat stressful, and thus they may have responded differently. In **Chapter 3**, there were only two response options in the science and maths misconceptions task, while in the other chapters there were four options so that participants could indicate their certainty. This aspect was introduced since the 50% accuracy in misconception trials observed in **Chapter 3** may have corresponded to guesses. The presence of four different responses would likely have changed the nature of the task, since participants had to come up with their answer and consider how certain they were. An initial investigation into the use of the response keys proved uninformative, as some participants always chose the definitely option, with others always choosing the probably option, with no link between certainty and performance. However, this was a cursory look and in order to better understand the responses, this should be investigated more formally in the future. Finally, the science and maths problems changed to different degrees between studies, and these tweaks may have inadvertently changed the extent to which inhibitory control was necessary to answer the problems. However, this was only relevant to a small sample of the problems.

It is possible that the results are due to the specific tasks used throughout the course of the studies. The science and maths task was a new task that had not undergone extensive validation, and as such may not be the best way of capturing misconceptions in adolescents. While the task was adapted over time and tweaked in response to teacher feedback, it nonetheless does not match the kind of problems that adolescents face in the classroom or in exams. The true/false response may have encouraged a more impulsive response, whereas a problem that requires the writing down of a response may have elicited behaviour that better reflects typical classroom reasoning. In addition it is possible that the inhibitory control tasks did not accurately capture the type of inhibition that is required for overcoming misconceptions. While the Hayling sentence completion task attempted to better reflect the inhibition required in the classroom, the possible use of effective strategies that may require little inhibition, may explain the lack of effects.

Finally, it is possible that the results are due in part to the socioeconomic backgrounds of the participants. While attempts were made across the studies to recruit participants from both state and independent schools, it is likely that those who volunteered to take part had a particular interest in science and maths, which might not reflect the adolescent population. The independent school samples likely have a higher socioeconomic background than state school samples, which may be one explanation for the minimal associations in **Chapters 4, 5 and 7,** as compared to **Chapter 3**, where all participants were from a state school. The findings across studies may therefore not be generalizable to the wider adolescent population, where performance may be lower and other factors like home environment may be particularly important.

Overall, the results of these studies provide some further support for the theory that inhibitory control enables the suppression of intuitive answers during reasoning about misconceptions (Dunbar et al., 2007; Mareschal, 2016), and that poor performance in these disciplines may be due to poor inhibition as opposed to poor knowledge or understanding (Houdé, 2000). This is in contrast to older views of learning which refer to the replacement, reorganisation, or restructuring of previously held knowledge (Posner et al., 1982; Vosniadou, 2007). The studies reported here extend the behavioural research by investigating misconceptions in particular, where previous studies have linked inhibitory control to general performance in science and maths. The studies also extend the neuroimaging findings to a broad range of educationally-relevant stimuli, and explicitly compare brain activation during misconception reasoning and inhibitory control, going beyond the reverse inference used in previous investigations.

### 8.1.2 Relational reasoning and science and maths

Prior research linking relational reasoning and science have typically used analogies as a tool to teach new science concepts. Studies have found that teaching by analogy can be effective compared to other modes of teaching (Jee et al., 2013; Matlen et al., 2009). Other work has shown that expert scientists use relational reasoning skills when reasoning about scientific concepts (Trickett et al., 2009). In maths, individual differences in relational reasoning ability have been shown to associate with better maths performance (Farrington-Flint et al., 2007; White et al., 1998). In addition, relational reasoning ability has been linked to executive functions (Grossnickle et al., 2016; Richland & Burchinal, 2013). It was therefore considered that executive functions should be controlled for in investigating the role of relational reasoning in science and maths.

**Chapters 6 and 7** sought to extend the literature in three ways: first, by examining the behavioural associations between relational reasoning and science and maths while controlling for shared associations with executive functions; second, by specifically linking relational reasoning to counterintuitive reasoning in science and maths; and third, by relating individual differences in brain activation during science and maths to relational reasoning performance. **Chapter 6** reported a neuroimaging study, analysing data from the same 34 participants in **Chapter 5**. The correlational data showed that better verbal analogical reasoning was associated with better science and maths performance (higher accuracy and faster RTs), and that better non-verbal matrix reasoning was associated with faster science RTs. When controlling for vocabulary and working memory, analogical reasoning remained a significant predictor of science performance. The lack of association with non-verbal matrix reasoning was attributed to the possibility that the WASI Matrix Reasoning subtest is not a very comprehensive measure of non-verbal relational reasoning. Experience administering the test suggested that there are some problems which most participants get right, some which most participants get wrong, and a few in the middle that lead to the bulk of variation in responses. It was this observation that led to the inclusion of a different test of relational reasoning in **Chapter 7** that was considered to be more comprehensive through its inclusion of different subscales, and more sensitive due to the time limit.

The brain analyses in **Chapter 6** showed activation during science reasoning in the parahippocampal gyrus (BA 37, 18, 27), paracentral lobule (BA 4, 5, 23), and cerebellum (BA 18), to correlate with non-verbal matrix reasoning. In maths, activation in the temporal lobe (BA 21, 20, 38) correlated with analogical reasoning performance. In both cases, these associations held when other significant predictors of brain activation in those regions were controlled for. None of these regions are those typically observed in the MD system (Duncan, 2010), so the increased activation of these regions cannot simply be attributed to general increased cognitive challenge. This was the first study to link relational reasoning to science and maths reasoning brain and behaviour when controlling for the effects of age, vocabulary, and executive function. **Chapter 6** also indicated a possible role for analogical reasoning in science misconception performance, although this was not supported by the neuroimaging data, which showed no links between misconception activation and relational reasoning individual differences. This is further support for the argument that misconception activation does not reflect the MD system, since MD networks tend to map well onto activation during relational reasoning tasks (Duncan, 2010).

**Chapter 7** sought to extend these findings, and reported a behavioural study of 120 participants, using a more comprehensive measure of non-verbal relational reasoning alongside the same verbal analogical reasoning task (which was altered slightly). The CCF relational reasoning test was chosen in place of the WASI Matrix Reasoning subtest used in **Chapter 6**. One advantage of this test is that it has a time limit for each subscale, and was thus suitable for use in schools with strict time requirements for testing. Experience observing participants complete this test indicated that there was greater variation in terms of which items were answered correctly, and as such it does seem to be a more comprehensive measure. Indeed the CCF showed greater links to science and maths performance than in **Chapter 6**, indicating increased sensitivity of this scale to individual differences, although this could be related to the larger sample size. A disadvantage of using the CCF was that this scale does not correspond to the four categories of relational reasoning proposed by Alexander and colleagues (2016). The TORR, which does tap all four categories (see section 1.4), was initially considered as the measure to be used in **Chapter 7**, but the lack of time limit meant participants may not have had time to finish within the allocated testing session time. There is no clear link between the scales in the CCF and the TORR, and it appears that each scale in the CCF incorporates aspects of different types of relational reasoning measured in the TORR.

Both verbal analogical reasoning and non-verbal relational reasoning were associated with higher accuracy in both science and maths when controlling for age, vocabulary, and working memory in **Chapter 7**. In addition, both measures were predictors of misconception accuracy when controlling for performance in control problems. This finding is particularly noteworthy because it suggests that relational reasoning has particular importance for counterintuitive reasoning, over and above vocabulary and executive functions. A final interesting finding was that analogical reasoning predicted misconception RTs, such that better analogical reasoning predicted slower RTs. This suggests that participants with better analogical reasoning took more time to answer the problems. This finding relates to the finding in **Chapter 3 and 5** that better response inhibition was associated with slower RTs, and as such is further support for the idea that slowing down during the resolution of counterintuitive problems may be a marker of better performance, rather than worse performance.

Since Go/No-Go and Stroop measures were not significant predictors of performance in **Chapters 6 and 7**, the interplay between inhibitory control, relational reasoning, and science and maths reasoning was hard to examine. In the reduced sample of participants who had data in VWM and the Hayling sentence completion test (measuring verbal semantic inhibition), the Hayling score was a significant predictor of science and maths accuracy. But when non-verbal reasoning was selected by the model, the Hayling score was no longer a significant predictor. This indicates that the CCF was able to explain the variance in science and maths accuracy better than the Hayling, and that relational reasoning ability may be contributing to performance in both. There were no other models that selected both the Hayling and relational reasoning measures, although the Hayling score did predict RT in overall science and maths, where no relational reasoning measures were significant. The latter finding may indicate that those with better verbal semantic inhibition were better able to respond quickly due to the verbal nature of the task. However, it is also possible that this association merely reflects the time requirements of the Hayling sentence completion test. The time taken to give an answer in the Hayling is a key part of the overall score used, and as such it may simply be that those who are faster on the Hayling (and thus get a better score) are also faster on the science and maths task.

Overall, these studies show that both verbal analogical and non-verbal relational reasoning are associated with performance in science and maths, even when controlling for age, vocabulary, and executive function. They extend the previous findings in linking individual differences in relational reasoning to science and maths in terms of both brain and behaviour, and in particular to misconceptions.

## 8.2 Reflection on educational neuroscience approach

Throughout the PhD I aimed to take an educational neuroscience approach, with five particular dimensions. I will first briefly consider the extent to which I was able to integrate each planned dimension into the PhD, and then consider further the debate around educational neuroscience, reflecting on my own experiences.

a. *Regularly meet with teachers to continually assess the educational relevance of the studies and findings.*
   Throughout the course of my PhD I was partnered with a school where members of staff were interested in the research I was addressing. I met with school staff on a regular basis, and explained my research to them. I had informal talks with teachers across disciplines, who often asked me how my research related to their subject, so this provided opportunities for me to link my findings to secondary school education.
b. *Spend a substantial amount of time in school to ensure a rounded understanding of the typical science and maths classroom, and the school context more widely.*
   I observed a total of 66 science and maths lessons at my partner school, which gave me a good understanding of typical lessons in these disciplines, which included taught material, group work, practical sessions, and independent work. Observing a wide range of teachers also highlighted the degree of freedom that teachers have in how they teach a lesson. This was useful in considering the kind of recommendations that might be helpful to teachers.

c. *Use stimuli that are relevant to science and maths content that 11- to 15-year-olds frequently encounter and need to understand for exam success.*
   I met with groups of science and maths teachers and showed them the stimuli for the main science and maths misconceptions task. Changes were made to the stimuli on the basis of their feedback, to ensure the problems were relevant to the content students had to know. In addition, the classroom studies in **Chapter 4** were designed with help from the teachers, who chose topics that were relevant for the year groups being tested.
d. *Adopt a suite of methods to address reasoning and learning at different levels of scientific understanding.*
   The PhD combines behavioural, fMRI and classroom-based studies. While a training study was initially planned for the PhD, this did not come to fruition as the interested school eventually dropped out. A training study would have helped to determine causal links between inhibitory control and counterintuitive reasoning, and as such may have enhanced the conclusions drawn from the PhD.
e. *Feed results back to teachers and pupils who took part in the research to ensure that their participation saw rewards and that the research had impact.*
   Results were given to teachers in the form of three presentations across two schools. In addition, I gave two talks to pupils who had taken part in the research, and have two more talks planned for next academic year. In all cases, I present the theory behind the research, the methods used, the results, and the possible educational implications. I also provide the opportunity for teachers and pupils to think about how they might change their teaching or learning practices in light of what I have presented. Teachers in particular have told me that they have found the process interesting and useful, and that the research chimes with their own experiences in the classroom.

As described in **Chapter 1** section 1.1, there is debate concerning the extent to which educational neuroscience can impact positively on education. According to some critics, neuroscience cannot inform education, as the distance is too great (Bishop, 2014; Bowers, 2016). Others have argued that integrating psychological and neural levels of explanation can lead to a greater understanding of the processes involved education, and have pointed to examples of progress in the field (Howard-Jones et al., 2016). My experience in attempting a project across different levels of description has led me to support the view that neuroscience and education can inform one another. While my research has not led to specific recommendations for teaching and learning, there are nonetheless some promising avenues for future research which may eventually lead to more solid recommendations for education. Critics have pointed to the lack of useful educational neuroscience research so far (Bowers, 2016), but the field is young, and this thesis has shown that it is not simply the case that one study can seamlessly integrate neuroscience and education, resulting in clear educational recommendations. Each area of research is an ongoing endeavour, and will require many different studies, with theories changing over time. The collaboration with teachers in the work presented in this thesis has shown that research can be bolstered by taking into consideration the expertise of teachers. Stimuli were altered based on teacher recommendations, and this granted the research greater ecological validity.

It has also been argued that simply finding out the neural underpinnings of an education-related cognitive process cannot reveal anything useful for education (Bishop, 2014). The neuroimaging findings presented in this thesis have shown that this argument does not accurately characterise the nature of all neuroimaging research in this field. Rather than simply documenting the brain regions involved in misconceptions, the aim here was to establish the overlap with inhibitory control, in order to better understand the cognitive mechanisms. In addition, the research was concerned with individual differences, in reporting how individual differences in recruitment of certain brain regions related to performance in relational reasoning tasks. This type of research does not directly lead to educational recommendations, but can inform later study design, and potentially highlight cognitive mechanisms that may lead to educational recommendations in the future.

Finally, working closely with teachers gave them the opportunity to shape the research in a small way according to their interests, and even enabled them to consider new approaches for the classroom. While the teachers are aware that the theories under examination in this project are being tested and thus not incontrovertible, teachers have told me that they have started to try out different techniques that draw on this research (for examples, see section 8.4). With the rise of teacher-led research (Dommett, Devonshire, & Churches, 2018), these kinds of collaborations between researchers and teachers may give teachers ideas that can then be tested more formally in their classrooms, contributing further to the scientific dialogue, without the need for external researchers to dictate the programme of investigation.

Overall, I found the experience of taking an educational neuroscience approach to be a positive one. I believe that a commitment to educational relevance has made the conclusions drawn more useful for education. Conducting different types of studies, and considering more than one level of explanation has led to a better understanding of the possible associations between different skills. There is no doubt that there is still plenty of work to be done in order to fully address the cognitive and neural bases of science and maths reasoning in adolescence. But this approach has paved the way for further research, and in the future it is hoped that there will be firmer educational recommendations in this area. This can only be achieved if teachers are involved as expert collaborators, and the research focusses on educational recommendations as a specific aim throughout, as opposed to simply an afterthought.

## 8.3 Future research

Of particular interest for future research is the extent to which cognitive training might lead to improved performance in science and maths. Attempts at training inhibitory control have proven successful. For example training 30 adults on the stop signal task over 10 sessions led to improved performance alongside associated neural changes in the IFG, compared to 30 adults who did not receive training (Berkman, Kahn, & Merchant, 2014). A study of 19 15- to 17-year-olds found improvement on the stop signal task, but with no transfer to a Go/No-Go task or risky behaviour (which was considered to reflect poor inhibitory control in the real world) as measured by a questionnaire (Beauchamp, Fisher, & Berkman, 2018). These examples follow a line of research aiming to improve executive functions, which have typically shown improvement in the task being practiced, and near but not far transfer (Melby-Lervåg & Hulme, 2013) despite early anticipation that academic-related skills such as science would improve in tandem (Klahr, Zimmerman, & Jirout, 2011). The proposed training would therefore not address inhibitory control or relational reasoning in isolation. Rather, training would occur within the context of science and maths, encouraging inhibitory control and relational reasoning within science and maths problems that require counterintuitive reasoning. Such an approach would reveal more about the educational relevance of the findings, and further allow for testing of possible causal associations.

To this end, I have been involved in the design and testing of an inhibitory control training programme in the context of primary school. The UnLocke programme (www.unlocke.org) aims to encourage inhibitory control during science and maths counterintuitive reasoning, through requiring pupils to 'stop and think' before giving their answer. This differs from the more traditional cognitive training approach which would have pupils practice inhibitory control in a separate programme, and examine any transfer to school-related performance. In addition, in the UnLocke programme, pupils are exposed to three types of reasoning about the counterintuitive concept: the correct reasoning, the incorrect reasoning that exemplifies the misconception, and different incorrect reasoning that is unrelated to the misconception. It is hoped that this approach will highlight to students both how to reason correctly, and the fact that incorrect reasoning can occur. The final data are still being collected for this study, and it relates to primary school rather than secondary school, but future research could test a similar approach in adolescents. Caution should be taken though, as it is possible that encouraging students to inhibit their initial responses may not always be beneficial; sometimes the first answer might be right, and students might change their answer erroneously.

The training of relational reasoning within the context of science and maths would also be of interest. As described in the Introduction (see section 1.4.3.1), there are already examples of analogical reasoning as a learning tool in science, which have shown success (Jee et al., 2013; Matlen et al., 2009). However, these examples have not focussed on adolescents, nor have they attempted to improve analogical reasoning skills in science. It is unclear from these studies how individual differences in analogical reasoning relate to the ability to learn through analogies, and whether learning in this manner improves analogical reasoning in other areas of science. Therefore, a key focus for future research could be the extent to which relational reasoning within the domain of science and maths can improve science and maths performance. This might involve explicitly highlighting relations between taught material, or altering the presentation of content such that relational reasoning requirements are increased. There have been attempts at relational reasoning training outside of the context of discipline-learning. One study found that 11- to 33-year-olds improved their relational reasoning skills following 20 days of online training, but that this did not transfer to other tested tasks, which included VWM and face perception (Knoll et al., 2016). It is interesting to note that while all age groups improved with training, the biggest improvement in relational reasoning was in late-adolescence (age 16 to 18 years). This suggests that this period of adolescence may be a good target for future interventions, and would coincide with final school examinations.

Other studies have similarly shown that performance on relational reasoning tests can be trained. One small-scale study of 28 7- to 10-year-olds examined the effect of eight weeks of training on different reasoning games (Mackey, Hill, Stone, & Bunge, 2011). The training led to improvement on the trained games, as well improvement on an untrained test of non-verbal matrix reasoning. One study examined 50 adults, half of whom received training in the Law School Admission Test (LSAT) in the USA, which heavily relies on reasoning, and half of whom did not (Mackey, Whitaker, & Bunge, 2012). Training was associated with higher performance on the LSAT, with accompanying decreased diffusivity in white matter connecting the frontal cortices to each other. A similar fMRI study of 51 adults found that training was associated with higher LSAT scores, and strengthened fronto-parietal functional connections (Mackey, Miller Singley, & Bunge, 2013). These neuroimaging studies provide convincing evidence that relational reasoning can be trained, with associated brain changes. Nonetheless, it is still unclear whether or not this transfers to improved aptitude with regards to school- and university-related outcomes.

These findings are particularly interesting when considered in the context of school entrance exams that use relational reasoning tests. The 11 Plus is an exam in the UK that grammar schools and some independent schools require students to sit before entry. The 11 Plus contains a verbal reasoning test and a non-verbal reasoning test. If the skills required to perform well on these tests improve through practice, it seems that they do not measure an underlying fixed ability, which calls into question their claim to indicate natural academic ability. Similarly, tests in the UK (such as the Law National Aptitude Test, LNAT) and USA (LSAT) for entry to university or college for some courses rely on relational reasoning. Future research could therefore investigate the extent to which performance on these entrance exams relates to course performance, taking into account the extent of training that was undertaken, and the type of training. Given the evidence presented above, it is unclear whether or not training would transfer to course performance, which would have important implications for the institutions that rely on these tests.

Beyond training studies, there is also further work to be done in establishing links between individual differences in different types of skill. The work presented here could be extended to examine more specific links between types of misconceptions and types of inhibition or reasoning. For instance, it is possible that different types of inhibitory control specifically allow the suppression of misconceptions of different origins. Similarly, it may be that different types of relational reasoning specifically enable reasoning about certain types of counterintuitive concepts. The studies presented here did not categorise types of misconception, and contained a mixture of those due to misleading perceptual cues, previously held beliefs, and prior experiences with the world (and it may prove difficult to tease apart the origins of misconceptions). Nonetheless, I have suggested that semantic inhibition may be more important for suppressing intuitive theories, and that response inhibition may be more important for allowing time to reason about the problem. These hypotheses could be tested further by grouping counterintuitive concepts according to their proposed origins, and assessing associations with semantic and response inhibition. An alternative might be to ask participants why they thought the misconception was true, and to group the problems according to each participant's reasoning.

One possibility that was not considered in the research in this thesis is that inhibitory control does not have a linear association with misconception performance. For instance, it may be that when answering a misconception problem when it is first encountered requires no inhibitory control, because the individual holds only the misconception, and does not hold the correct concept. Thus, there is nothing to inhibit. Later, when the correct, counterintuitive concept has been learnt, the individual must inhibit their intuitive response to get the answer right. Later still, the correct concept may be fully automatized, and so again, inhibitory control is not necessary (or is necessary to a lesser extent), since the correct response has been successfully assimilated. It is possible that this non-linear association led to the weak inhibitory control effects observed in the studies within this thesis. In order to address this question, future research could assess the presence of specific misconceptions in individuals, ideally over time, and examine the link between inhibitory control and misconception performance once the correct answer has recently been learnt. This approach would move the learning studies of **Chapter 4** into new territory, taking a more longitudinal approach and ideally with a wider range of participant performance to avoid ceiling effects, and a broader set of concepts to increase the ecological validity.

In addition, different types of science and maths ability could be examined more closely. For example, although not focussing on counterintuitive reasoning, previous research found an association between inhibition and procedural maths, but not between inhibition and conceptual or factual maths (Gilmore et al., 2015). Maths could be measured by these three components, with control problems and misconception problems within each component. Science could be measured according to the three core skills proposed by Tolmie and colleagues (2016): prediction, description, and explanation, again with control problems and misconception problems in each. As in the studies reported in this thesis, both response and semantic inhibition could be measured. Relational reasoning could be measured according to the four components posited by Alexander and colleagues (2016): analogical, anomalous, antinomous, and antithetical. While these components are thought to be best combined to get an overall measure of relational reasoning ability, investigating the possibility that they show different patterns of association to science and maths may lead to important insights. In addition, working memory (visuospatial and verbal) and vocabulary could be measured, to establish shared associations and mediating relations. A large-scale, comprehensive investigation of this kind would lead to more specific conclusions than presented here, possibly with associated specific recommendations for education.

## 8.4 Educational implications

In order to support the development of counterintuitive reasoning, the challenge for teachers lies in recognising misconceptions, remedying them, and all the while building on correct prior knowledge (Klahr et al., 2011). While the results presented in this thesis do not themselves lead to concrete recommendations for teaching and learning, they do highlight some possible approaches that could be evaluated in the future or tried by teachers in their classrooms. The first suggestion is to encourage students not to give their first answer to a problem, but instead to spend more time reasoning about the problem presented. This approach would be consistent with the findings that better response inhibition and better analogical reasoning were associated with longer RTs.

While quick answers are sometimes encouraged in lessons (particularly in maths, when fast mental calculations can be seen as important), taking time to ‘stop and think’ may be beneficial and lead to more time for considering the correct answer. This idea was suggested in the context of science education by Rowe (1986), alongside evidence from a range of studies that both comprehension and attitude improved through introducing ‘wait time’. Nonetheless, experience in school suggests that the wait time approach has not been widely adopted. This approach may have particular importance for individuals with deficits in inhibitory control, related to a developmental disorder such as ADHD. These students may require extra support in learning to suppress their first intuitive answer.

The second suggestion is to raise awareness of misconceptions within students, telling them explicitly about incorrect answers and why they may be appealing but wrong. It may be that this can help students to apply their cognitive skills more effectively. For instance, if a student knows that they hold incorrect intuitions in science, they might more readily suppress their initial response in a science lesson and take more time to reason and consider other answers. This suggestion is consistent with recommendations to use refutation texts, which explicitly address misconceptions and explain why they are wrong in detail (Ecker, Swire, & Lewandowsky, 2014). Here, the idea is to go further than simply saying the information is untrue (a plain retraction) and to include supporting evidence explaining why it is wrong, and if possible, where the misconception came from. As long as the refutation text is clear, well-constructed, and the students engage fully in the material (as opposed to skim reading) this is more effective than a plain retraction (Ecker et al., 2014).

Ryan and Williams (2007) also argued that simply correcting errors is unproductive, and suggested a ‘dialogic pedagogy’ approach in maths. The starting point for this approach is when there is some kind of disagreement between students about something mathematical, which is followed by the teacher encouraging argument in discussion. It is key that students are encouraged to listen to each other, and to be open to other points of view. Students first articulate their point of view, then ideally reformulate their ideas based on what others have said, then reflect on why their initial reasoning was incorrect, and finally reach a resolution. While this approach was proposed in the context of maths by Ryan and Williams (2007), the same strategy could be used in science too, when disagreements occur. Indeed, Driver and colleagues (2015) highlight the importance of peer discussions in science, showing that such conversations help students to clarify their own thoughts and to build on their collective understanding. Although not mentioned by the authors, as in maths, the presence of a teacher would help to ensure misconceptions are not propagated when discussions relate to counterintuitive concepts.

Stavy and Tirosh (2000) suggested the ‘conflict teaching’ approach to help students overcome misconceptions. In this approach, students are presented with a problem known to elicit an incorrect (intuitive) response, triggering a misconception, followed by a second problem that contradicts their first response. There is some evidence that this approach is effective and that performance following this approach improves (Stavy & Tirosh, 2000). No mechanism is posited for conflict teaching, but it seems likely that it would highlight students’ own tendencies to answer intuitively, thus raising awareness of the need to go beyond initial responses.

Arguments have also been made for supporting relational reasoning in education. Analogical reasoning has been shown to help students learn in science (Jee et al., 2013; Matlen et al., 2009), and extending this to maths learning has been suggested as it may be especially helpful in the learning of abstract concepts (Richland et al., 2007). It has also been shown that explanations of analogies in learning must be made explicit by teachers, as relations may not always be obvious to the learner (Vendetti et al., 2015). Younger learners would need support in applying analogies since relational reasoning continues to develop through the school years (Richland & Burchinal, 2013).

Overall, the literature and this thesis indicate that cognitive skills play an important role in reasoning and learning in science and maths. Teachers are not always aware of the role of these factors in successful school performance (Gilmore & Cragg, 2014), so highlighting their importance to educators and learners is likely a worthwhile first step.

## 8.5 Conclusion

The evidence presented in thesis suggests that inhibitory control and relational reasoning are two cognitive skills that enable adolescents to reason effectively in science and maths. In particular, these skills seem to have a specific role in reasoning about counterintuitive concepts, where misconceptions may be held. The studies presented here are novel in: assessing performance in a broad range of educationally-relevant science and maths problems; considering inhibitory control in the context of learning new counterintuitive concepts in the classroom; explicitly examining overlapping neural activation in inhibitory control and counterintuitive reasoning; associating individual differences in neural activation during science and maths with relational reasoning behavioural performance; and finally, in considering the role of both relational reasoning and inhibitory control when controlling for the effects of age, vocabulary, and working memory. Overall, the results from this thesis indicate that success in science and maths relies not only on content knowledge, but on inhibitory control and relational reasoning ability.

# References

Adleman, N. E., Menon, V., Blasey, C. M., White, C. D., Warsofsky, I. S., Glover, G. H., & Reiss, A. (2002). A developmental fMRI study of the Stroop color-word task. *NeuroImage, 16*(1), 61–75.

Alexander, P. A. (2016). Relational thinking and relational reasoning: Harnessing the power of patterning. *Npj Science of Learning*, *1*, 16004.

Alexander, P. A., Dumas, D., Grossnickle, E. M., List, A., & Firetto, C. M. (2016). Measuring relational reasoning. *The Journal of Experimental Education*, *84*(1), 119–151.

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental Psychology*, *50*(10), 2368–2379. http://doi.org/10.1037/a0037493

Alloway, T. P. (2007). *Automated working memory assessment*. London, UK: Pearson Assessment.

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences*, *113*(18), 4909–4917.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, *8*(2), 71–82.

Ansari, D., & Coch, D. (2006). Bridges over troubled waters: Education and cognitive neuroscience. *Trends in Cognitive Sciences*, *10*(4), 146–151.

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *The Neuroscientist*, *13*(3), 214–228.

Arsalidou, M., Pawliw-Levac, M., Sadeghi, M., & Pascual-Leone, J. (2018). Brain areas associated with numbers and calculations in children: Meta-analyses of fMRI studies. *Developmental Cognitive Neuroscience*, *30*(July 2017), 239–250. http://doi.org/10.1016/j.dcn.2017.08.002

Baker, S. T., Gjersoe, N. L., Sibielska-Woch, K., Leslie, A. M., & Hood, B. M. (2011). Inhibitory control interacts with core knowledge in toddlers' manual search for an occluded object. *Developmental Science*, *14*(2), 270–279.

Banich, M. T., & Depue, B. E. (2015). Recent advances in understanding neural systems that support inhibitory control. *Current Opinion in Behavioral Sciences*, *1*, 17–22.

Beauchamp, K. G., Fisher, P. A., & Berkman, E. (2018). Brief, computerized inhibitory control training to leverage adolescent neural plasticity: A pilot effectiveness trial. *Applied Neuropsychology: Child*.

Bedard, A.-C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental Neuropsychology*, *21*(1), 93–111.

Berkman, E. T., Kahn, L. E., & Merchant, J. S. (2014). Training-induced changes in inhibitory control network activity. *Journal of Neuroscience*, *34*(1), 149–157.

Bishop, D. V. M. (2014). What is educational neuroscience? Retrieved September 17, 2018, from http://deevybee.blogspot.com/2014/01/what-is-educational-neuroscience.html

Blair, C. (2016). Developmental science and executive function. *Current Directions in Psychological Science*, *25*(1), 3–7.

Blos, J., Chatterjee, A., Kircher, T., & Straube, B. (2012). Neural correlates of causality judgment in physical and social context - The reversed effects of space and time. *NeuroImage*, *63*(2), 882–893.

Borst, G., Poirel, N., Pineau, A., Cassotti, M., & Houdé, O. (2013). Inhibitory control efficiency in a Piaget-like class-inclusion task in school-age children and adults: A developmental negative priming study. *Developmental Psychology*, *49*(7), 1366–1374.

Bowers, J. S. (2016). The practical and principled problems with educational neuroscience. *Psychological Review*, *123*, 600–612.

Brault Foisy, L.-M., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education*, *4*, 26–36.

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

Bruer, J. T. (1997). Education and the brain: A bridge too far. *Educational Researcher*, *26*(8), 4–16.

Burgess, P. W., & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, *34*(4), 263–273.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton tests*. London, UK: Pearson Assessment.

Burnett, S., Bird, G., Moll, J., Frith, C., & Blakemore, S.-J. (2008). Development during adolescence of the neural processing of social emotion. *Journal of Cognitive Neuroscience, 21*(9), 1736–1750.

Butterworth, B., & Tolmie, A. (2014). Introduction. In D. Mareschal, B. Butterworth, & A. Tolmie (Eds.), *Educational Neuroscience* (1st ed., pp. 1–12). Chichester, UK: John Wiley & Sons, Ltd.

Butterworth, B., & Varma, S. (2013). Mathematical development. In D. Marescal, B. Butterworth, & A. Tolmie (Eds.), *Educational Neuroscience* (pp. 201–236). Chichester, UK: John Wiley & Sons, Ltd.

Carver, A. C., Livesey, D. J., & Charles, M. (2001). Age related changes in inhibitory control as measured by stop signal task performance. *International Journal Neuroscience*, *107*, 43–61.

Cattell, R. B., & Cattell, A. K. S. (1959). Handbook for the culture fair intelligence test. Illinois, USA: The Institute of Personality & Ability Testing.

Cauley, S. F., Polimeni, J. R., Bhat, H., Wang, D., Wald, L. L., & Setsompop, K. (2014). Inter-slice leakage artifact reduction technique for simultaneous multi-slice acquisitions. *Magnetic Resonance in Medicine*, *72*(1), 93–102.

Chee, M. W. L., Venkatraman, V., Westphal, C., & Siong, S. C. (2003). Comparison of block and event-related fMRI designs in evaluating the word-frequency effect. *Human Brain Mapping*, *18*(3), 186–193.

Cipolotti, L., Spanò, B., Healy, C., Tudor-Sfetea, C., Chan, E., White, M., … Bozzali, M. (2016). Inhibition processes are dissociable and lateralized in human prefrontal cortex. *Neuropsychologia*, *93*(March), 1–12.

Comalli, P. E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *The Journal of Genetic Psychology*, *100*, 47–53.

Corr, P. J. (2006). Neuroimaging. In *Understanding biological psychology*. Malden, USA: Blackwell Publishing Ltd.

Cragg, L. (2016). The development of stimulus and response interference control in midchildhood. *Developmental Psychology*, *52*(2), 1–18.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63–68.

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, *162*, 12–26.

Criaud, M., & Boulinguez, P. (2013). Have we been asking the right questions when assessing response inhibition in go/no-go tasks with fMRI? A meta-analysis and critical review. *Neuroscience and Biobehavioral Reviews*, *37*(1), 11–23.

Crone, E. A., Wendelken, C., Van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, *12*(1), 55–66.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078.

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.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, *12*, 45–75.

Department for Education. (2013a). *Mathematics programmes of study: Key stage 3*.

Department for Education. (2013b). *Science programmes of study: Key stage 3*.

Diamond, A. (2013). Executive functions. *The Annual Review of Psychology*, *64*, 135–168.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959–964.

Dick, F., Lloyd-Fox, S., Blasi, A., Elwell, C., & Mills, D. (2013). Neuroimaging methods. In D. Mareschal, B. Butterworth, & A. Tolmie (Eds.), *Educational neuroscience*. Chichester, UK: John Wiley & Sons, Ltd.

Dommett, E., Devonshire, I., & Churches, R. (2018). Bridging the gap between evidence and classroom "clinical practice": The potential of teacher-led randomised controlled trials to advance the science of learning. *Impact*, *2*, 64–67.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (2015). *Making sense of* *secondary science: Research into children's ideas* (Classic). New York: Routledge.

Dumontheil, I. (2014). Development of abstract thinking during childhood and adolescence: The role of rostrolateral prefrontal cortex. *Developmental Cognitive Neuroscience, 10*, 57–76.

Dumontheil, I., Houlton, R., Christoff, K., & Blakemore, S.-J. (2010). Development of relational reasoning during adolescence. *Developmental Science*, *13*(6), F15–F24.

Dumontheil, I., Wolf, L. K., & Blakemore, S.-J. (2016). Audience effects on the neural correlates of relational reasoning in adolescence. *Neuropsychologia*, *87*, 85–95.

Dunbar, K. N., Fugelsang, J. A., & Stein, C. (2007). Do naïve theories ever go away? Using brain and behavior to understand changes in concepts. In M. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 193–205). Mahwah, New Jersey, USA: Lawrence Erlbaum Associates.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179.

Durston, S., Thomas, K. M., Yang, Y. H., Ulug, a M., Zimmerman, R. D., Casey, B. J., … Casey, B. J. (2002). A neural basis for the development of inhibitory control. *Developmental Science*, *5*(4), F9–F16.

Ecker, U. K. H., Swire, B., & Lewandowsky, S. (2014). Correcting misinformation. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. Cambridge, MA: The MIT Press.

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *PNAS*, *113*(28), 7900–7905.

Farrington-Flint, L., Canobi, K. H., Wood, C., & Faulkner, D. (2007). The role of relational reasoning in children's addition concepts. *British Journal of Developmental Psychology*, *25*(2), 227–246.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Ferrer, E., O'Hare, E. D., & Bunge, S. A. (2009). Fluid reasoning and the developing brain. *Frontiers in Neuroscience*, *3*, 46–51.

Field, A. (2012). Linear models: Looking for bias. Retrieved August 10, 2016, from http://www.statisticshell.com/docs/linearmodelsbias.pdf

Flandin, G., & Friston, K. J. (2017). Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Human Brain Mapping*.

Friston, K. J. (2007). Statistical parametric mapping. In K. J. Friston, J. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statiscial parametric mapping: The analysis of functional brain images*. London, UK: Academic Press.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, *39*(1), 47–59. http://doi.org/10.2307/1130388

Gilmore, C., & Cragg, L. (2014). Teachers' understanding of the role of executive functions in mathematics learning. *Mind, Brain, and Education*, *8*(3), 132–136.

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *ZDM Mathematics Education*, *47*, 771–782.

Glady, Y., French, R. M., & Thibaut, J. P. (2017). Children's failure in analogical reasoning tasks: A problem of focus of attention and information integration? *Frontiers in Psychology*, *8*, 1–13.

Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate number sense, that counts. *Psychological Science*, *25*(3), 789–798.

Goldacre, B. (2013). *Building evidence into education*. London, UK.

Goswami, U. (1992). *Analogical reasoning in children*. Hove, UK: Lawrence Erlbaum Associates.

Green, C. T., Bunge, S. A., Briones Chiongbian, V., Barrow, M., & Ferrer, E. (2017). Fluid reasoning predicts future mathematical performance among children and adolescents. *Journal of Experimental Child Psychology*, *157*, 125–143.

Grossnickle, E. M., Dumas, D., Alexander, P. A., & Baggetta, P. (2016). Individual differences in the process of relational reasoning. *Learning and Instruction*, *42*, 141–159.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1264–1282.

Henriksson, L., Karvonen, J., Salminen-Vaparanta, N., Railo, H., & Vanni, S. (2012). Retinotopic maps, spatial tuning, and locations of human visual areas in surface coordinates characterized with multifocal and blocked fMRI designs. *PLoS ONE*, *7*(5), e36859.

Holland, S. K., Altaye, M., Robertson, S., Byars, A. W., Plante, E., & Szaflarski, J. P.

(2014). Data on the safety of repeated MRI in healthy children. *NeuroImage: Clinical*, *4*, 526–530.

Hood, B., Cole-Davies, V., & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. *Developmental Psychology*, *39*(1), 61–70.

Houdé, O. (2000). Inhibition and cognitive development: Object, number, categorization, and reasoning. *Cognitive Development*, *15*, 63–73.

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

Howard-Jones, P., Varma, S., Ansari, D., Butterworth, B., De Smedt, B., Goswami, U., … Thomas, M. S. C. (2016). The principles and practices of educational neuroscience: Comment on Bowers (2016). *Psychological Review*, *123*(5), 620–627.

Huizinga, M., Dolan, C. V, & van der Molen, M. W. (2006). Age-related change in executive function: developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–36.

Humphrey, G., & Dumontheil, I. (2016). Development of risk-taking, perspective-taking, and inhibitory control during adolescence. *Developmental Neuropsychology*, *5641*, 1–18.

Jablansky, S., Alexander, P. A., Dumas, D., & Compton, V. (2015). Developmental differences in relational reasoning among primary and secondary school students. *Journal of Educational Psychology*, *108*(4), 592–608.

Jaeger, A. (2013). Inhibitory control and the adolescent brain: A review of fMRI research. *Psychology and Neuroscience*, *6*(1), 23–30.

Jahn, G., Wendt, J., Lotze, M., Papenmeier, F., & Huff, M. (2012). Brain activation during spatial updating and attentive tracking of moving targets. *Brain and Cognition*, *78*(2), 105–113. http://doi.org/10.1016/j.bandc.2011.12.001

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., & Sageman, B. (2013). Finding faults: Analogical comparison supports spatial concept learning in geoscience. *Cognitive Processing*, *14*(2), 175–187.

Kail, R. (1993). Processing time decreases globally at an exponential rate during childhood and adolescence. *Journal of Experimental Child Psychology*.

Khng, K. H., & Lee, K. (2009). Inhibiting interference from prior knowledge: Arithmetic intrusions in algebra word problem solving. *Learning and Individual Differences*, *19*(2), 262–268.

Khng, K. H., & Lee, K. (2014). The relationship between Stroop and stop-signal measures of inhibition in adolescents: Influences from variations in context and measure estimation. *PLoS ONE*, *9*(7), e101356.

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science*, *333*(6045), 971–975.

Knoll, L. J., Fuhrmann, D., Sakhardande, A. L., Stamp, F., Speekenbrink, M., & Blakemore, S.-J. (2016). A Window of Opportunity for Cognitive Training in Adolescence. *Psychological Science*, 1–12.

Koerber, S., Mayer, D., Osterhaus, C., Schwippert, K., & Sodian, B. (2015). The development of scientific thinking in elementary school: A comprehensive inventory. *Child Development*, *86*(1), 327–336.

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development*, *84*(6), 1933–1953.

Leech, R., Mareschal, D., & Cooper, R. P. (2007). Relations as transformations: Implications for analogical reasoning. *Quarterly Journal of Experimental Psychology*, *60*(7), 897–908.

Leech, R., Mareschal, D., & Cooper, R. P. (2008). Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences*, *31*, 357–414.

LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, *81*(6), 1753–1767.

Leon-Carrion, J., García-Orza, J., & Pérez-Santamaría, F. J. (2004). Development of the inhibitory component of the executive functions in children and adolescents. *International Journal of Neuroscience*, *114*(10), 1291–1311.

Linzarini, A., Houdé, O., & Borst, G. (2015). When Stroop helps Piaget: An inter-task positive priming paradigm in 9-year-old children. *Journal of Experimental Child Psychology*, *139*, 71–82.

Lortie-Forgues, H., Tian, J., & Siegler, R. S. (2015). Why is learning fraction and decimal arithmetic so difficult? *Developmental Review*, *38*, 201–221.

Lubin, A., Vidal, J., Lanoë, C., Houdé, O., & Borst, G. (2013). Inhibitory control is needed for the resolution of arithmetic word problems: A developmental negative priming study. *Journal of Educational Psychology*, *105*(3), 701–708.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*(5), 1357–1372. Retrieved from http://www.wpic.pitt.edu/research/lncd/articles/Luna_2004_ChildDevel.pdf

Luna, B., Marek, S., Larsen, B., Tervo-Clemmens, B., & Chahal, R. (2015). An integrative model of the maturation of cognitive control. *Annual Review of Neuroscience*, *38*(1), 151–170.

Luo, Q., Perry, C., Peng, D., Jin, Z., Xu, D., Ding, G., & Xu, S. (2003). The neural substrate of analogical reasoning: an fMRI study. *Brain Research. Cognitive Brain Research*, *17*(3), 527–34. http://doi.org/10.1016/S0926-6410(03)00167-8

Mackey, A. P., Hill, S. S., Stone, S. I., & Bunge, S. A. (2011). Differential effects of reasoning and speed training in children. *Developmental Science*, *14*(3), 582–590.

Mackey, A. P., Miller Singley, A. T., & Bunge, S. A. (2013). Intensive reasoning training alters patterns of brain connectivity at rest. *The Journal of Neuroscience*, *33*(11), 4796–4803.

Mackey, A. P., Whitaker, K. J., & Bunge, S. A. (2012). Experience-dependent plasticity in white matter microstructure: Reasoning training alters structural connectivity. *Frontiers in Neuroanatomy*, *6*, 1–9.

Magis-Weinberg, L., Blakemore, S.-J., & Dumontheil, I. (2017). Social and nonsocial relational reasoning in adolscence and adulthood. *Journal of Cognitive Neuroscience*, *29*(10), 1739–1754.

Mareschal, D. (2016). The neuroscience of conceptual learning in science and mathematics. *Current Opinion in Behavioral Sciences*, *10*, 114–118.

Marsh, R., Zhu, H., Schultz, R. T., Quackenbush, G., Royal, J., Skudlarski, P., & Peterson, B. S. (2006). A developmental fMRI study of self-regulatory control. *Human Brain Mapping*, *27*(11), 848–863.

Marsh, R., Zhu, H., Schultz, R. T., Quackenbush, G., Royal, J., Skudlarski, P., & Peterson, B. S. (2008). A developmental fMRI study of self-regulatory control. *October*, *141*(4), 520–529.

Mason, R. A., & Just, M. A. (2016). Neural representations of physics concepts. *Psychological Science*, *27*(6), 904–913.

Masson, S., Potvin, P., Riopel, M., & Brault Foisy, L.-M. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education*, *8*(1), 44–55.

Matlen, B., Vosniadou, S., Jee, B., & Ptouchkina, M. (2009). Enhancing the comprehension of science text through visual analogies. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*(33), 2910–2915.

Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction*, *29*, 43–55.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291.

Merkley, R., Thompson, J., & Scerif, G. (2016). Of huge mice and tiny elephants: Exploring the relationship between inhibitory processes and preschool math skills. *Frontiers in Psychology*, *6*, 1903.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Monette, S., Bigras, M., & Guay, M. C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *Journal of Experimental Child Psychology*, *109*(2), 158–173.

Morooka, T., Ogino, T., Takeuchi, A., Hanafusa, K., Oka, M., & Ohtsuka, Y. (2012). Relationships between the color-word matching Stroop task and the Go/NoGo task: Toward multifaceted assessment of attention and inhibition abilities of children. *Acta Medica Okayama*, *66*(5), 377–386.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220–246.

Ordaz, S. J., Foran, W., Velanova, K., & Luna, B. (2013). Longitudinal growth curves of brain function underlying inhibitory control through adolescence. *Journal of Neuroscience*, *33*(46), 18109–18124.

Parsons, R. (2014). *KS3 maths: Complete study and practice*. Newcaste upon Tyne, UK, UK: Coordination Group Publications Ltd.

Parsons, R., & Gannon, P. (2014). *KS3 science: Complete study and practice*. Newcastle upon Tyne, UK, UK: Coordination Group Publications Ltd.

Peters, L., & De Smedt, B. (2018). Arithmetic in the developing brain: A review of brain imaging studies. *Developmental Cognitive Neuroscience*, *30*(November 2016), 265–279. http://doi.org/10.1016/j.dcn.2017.05.002

Petersen, I. T., Hoyniak, C. P., McQuillan, M. E., Bates, J. E., & Staples, A. D. (2016). Measuring the development of inhibitory control: The challenge of heterotypic continuity. *Developmental Review*, *40*, 25–71.

Petersen, S. E., & Dubis, J. W. (2012). The mixed block/event-related design. *NeuroImage*, *62*(2), 1177–1184.

Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology*, *31*(2), 153–179.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227.

Rhodes, S. M., Booth, J. N., Campbell, L. E., Blythe, R. A., Wheate, N. J., & Delibegovic, M. (2014). Evidence for a role of executive functions in learning biology. *Infant and Child Development*, *23*(1), 67–83.

Rhodes, S. M., Booth, J. N., Palmer, L. E., Blythe, R. A., Delibegovic, M., & Wheate, N. J. (2016). Executive functions predict conceptual learning of science. *British Journal of Developmental Psychology*, *34*(2), 261–275.

Richards, J. E., Sanchez, C., Phillips-Meek, M., & Xie, W. (2016). A database of age-appropriate average MRI templates. *NeuroImage*, *124*, 1254–1259.

Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science*, *24*(1), 87–92.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *94*(3), 249–273.

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science*, *316*(5828), 1128–1129.

Ridderinkhof, K. R., & van der Molen, M. W. (1995). A psychophysiological analysis of developmental differences in the ability to resist inteference. *Child Development*, *66*(4), 1040–1056.

Rivera, S. M. 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.

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, *12*, 191–200.

Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education*, *37*, 43–50.

Rubia, K., Smith, A. B., Taylor, E., & Brammer, M. (2007). Linear age-correlated functional development of right inferior fronto-striato-cerebellar networks during response inhibition and anterior cingulate during error-related processes. *Human Brain Mapping*, *28*(11), 1163–1177.

Ryan, J., & Williams, J. (2007). *Children's mathematics 4-15: Learning from errors and misconceptions*. Maidenhead, UK, UK: Open University Press.

Samarapungavan, A., Mantzicopoulos, P., Patrick, H., & French, B. (2009). The development and validation of the Science Learning Assessment (SLA). *Journal of Advanced Academics*, *20*(3), 502–535.

Siegel, J. S., Power, J. D., Dubis, J. W., Vogel, A. C., Church, J. A., Schlaggar, B. L., & Petersen, S. E. (2014). Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Human Brain Mapping*, *35*(5), 1981–1996.

Siegler, R. S. (1998). *Children's Thinking* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. New Jersey: Lawrence Erlbaum Associates, Inc.

Simmonds, A. (2014). *How neuroscience is affecting education: Report of teacher and parent surveys*. London.

Simmonds, D. J., Pekar, J. J., & Mostofsky, S. H. (2008). Meta-analysis of go/no-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia*, *46*(1), 224–232.

Singh-Curry, V., & Husain, M. (2009). The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia*, *47*(6), 1434–1448.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96.

Starkey, P., & Cooper, R. (1980). Perception of numbers by human infants. *Science*, *210*(4473), 1033–1035.

Starkey, P., & Cooper, R. G. (1995). The development of subitizing in young children. *British Journal of Developmental Psychology, 13*(4), 399–420.

Stavy, R., & Babai, R. (2010). Overcoming intuitive interference in mathematics: Insights from behavioral, brain imaging and intervention studies. *ZDM Mathematics Education*, *42*(6), 621–633.

Stavy, R., & Tirosh, D. (2000). *How students (mis-)understand science and mathematics: Intuitive Rules*. New York: Teachers College Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662.

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.

The FIL methods group. (2015). fMRI model specification. In *SPM12 Manual*. London, UK.

The Royal Society. (2011). *Brain Waves Module 2: Neuroscience: Implications for education and lifelong learning*. London, UK.

The Royal Society. (2014). *Vision for science and mathematics education*. London, UK.

Tolmie, A. K., Ghazali, Z., & Morris, S. (2016). Children's science learning: A core skills approach. *British Journal of Educational Psychology*, *86*(3), 481–497.

Trickett, S. B., Trafton, J. G., & Schunn, C. Dc. D. (2009). How do scientists respond to anomalies? Different strategies used in basic and applied science. *Topics in Cognitive Science*, *1*(4), 711–729.

Turner, R. (2016). Uses, misuses, new uses and fundamental limitations of magnetic resonance imaging in cognitive science. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1705), 20150349.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*(5), 427–449.

Varma, S., McCandliss, B. D., & Schwartz, D. L. (2008). Scientific and pragmatic challenges for bridging education and neuroscience. *Educational Researcher*, *37*(3), 140–152.

Vendetti, M. S., Matlen, B. J., Richland, L. E., & Bunge, S. A. (2015). Analogical reasoning in the classroom: Insights from cognitive science. *Mind, Brain, and Education*, *9*(2), 100–106.

Verbruggen, F., Liefooghe, B., & Vandierendonck, A. (2004). The interaction between stop signal inhibition and distractor interference in the flanker and Stroop task. *Acta Psychologica*, *116*(1), 21–37.

Visser, M., Jefferies, E., Embleton, K. V, & Ralph, M. A. L. (2012). Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: Distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *Journal of Cognitive Neuroscience*, *24*(8), 1766–1778.

Vosniadou, S. (2007). Conceptual change and education. *Human Development*, *50*(1), 47–54.

Vosniadou, S., Pnevmatikos, D., Makris, N., Eikospentaki, K., & Chountala, A. (2018). The recruitment of shifting and inhibition in on-line science and mathematics tasks. *Cognitive Science*, *42*(6), 1860–1886.

Watanabe, J., Sugiura, M., Sato, K., Sato, Y., Maeda, Y., Matsue, Y., … Kawashima, R. (2002). The human prefrontal and parietal association cortices are involved in NO-GO performances: An event-related fMRI study. *NeuroImage*, *17*(3), 1207–1216.

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

Whitaker, K. J., Vendetti, M. S., Wendelken, C., & Bunge, S. A. (2017). Neuroscientific insights into the development of analogical reasoning. *Developmental Science*, *21*(2), e12531.

White, C. S., Alexander, P. A., & Daugherty, M. (1998). The relationship between young children’s analogical reasoning and mathematical learning. *Mathematical Cognition*, *4*(2), 103–123.

Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999).

Development of inhibitory control across the life span. *Developmental Psychology, 35*(1), 205–213.

Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience, 1*, 1–8.

Xu, J., Moeller, S., Auerbach, E. J., Strupp, J., Smith, S. M., Feinberg, D. A., … Uğurbil, K. (2013). Evaluation of slice accelerations using multiband echo planar imaging at 3T. *NeuroImage*, *83*, 991–1001.

Zaitchik, D., Iqbal, Y., & Carey, S. (2014). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development*, *85*(1), 160–175.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*(1), 99–149.

# Appendix 1

Information sheets and consent forms from the study described in **Chapter 3**.

Centre for Educational Neuroscience

Birkbeck UNIVERSITY OF LONDON

**Department of Psychological Sciences**
**Birkbeck, University of London**
**Parent/Guardian Consent Form**

**Study title:** The development of science and maths reasoning in relation to inhibitory control
**Researcher:** Annie Brookman (abrook07@mail.bbk.ac.uk)
**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)
**Ethics number:** 141552

I have had the details of the study explained to me and willingly consent for my child to take part.

My questions have been answered to my satisfaction and I understand that I may ask further questions at any time.

I understand that my child will remain anonymous and that all the information given will be used for this study only.

I understand that I may withdraw my consent for the study at any time without giving any reason.

I understand that all information given will remain confidential except in the highly unlikely event that the researcher has a serious concern regarding a child protection issue.

I confirm that my child does not have a history of any developmental or neurological disorders (such as dyspraxia, epilepsy, or ADHD).

**Participant**

Name of child: ………………………………………………… Child date of birth:………….
Class: ………………
Name of Parent/Guardian: ……………………………… Signature: ……………………..
Date: ………………

**Researcher**

Name of researcher: …………………………………………………
Researcher signature: ……………………………………..... Date: ………………

**Department of Psychological Sciences**

**Birkbeck, University of London**

**Pupil Consent Form (ages 13-15)**

**Study title:** The development of science and maths reasoning in relation to inhibitory control

**Researcher:** Annie Brookman (abrook07@mail.bbk.ac.uk)

**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)

**Ethics number:** 141552

I have had the details of the study explained to me and I'm happy to take part.

I'm happy that my questions have been answered and I understand I can ask more questions at any time.

I understand that my name will not be linked to my scores on the different tests and that my results will only be used in this study and not given to anyone else.

I understand that the information collected about me will not be given to anyone else, unless the researcher has serious worries about my safety.

I understand that I can say 'stop, I don't want to do this anymore', at any time, without having to give a reason.

**Participant**

Name: ………………………………………………

Signature: …………………………………

Date: ………………

**Researcher**

Name: ………………………………………………

Signature: …………………………………

Date: ………………

**Department of Psychological Sciences**

**Birkbeck, University of London**

**Parent/Guardian Information Sheet**

**Study title:** The development of science and maths reasoning in relation to inhibitory control

**Researcher:** Annie Brookman (abrook07@mail.bbk.ac.uk)

**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)

**Ethics number:** 141552

Dear Parents/Guardians,

We would like to invite your child to participate in a research project. Participation is totally voluntary; choosing not to let your child take part will not disadvantage you or them in any way. Before you decide whether or not you want your child to take part, it is important for you to read the following information carefully and discuss it with others if you wish. Please contact the researcher on this project, Annie Brookman (email: abrook07@mail.bbk.ac.uk, phone: 020 7079 0703), if there is anything that is not clear, or if you would like more information.

**What is the study about?**

We are investigating the development of science and maths reasoning in pupils in years 7 to 10. We are interested in how understanding of maths and science relates to pupils' abilities to inhibit behaviours. Research suggests that adults need to inhibit their prior beliefs when reasoning about science and maths, but this has not been investigated in younger age groups before. We are therefore interested in seeing how maths and science reasoning relates to inhibition throughout early secondary school. We intend to see pupils for this project before the summer holidays.

**Why has my child been chosen?**

Your child's school has agreed to take part in our project, and information sheets have been sent to parents of all children in years 7 to 10.

**What will happen to my child if I agree to take part?**

We will be asking your child to complete several tasks. Most will be performed on a laptop. For example, your child will be asked to look at coloured squares on the screen and press a button corresponding to where on the screen the square is, but only when the square is a certain colour. Another task is a measure of language, and will require your child to tell the researcher the meaning of different words.

If you agree for your child to participate, s/he will be tested in a quiet room at school by the trained researcher, Annie Brookman (a Masters student at Birkbeck, University of London). The tasks will be fully explained and any questions answered. Rest breaks will be provided and the researcher will be present throughout the whole experiment. Your child will be made aware that if they wish to stop the experiment for any reason at any point during the study, it will be discontinued without them having to give a reason. The study session will take no longer than an hour and will be during the school day. We are working with the school to ensure that minimal disruption to your child's school day will occur, and the timing of the study sessions do not interfere with any examination periods. Finally, we will ask your child's science and maths teachers for their average grades in these subjects.

Risks: There are not thought to be any risks associated with any of the tasks your child will be asked to take part in.

Benefits: The study is not intended to give direct benefit to the participants themselves. However, it is hoped that any knowledge gained as a result of the study will be able to help inform further studies into maths and science reasoning and learning. Ultimately we are aiming to help inform educational practice in these areas.

**What will happen to the results?**

The results of the study will be used as part of Annie Brookman's Masters dissertation and will not be shared with the school. Your son or daughter will not be identifiable in the dissertation or any publication that might ensue, which will report group averages. Your child's involvement in the study will remain confidential except in the highly unlikely event that the researcher has a serious concern regarding a child protection issue. All data will be collected and stored in accordance with the Data Protection Act 1998.

Your child does not have to take part in this study if you do not want them to or if they do not want to. Participation is entirely voluntary. You/your child are free to decline to enter or to withdraw from the study at any time without having to give a reason. Participation in this study will in no way affect your/your child's legal rights, or current or future medical treatment. The project has received ethical approval from the Department of Psychological Sciences Research Ethics Committee of Birkbeck, University of London.

**How do we take part?**

If you decide that your child can take part, we request that you sign the consent form that is enclosed with this information pack. Please return the signed consent form to your child's tutor, who will pass it on to the researcher. Even after signing the consent form you are still free to withdraw your child from the study at any time and without giving a reason. If you give consent for your child to take part in the study the researcher will work with the school to book a convenient time for your child to take part in the study.

Thank you very much for your interest in our research.  If you have any further questions please contact Annie Brookman. If you'd like to speak to the supervisor of this project directly, please contact Dr Iroise Dumontheil.

Researcher:
Annie Brookman
Tel: 020 7079 0703
Email: abrook07@mail.bbk.ac.uk

Supervisor:
Dr Iroise Dumontheil
Tel: 020 3073 8008
Email: i.dumontheil@bbk.ac.uk

**Department of Psychological Sciences**

**Birkbeck, University of London**

## Pupil Information Sheet (ages 11-15)

The tasks will take about an hour during the school day.

You will do the tasks with Annie Brookman, who is a student at Birkbeck, University of London.

If you want to take part, we may ask you to sign a form. If you want to stop at any time, just say that you don't want to take part. You don't have to give a reason and nobody will mind.


You will take part in some tasks on a laptop, one in a book, and one where you have to say the meanings of words.

Nobody will know what answers you gave, and the results will be anonymous, so your name will not be stored with your answers.

We are interested in how pupils in secondary school think about science and maths. These tasks will help us to understand more about these skills.

The study will take place in a quiet room in your school.

### Important points

- You can stop each task whenever you like, without being questioned.
- The information collected about you will not be given to anyone else, unless the researcher has serious worries about your safety.
- If you have any questions you can ask the researcher: Annie Brookman, Tel: 020 7079 0703, Email: abrook07@mail.bbk.ac.uk

Information sheets and consent forms from the studies described in **Chapter 4**.

**Department of Psychological Sciences**
**Birkbeck, University of London**

**Parent/Guardian Information Sheet**

**Study title:** The role of inhibitory control and working memory in learning new concepts in [science/maths]
**Researcher:** Annie Brookman-Byrne (abrook07@mail.bbk.ac.uk)
**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)
**Ethics approval number:** 151631

Dear Parent/Guardian,

[School name] is working with researchers at Birkbeck, University of London, to investigate how pupils learn and reason about [science/maths]. Your child's class has been chosen to take part in a study that looks at the skills involved in learning new concepts in [science/maths].

**What is the study about?**
We are interested in how learning new [science/maths] concepts relates to a cognitive skill called inhibitory control. This is the ability to withhold an automatic response, and is known to develop throughout adolescence. Research suggests that we need to inhibit prior beliefs when reasoning about science, but the relationship between inhibitory control and [science/maths] *learning* has not been investigated before. We also interested in learning how the role of working memory, the ability to hold and manipulate information in mind, is related to learning in [science/maths]. We are therefore researching this relationship in [School name] pupils.

**What will happen during the study?**
The whole class will take part in some online tasks in a school computer room during the lunchbreak on a normal school day. There will be two tasks and the session will take around 15-20 minutes in total. Annie Brookman-Byrne, a PhD student at Birkbeck, will be there to explain the tasks and answer any questions. The tasks will be carried out individually by each pupil at the same time. In these tasks, your child will be asked to press a key as quickly as possible in response to colours, numbers and dots seen on the screen.

During a [science/maths] lesson, the whole class will be given a short test of their understanding of different concepts in [science/maths] that will last around 10 minutes. Annie will then observe some of the [science/maths] lessons. Two weeks later they will be given another [science/maths] test in a lesson, to see what they have learnt. Finally, approximately three weeks later, they will be given another [science/maths] test during a lesson. This will allow us to see how well they have learnt the information given to them in their lessons. The study will take place after the Easter holidays.

We will also ask the [science/maths] teacher for each member of the class's [science/maths] marks to help us gain a full picture of [science/maths] performance. After the study, we may decide that we would like to do more tasks with the class, to explore the skills involved in [science/maths] learning further. If this is the case, we will contact you again with another information sheet and consent form.

**What will happen to the results?**

The results of the study will be used as part of Annie Brookman-Byrne's PhD thesis, published in scientific journals, and presented at conferences. Your child will not be identifiable in the thesis or any publication that might ensue, which will report group averages. Your child's involvement in the study will remain confidential except in the highly unlikely event that the researcher has a serious concern regarding a child protection issue. All data will be collected and stored in accordance with the Data Protection Act 1998. You and your child are free to withdraw from the study at any time without having to give a reason, by getting in touch with Annie or her supervisor (see below). The project has received ethical approval from the Department of Psychological Sciences Research Ethics Committee of Birkbeck, University of London.

**How do we take part?**

Attached is an opt-out consent form.

- *If you agree for your child to take part, you do not have to do anything.*
- *If you do not wish for your child to take part in this research, please complete the form and return it to the school.*

If you have any questions about the study, please contact us (see below).

Thank you!

**Researcher:**

Annie Brookman-Byrne, Tel: 020 7079 0777, Email: abrook07@mail.bbk.ac.uk

**Supervisor:**

Dr Iroise Dumontheil, Tel: 020 3073 8008, Email: i.dumontheil@bbk.ac.uk

**Lab website:**

http://sites.google.com/site/idcnlab/

Centre for Educational Neuroscience

Birkbeck UNIVERSITY OF LONDON

**Department of Psychological Sciences**
**Birkbeck, University of London**

**Pupil Information Sheet**

The study will take place after half term.

Your whole class will take part, but you will do each task individually.

You will complete two tasks on a computer, and some short maths tests during maths lessons.

The computer tasks will take about fifteen minutes, and the maths tests will take about ten minutes each.

We may also ask your teacher for your maths marks.


If you want to stop at any time, just say that you don't want to take part. You don't have to give a reason and nobody will mind.

Nobody will know what answers you gave, and the results will be anonymous, so your name will not be stored with your answers.

We are interested in how pupils in secondary school learn new concepts in maths. These tasks will help us to understand more about how pupils learn.

The study will take place in your maths classroom and possibly a school computer room.

**Important points**

- You can stop each task whenever you like, without being questioned
- The information collected about you will not be given to anyone else, unless the researcher has serious worries about your safety
- If you have any questions you can ask the researcher:
  Annie Brookman-Byrne, Tel: 020 7079 0777, Email: abrook07@mail.bbk.ac.uk

Centre for Educational Neuroscience

Birkbeck UNIVERSITY OF LONDON

**Department of Psychological Sciences**
**Birkbeck, University of London**

**Parent/Guardian opt-out consent form**

**Study title:** The role of inhibitory control and working memory in learning new concepts in [science/maths]
**Researcher:** Annie Brookman-Byrne (abrook07@mail.bbk.ac.uk)
**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)
**Ethics approval number:** 151631

I have had the details of the study explained to me. My questions have been answered to my satisfaction and I understand that I may ask further questions at any time.

I understand that my child will remain anonymous and that all the information given will be used for this study only.

I understand that all information given will remain confidential except in the highly unlikely event that the researcher has a serious concern regarding a child protection issue.

I understand that we may be contacted to take part in further research, and that if it is the case, I will receive further information and will be able to decide whether I would like my child to take part.

I understand that I may withdraw my consent for the present study at any time and will not need to justify my decision.

As the whole class is taking part in the study, we are using an **opt-out consent** approach for this research. This means that your child will be taking part in the research, **EXCEPT** if you **OPT-OUT** by filling in the form below and returning it to [School name].

-----------------------------------------------------------------------------------------------

To be completed by a parent or guardian who **DOES NOT AGREE** to their child taking part in the research study:

I **DO NOT** wish for my child to take part in the above study ☐

Please use BLOCK CAPITALS

Parent/Guardian name ……………………………………………………………………………………………………

Child full name ……………………………………………………………………………………………………………

Signature of Parent/Guardian ……………………………………………………… Date ……………………………

Department of Psychological Sciences
Birkbeck, University of London

**Pupil consent form**

**Study title:** The role of inhibitory control and working memory in learning new concepts in [science/maths]
**Researcher:** Annie Brookman-Byrne (abrook07@mail.bbk.ac.uk)
**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)
**Ethics approval number:** 151631

I have had the details of the study explained to me and I'm happy to take part.

I'm happy that my questions have been answered and I understand I can ask more questions at any time.

I understand that my name will not be linked to my scores on the different tests and that my results will only be used in this study and not given to anyone else.

I understand that the information collected about me will not be given to anyone else, unless the researcher has serious worries about my safety.

I understand that I can say 'stop, I don't want to do this anymore', at any time, without having to give a reason.

**Participant**

Name: ………………………………………………………………………………

Signature: ………………………………………………………………………

Date: …………………………………………………………………

**Researcher**

Name: ………………………………………………………………………………

Signature: ………………………………………………………………………

Date: …………………………………………………………………

**Department of Psychological Sciences**
**Birkbeck, University of London**

**Parent/Guardian Information Sheet – Follow up test**

**Study title:** The role of inhibitory control and working memory in learning new concepts in [science/maths]
**Researcher:** Annie Brookman-Byrne (abrook07@mail.bbk.ac.uk)
**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)
**Ethics approval number:** 151631

Dear Parent/Guardian,

You may remember that last academic year your child took part in a study with researchers at Birkbeck, University of London, to investigate how pupils learn and reason about [science/maths]. We would like to see your child for a further five minutes to complete one more computer test.

**What is the study about?**
We are interested in how learning new [science/maths] concepts relates to a cognitive skill called inhibitory control. This is the ability to withhold an automatic response, and is known to develop throughout adolescence. Research suggests that we need to inhibit prior beliefs when reasoning about science, but the relationship between inhibitory control and [science/maths] *learning* has not been investigated before. We are also interested in learning how the role of working memory, the ability to hold and manipulate information in mind, is related to learning in [science/maths]. We are therefore researching this relationship in [School name] pupils and would like to test this ability in pupils who have already taken part.

**What will happen during the study?**
There will be one task to complete for about 5 minutes. Annie Brookman-Byrne, a PhD student at Birkbeck, will explain the task and answer any questions. The task will take place either one-to-one with Annie, or in a computer room as a class. Your child will be asked to click the mouse as quickly as possible in response to dots seen on the screen.

**What will happen to the results?**
The results of the study will be used as part of Annie Brookman-Byrne's PhD thesis, published in a scientific journal, and presented at conferences. Your child will not be identifiable in the thesis or any publication that might ensue, which will report group averages. Your child's involvement in the study will remain confidential except in the highly unlikely event that the researcher has a serious concern regarding a child protection issue. All data will be collected and stored in accordance with the Data Protection Act 1998. You and your child are free to withdraw from the study at any time without having to give a reason, by getting in touch with Annie or her supervisor (see below). The project has received ethical approval from the Department of Psychological Sciences Research Ethics Committee of Birkbeck, University of London.

**How do we take part?**
Attached is an opt-out consent form.

- *If you agree for your child to take part, you do not have to do anything.*
- *If you do not wish for your child to take part in this research, please complete the form and return it to the school.*

If you have any questions about the study, please contact us (see below).

Thank you!

**Researcher:**
Annie Brookman-Byrne, Tel: 020 7079 0777, Email: abrook07@mail.bbk.ac.uk

**Supervisor:**
Dr Iroise Dumontheil, Tel: 020 3073 8008, Email: i.dumontheil@bbk.ac.uk

**Lab website:**
http://sites.google.com/site/idcnlab/

Centre for Educational Neuroscience

Birkbeck UNIVERSITY OF LONDON

**Department of Psychological Sciences**
**Birkbeck, University of London**

**Pupil Information Sheet – Follow up test**

Last academic year you took part in a research project about learning new concepts in science. We would like to see you for another five minutes to complete one final test.

We are interested in how pupils in secondary school learn new concepts in science. There are no more science tests to do, we are just inviting you to take part in a computer game where you have to respond to dots on a screen by clicking the mouse.

This will either take place one-to-one with the researcher, Annie, or as a whole class in the computer room.

**Important points**

- You can stop the task whenever you like, without being questioned
- The information collected about you will not be given to anyone else, unless the researcher has serious worries about your safety
- If you have any questions you can ask the researcher:
  Annie Brookman-Byrne, Tel: 020 7079 0777, Email: abrook07@mail.bbk.ac.uk

Information sheets and consent forms from the study described in **Chapters 5 and 6**.

**Department of Psychological Sciences**
**Birkbeck, University of London**
**Parent/Guardian Information Sheet**

**Study title:** Neuroimaging study of the role of executive functions in science and maths reasoning
**Researcher:** Annie Brookman (abrook07@mail.bbk.ac.uk), Jack White (jwhite18@mail.bbk.ac.uk)
**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)
**Ethics approval number:** 1602-005

Dear Parent(s)/Guardian(s),
We would like to invite your child to participate in a research project. Participation is totally voluntary; choosing not to let your child take part will not disadvantage you or them in any way. Before you decide whether or not you want your child to take part, it is important for you to read the following information carefully and discuss it with others if you wish. Please contact the researchers on this project, Annie Brookman (abrook07@mail.bbk.ac.uk; 020 7079 0703), Jack White (jwhite18@mail.bbk.ac.uk), or Dr. Iroise Dumontheil (i.dumontheil@bbk.ac.uk; 020 3073 8008) if there is anything that is not clear, or if you would like more information.

**What is the study about?**
We are investigating the development of science and maths reasoning in pupils in Years 7 to 10. We are interested in how understanding maths and science relates to pupils' executive functions, the set of mental mechanisms that allow us to keep important information in mind and use it to achieve our goals. Executive functions are still developing through adolescence, and are thought to have specific influence on science and maths reasoning. We are therefore interested in seeing how maths and science reasoning relates to executive functions throughout early secondary school using computerised tasks and magnetic resonance imaging (MRI).

**Why study this with an MRI machine?**
Functional magnetic resonance imaging (fMRI) allows us to view brain activity in real time while your child is carrying out some simple science, maths and executive function tasks. This will help us see if their brain uses similar brain regions for the different types of tasks. MRI scanners are basically a very large and strong magnet, which is completely harmless for your child. You can watch a video of someone being scanned at BUCNI here:
http://www.drru-research.org/pages/what-its-like-to-have-a-brain-scan.html

**Why has my child been chosen?**
Your child's school has agreed to take part in our project, and information sheets have been sent to parents of children in Years 7 to 10.

**What will happen to my child if I agree to take part?**

We will be asking your child to complete several simple tasks. Some will be performed on a laptop and others will be performed while inside the MRI scanner. For example, they may be asked to view different scientific and mathematical pictures and sentences, and asked whether they are true or false. They will be asked to click on a button to make their choice. In other tasks they may be asked to remember numbers, or to press a button as quickly as possible if they see a green square, but not press any button if they see a red square on the screen. Some tasks will be completed by your child outside the scanner and may include telling us the meanings of words or choosing which picture completes a pattern.

**Where and when will this happen?**
If you agree for your child to participate, s/he will be tested at the Birkbeck and UCL Centre for Neuroimaging (BUCNI), in central London, by trained researchers, Annie Brookman (a PhD student at Birkbeck College, University of London), Jack White (a Master's student at Birkbeck), and their supervisor, Dr. Iroise Dumontheil (Reader in Cognitive Neuroscience at Birkbeck). The tasks will be fully explained and any questions answered. Rest breaks will be provided and a researcher will be present throughout the whole experiment. Your child will be made aware that if they wish to stop the experiment for any reason at any point during the study, it will be discontinued without them having to give a reason. The study session will take no longer than two hours and will be during the day over the Easter holidays or on a weekend in April or May of this year.

In some cases, we may like to see your child again, and ask them to complete some additional tasks at school. These will be done on a laptop or using pen and paper. If we would like to see your child in school, we will give you a separate consent form to sign. The researchers will agree a convenient time to see your child at the school, within normal school hours.

**Are there any risks involved in taking part?**
The MRI scanning procedure requires that your child be confined in a small partially enclosed space. Some individuals find this to be uncomfortable and may exhibit symptoms of claustrophobia including nervousness, sweating or other minor discomfort. For this reason we cannot scan children who are uncomfortable in small spaces, like elevators.

The sound of the MRI scanner can be quite loud; your child will be given special earplugs to minimise the noise. In addition, the scanner is a very strong magnet, which means it attracts certain metals. Therefore, people with these metals within their bodies (such as dental braces, pacemakers, infusion pumps, aneurysm clips, metal prostheses, joints, rods, or plates) will be excluded from the study. The "metal" in dental fillings is less responsive to magnetism and is therefore allowed, but we cannot scan children who are wearing dental braces.

There are no other known side effects resulting from exposure to the MRI scan. The project has received ethical approval from the Department of Psychological Sciences Research Ethics Committee of Birkbeck, University of London.

**What are the benefits of taking part?**
The study is not intended to give direct benefit to the participants themselves. However, it is hoped that any knowledge gained as a result of the study will be able to help inform further studies into maths and science reasoning and learning. Ultimately we are aiming to help inform educational practice in these areas.

Participants will be compensated £10/hour of their time for taking part and will receive a picture of their brain. Travel cost for the participant and their parent/guardian will also be reimbursed.

**What will happen to the results?**

The results of the study will be used as part of Annie Brookman's PhD dissertation, Jack White's Master's dissertation, and research publications. Individual children's results will be anonymised and will not be shared with their school. Your child will not be identifiable in the dissertations or any publication that might ensue, which will report group averages. Your child's involvement in the study will remain confidential except in the highly unlikely event that a researcher has a serious concern regarding a child protection issue. All data will be collected and stored in accordance with the Data Protection Act 1998.

**Can we change our minds?**

Yes. Your child does not have to take part in this study if you do not want them to or if they do not want to. Participation is entirely voluntary. You and your child are free to decline to enter or to withdraw from the study at any time without having to give a reason. Participation in this study will in no way affect your legal rights or those of your child's, or current or future medical treatment.

**How do we take part?**

If you decide that your child can take part, please contact Annie Brookman: abrook07@mail.bbk.ac.uk; 020 7079 0703, who is running this project. She will discuss the study further with you, and will ask you to go through an MRI pre-screening form to check that there is no reason (such as claustrophobia, or dental braces) for your child to not take part, and will book a date and time for your child to do the testing session at BUCNI.

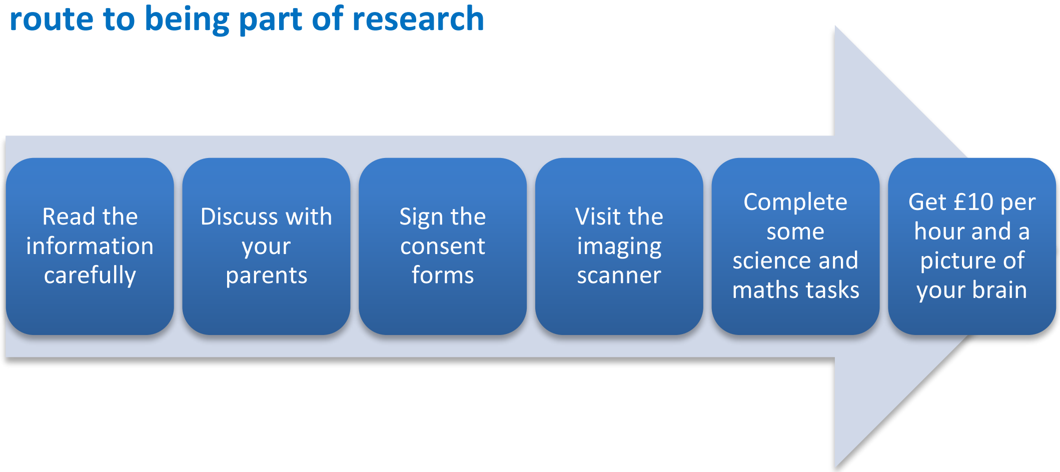
On the day we will ask you to sign a consent form stating that you understand what the study is about and that you agree for your child to take part. Even after signing the consent form you are still free to withdraw your child from the study at any time and without giving a reason. We will also ask you to complete and sign the MRI pre-screening form, and the operator running the MRI scanner will check that your child has removed all metal from around his or her body.

Thank you very much for your interest in our research. If you have any further questions please contact Annie Brookman or Jack White. If you'd like to speak to the supervisor of this project directly, please contact Dr Iroise Dumontheil.

**Researchers:**
Annie Brookman, Tel: 020 7079 0703, Email: abrook07@mail.bbk.ac.uk
Jack White, Email: jwhite18@mail.bbk.ac.uk
**Supervisor:**
Dr Iroise Dumontheil, Tel: 020 3073 8008, Email: i.dumontheil@bbk.ac.uk,
**Lab website:**
http://sites.google.com/site/idcnlab/

**Department of Psychological Sciences**
**Birkbeck, University of London**
**Participant Information Sheet**

## Your route to being part of research





**Important points**

- You can stop each task whenever you like, without being questioned
- The information collected about you will not be given to anyone else, unless the researcher has serious worries about your safety
- If you have any questions you can ask the researcher:
  Annie Brookman, Tel: 020 7079 0703, Email: abrook07@mail.bbk.ac.uk

**Department of Psychological Sciences**
**Birkbeck, University of London**
**Parent/Guardian Consent Form (fMRI)**

**Study title:** Neuroimaging study of the role of executive functions in science and maths reasoning
**Researchers:** Annie Brookman (abrook07@mail.bbk.ac.uk), Jack White (jwhite18@mail.bbk.ac.uk)
**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)
**Ethics approval number:** 1602-005

- I have had the details of the study explained to me and willingly consent for my child to take part.
- My questions have been answered to my satisfaction and I understand that I may ask further questions at any time.
- I understand that my child will remain anonymous and that all the information given will be used for this study only.
- I understand that I may withdraw my consent for the study at any time without giving any reason.
- I understand that all information given will remain confidential except in the highly unlikely event that the researcher has a serious concern regarding a child protection issue.
- I confirm that my child does not have a history of any developmental or neurological disorders (such as dyspraxia, epilepsy, or ADHD).
- I confirm that during the testing my child will not have any metal in or about their person.

**Signature and Acknowledgment**

__________________________________________ _____________ ____________
NAME OF CHILD (PLEASE PRINT) DATE OF BIRTH SCHOOL YEAR

________________________________________ ________________________________
NAME OF PARENT/CARER (PLEASE PRINT) EMAIL

________________________________________ ________________________________
SIGNATURE OF PARENT/CARER TODAY'S DATE

**Investigator's Statement**

I …………………………………………………………………. confirm that I have carefully explained the purpose of the study to the participant and his/her parent/carer and outlined any reasonably foreseeable risks or benefits (where applicable).

Signed: Date:

Centre for Educational Neuroscience

Birkbeck UNIVERSITY OF LONDON

**Department of Psychological Sciences**
**Birkbeck, University of London**

**Participant Consent Form (11-15y) (fMRI)**

**Study title:** Neuroimaging study of the role executive functions in science and maths reasoning
**Researcher:** Annie Brookman (abrook07@mail.bbk.ac.uk), Jack White (jwhite18@mail.bbk.ac.uk)
**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)
**Ethics approval number:** 1602-005

- I have had the details of the study explained to me and I'm happy to take part.
- I'm happy that my questions have been answered and I understand I can ask more questions at any time.
- I understand that my name will not be linked to my scores on the different tests and that my results will only be used in this study and not given to anyone else.
- I understand that the information collected about me will not be given to anyone else, unless the researcher has serious worries about my safety.
- I understand that I can say 'stop, I don't want to do this anymore', at any time, without having to give a reason.
- I understand that I should not have any metal (jewellery, piercings or clothing) in or on my body during the time of testing.

**Signature and Acknowledgment**

______________________________ ____________ ____________
NAME DATE OF BIRTH SCHOOL YEAR

______________________ ______________________
SIGNATURE TODAY'S DATE

**Investigator's Statement**

I ……………………………………………………………………. confirm that I have carefully explained the purpose of the study to the participant and his/her parent/carer and outlined any reasonably foreseeable risks or benefits (where applicable).

Signed: Date:

Centre for Educational Neuroscience

Birkbeck UNIVERSITY OF LONDON

**Department of Psychological Sciences**

**Birkbeck, University of London**

**Parent/Guardian Consent Form**

**Study title:** Relational reasoning in science and maths

**Researchers:** Annie Brookman-Byrne (abrook07@mail.bbk.ac.uk), Adrian Woodley-Cooper (woodleycooper@btinternet.com), Sara Kapika (s.m.kapika@student.vu.nl)

**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)

**Ethics number:** 171807

I have had the details of the study explained to me and willingly consent for my child to take part.

My questions have been answered to my satisfaction and I understand that I may ask further questions at any time.

I understand that the researchers will obtain my child's average grades in science and maths for the year, and their [MidYIS OR CAT OR non-verbal and verbal reasoning] test scores from the school.

I understand that my child's data will remain confidential. I understand that anonymous average scores will be put online for other scientists to use, but no information that could allow my child to be identified will be shared.

I understand that I may withdraw my consent for the study up to one year after the end of the project without giving any reason. After this date my child's data will no longer be linked to their name.

I understand that all information given will remain confidential except in the highly unlikely event that the researcher has a serious concern regarding a child protection issue.

I confirm that my child does not have a history of any developmental or neurological disorders (such as dyspraxia, epilepsy, or ADHD).

**Participant**

Child's name: ………………………………….. Child's date of birth: ………………

Sex: Female / Male School year: 7 / 8 / 9 / 10 Class: ……………

Parent/Guardian name: ……………………………………………….

Parent/Guardian signature: …………………………………. Today's date: ………………………

**For researcher to fill in**

Name of researcher: ……………………………………………….

Researcher signature: ……………………………………….. Date: ………………………

Please return this form to [School name]

Centre for Educational Neuroscience

Birkbeck UNIVERSITY OF LONDON

**Department of Psychological Sciences**
**Birkbeck, University of London**
**Pupil Consent Form (ages 13-15)**

**Study title:** Relational reasoning in science and maths
**Researchers:** Annie Brookman-Byrne (abrook07@mail.bbk.ac.uk), Adrian Woodley-Cooper (woodleycooper@btinternet.com), Sara Kapika (s.m.kapika@student.vu.nl)
**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)
**Ethics number:** 171807

I have had the details of the study explained to me and I'm happy to take part.

I'm happy that my questions have been answered and I understand I can ask more questions at any time.

I understand that the researchers will obtain my average grades in science and maths for the year and [MidYIS OR CAT OR non-verbal and verbal reasoning] test scores from the school.

I understand that my name will not be linked to my scores on the different tests. I understand that my scores will be stored online so that other scientists can do research on them but my my name or date of birth will not be shared.

I understand that my personal information will not be given to anyone else, unless the researcher has serious worries about my safety.

I understand that I can say 'stop, I don't want to do this anymore', at any time, without having to give a reason, and that I can ask for my data to be deleted up to one year after the end of the project.

**Participant**

Name: ………………………………………………….

Signature: ……………………………………..

Date: …………………

**For researcher to fill in**

Researcher name: ………………………………………………….

Signature: ……………………………………..

Date: …………………

**Department of Psychological Sciences**
**Birkbeck, University of London**
**Parent/Guardian Information Sheet**

**Study title:** Relational reasoning in science and maths

**Researchers:** Annie Brookman-Byrne (abrook07@mail.bbk.ac.uk), Adrian Woodley-Cooper (woodleycooper@btinternet.com), Sara Kapika (s.m.kapika@student.vu.nl)

**Supervisor:** Dr Iroise Dumontheil (i.dumontheil@bbk.ac.uk)

**Ethics number:** 171807

Dear Parents/Guardians,

We would like to invite your child to participate in a research project. Participation is voluntary; choosing not to let your child take part will not disadvantage you or them in any way. Before you decide whether or not you want your child to take part, it is important for you to read the following information carefully and discuss it with others if you wish. Please contact the researcher on this project, Annie Brookman-Byrne (email: abrook07@mail.bbk.ac.uk, phone: 020 7079 0777), if there is anything that is not clear, or if you would like more information.

**What is the study about?**

We are investigating the development of science and maths reasoning in pupils in years 7 to 10. We are trying to find out how important relational reasoning is for science and maths. Research suggests that the ability to see relations between items is important when solving science and maths problems. The role of relational reasoning in science and maths has not been investigated before. We are therefore interested in discovering the associations between relational reasoning and science and maths throughout early secondary school.

**Why has my child been chosen?**

[School name] have agreed to take part in Annie's PhD project, and information sheets have been sent to parents of all children in years 7 to 10.

**What will happen to my child if I agree to take part?**

We will ask your child to complete several tasks. Most will be performed on a laptop. For example, your child will be asked to look at science and maths statements and decide whether they are true or false. Another task requires pressing buttons as quickly as possible when coloured squares appear on the screen.

If you agree for your child to participate, they will be tested in a quiet room at school by the trained researcher, Annie Brookman-Byrne (a PhD student at Birkbeck, University of London) or Adrian Woodley-Cooper or Sara Kapika (also from Birkbeck, University of London) who have been trained by Annie. The tasks will be fully explained and any questions answered. Rest breaks will be provided and the researcher will be present throughout the whole experiment. Your child will be made aware that if they wish to stop the experiment for any reason at any point during the study, it will be discontinued without them having to give a reason. The study session will take approximately 45 minutes during the school day. We are working with the school to ensure that minimal disruption to your child's day will occur, and the timing of the study sessions do not interfere with any important lessons or examination periods. Finally, the school will provide the researchers with your child's science and maths average grade for the current year, and your child's [MidYIS or CAT or verbal and non-verbal reasoning] scores.

The study is not intended to give direct benefit to the participants themselves. However, it is hoped that any knowledge gained as a result of the study will be able to help inform further studies into science and maths reasoning and learning. Ultimately we are aiming to help inform educational practice in these areas. There are not thought to be any risks associated with any of the tasks your child will be asked to take part in.

**What will happen to the results?**

The results of the study will be used as part of Annie Brookman-Byrne's PhD thesis and university students' dissertations, and individual results will not be shared with the school. Your son or daughter will not be identifiable in the thesis or any publication that might ensue, which will report group averages and compare performance in the different tasks. Your child's involvement in the study will remain confidential except in the highly unlikely event that the researcher has a serious concern regarding a child protection issue. All data will be collected and stored in accordance with the Data Protection Act 1998. All raw data and consent forms will be destroyed 10 years after the end of the study. Anonymised summary data (average scores on the different tests) will be stored online at the end of the study so that other scientists can check our work or perform their own analyses.

Your child does not have to take part in this study if you do not want them to or if they do not want to. Participation is entirely voluntary. You and your child are free to decline to enter. If you decide that your child can take part, you are free to later withdraw their data from the study up to one year after the end of the project without having to give a reason. After this time, it will no longer be possible to link their name to their data. Participation in this study will in no way affect your or your child's legal rights, or current or future medical treatment. The project has received ethical approval from the Department of Psychological Sciences Research Ethics Committee of Birkbeck, University of London.

**How do we take part?**

If you decide that your child can take part, please sign the consent form that is enclosed with this information pack. Please return the signed consent form to your child's tutor, who will pass it on to Annie. Even after signing the consent form you are still free to withdraw your child from the study by contacting Annie or the school without giving a reason. If you give consent for your child to take part in the study Annie will work with the school to book a convenient time for your child to take part in the study. Your child's participation will take place by the end of the current school year. Unfortunately, we are unable to include children who have any diagnosed developmental or neurological disorders (such as dyspraxia, epilepsy, or ADHD).

Thank you very much for your interest in our research. If you have any further questions please contact Annie Brookman-Byrne. If you'd like to speak to the supervisor of this project directly, please contact Dr Iroise Dumontheil.

**Researchers:** Annie Brookman-Byrne, tel: 020 7079 0777, email: abrook07@mail.bbk.ac.uk
Adrian Woodley-Cooper (woodleycooper@btinternet.com), Sara Kapika (s.m.kapika@student.vu.nl)
**Supervisor:** Dr Iroise Dumontheil, tel: 020 3073 8008, email: i.dumontheil@bbk.ac.uk
**Lab website:** http://sites.google.com/site/idcnlab/

**Department of Psychological Sciences**
**Birkbeck, University of London**

**Pupil Information Sheet (ages 11-15)**

The tasks will take about 45 minutes during the school day at a convenient time.

You will do the tasks with Annie, Adrian, or Sara who are from Birkbeck, University of London.

Anonymous average data, without your name or date of birth, will be put online.

If you want to take part, we may ask you to sign a form. If you want to stop at any time, just say that you don't want to take part. You don't have to give a reason and nobody will mind.


You will take part in some tasks on a laptop and one using pen and paper.

Nobody will know what answers you gave, and the results will be confidential, so your name will not be stored with your answers.

We are interested in how pupils in secondary school think about science and maths. These tasks will help us to understand more about these skills.

The study will take place in a quiet room at school.

**Important points**

- You can stop each task whenever you like, without being questioned.
- The information collected about you will not be given to anyone else, unless the researcher has serious worries about your safety.
- If you have any questions you can ask the main researcher Annie Brookman-Byrne, telephone: 020 7079 0777, email: abrook07@mail.bbk.ac.uk

# Appendix 2

Regression analyses for brain activation in **Chapter 6**. Age, vocabulary, and executive function measures were entered stepwise in block 1, and the relevant relational reasoning measure was entered stepwise in block 2.

**Table A.2.1.** Regression models for activation in the cluster that has its peak in the parahippocampal gyrus for the Science > Arrows contrast covaried with WASI Matrix Reasoning. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variable** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Activation in the parahippocampal gyrus cluster | | | | |
| **Model 1a** | **Constant** | | **2.35** | **.025** |
| $F(1, 31) = 7.81$, $p = .009$, $R^2 = 20\%$ | **VWM** | **.45** | **2.79** | **.009** |
| **Model 1b** | Constant | | 1.47 | .151 |
| $F(2, 30) = 6.74$, | **VWM** | **.36** | **2.31** | **.028** |
| $p = .004$, $R^2 = 31\%$, $\Delta R^2 = 10.9\%$ | **Simple Go RT cost** | **.34** | **2.18** | **.037** |
| **Model 1c** | **Constant** | | **-5.06** | **< .001** |
| $F(3, 29) = 22.41$, | VWM | .10 | 0.92 | .368 |
| $p < .001$, $R^2 = 70\%$ | Simple Go RT cost | .22 | 2.01 | .054 |
| $\Delta R^2 = 38.9\%$ | **WASI Matrix Reasoning** | **.70** | **6.12** | **< .001** |

**Table A.2.2.** Regression models for activation in the cluster that has its peak in the paracentral lobule for the Science > Arrows contrast covaried with WASI Matrix Reasoning. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variable** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Activation in the paracentral lobule | | | | |
| **Model 2a** | Constant | | **0.87** | **.389** |
| $F(1, 31) = 4.92$, $p = .034$, $R^2 = 14\%$ | **Complex Go RT cost** | **.37** | **2.22** | **.034** |
| **Model 2b** | Constant | | -0.82 | .421 |
| $F(2, 30) = 5.08$, | **Complex Go RT cost** | **.39** | **2.44** | **.021** |
| $p = .013$, $R^2 = 25\%$ | **VWM** | **.34** | **2.16** | **.039** |
| **Model 2c** | **Constant** | | **-4.47** | **< .001** |
| $F(3, 29) = 12.09$, | **Complex Go RT cost** | **.30** | **2.43** | **.022** |
| $p < .001$, $R^2 = 56\%$ | VWM | .08 | 0.62 | .543 |
| $\Delta R^2 = 30\%$ | **WASI Matrix Reasoning** | **.61** | **4.44** | **< .001** |

**Table A.2.3.** Regression models for activation in the cluster that has its peak in crus I of the cerebellar hemisphere for the Science > Arrows contrast covaried with WASI Matrix Reasoning. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variable** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Activation in crus I of cerebellar hemisphere | | | | |
| **Model 3a** | Constant | | -1.88 | .069 |
| $F(1, 31) = 10.26$, $p = .003$, $R^2 = 25\%$ | **VS-WM** | **.50** | **3.20** | **.003** |
| **Model 3b** | **Constant** | | **-4.58** | **< .001** |
| $F(2, 30) = 15.97$, | VS-WM | .24 | 1.73 | .095 |
| $p < .001$, $R^2 = 52\%$, $\Delta R^2 = 26.7\%$ | **WASI Matrix Reasoning** | **.58** | **4.07** | **< .001** |

**Table A.2.4.** Regression models for activation in the cluster that has its peak in the middle temporal gyrus for the Maths > Arrows contrast covaried with analogical reasoning. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| Dependent variable | Independent variables | β | *t* | *p* |
|---|---|---|---|---|
| Activation in the middle temporal gyrus | | | | |
| **Model 4a** $F(1, 31) = 7.60$, $p = .010$, $R^2 = 20\%$ | Constant | | -0.96 | .343 |
| | **Simple Go RT cost** | **.44** | **2.76** | **.010** |
| **Model 4b** $F(2, 30) = 17.33$, $p < .001$, $R^2 = 54\%$ $\Delta R^2 = 33.9\%$ | **Constant** | | **-4.85** | **< .001** |
| | Simple Go RT cost | .19 | 1.44 | .161 |
| | **Analogical reasoning** | **.63** | **4.68** | **< .001** |

# Appendix 3

Additional regression analyses for **Chapter 7**. These analyses were run on the subsample of participants who had data for all tasks, $N = 73$.

**Table A.3.1.** Regression models for science and maths accuracy. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variable** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science and maths accuracy | | | | |
| **Model 11a** | **Constant** | | **2.92** | **.005** |
| $F(2, 69) = 7.02$, | **Age (months)** | **.26** | **2.37** | **.021** |
| $p = .002$, $R^2 = 17\%$ | **VWM total** | **.30** | **2.69** | **.009** |
| **Model 11b** | **Constant** | | **2.20** | **.031** |
| $F(3, 68) = 6.84$, | **Age (months)** | **.22** | **2.03** | **.046** |
| $p < .001$, $R^2 = 23\%$ | **VWM total** | **.30** | **2.83** | **.006** |
| $\Delta R^2 = 6.3\%$ | **Hayling overall scaled score** | **.25** | **2.36** | **.021** |
| **Model 11c** | **Constant** | | **2.84** | **.006** |
| $F(4, 67) = 19.34$, | Age (months) | .12 | 1.34 | .184 |
| $p < .001$, $R^2 = 54\%$ | VWM total | .07 | 0.80 | .425 |
| $\Delta R^2 = 30.4\%$ | Hayling overall scaled score | .05 | 0.51 | .611 |
| | **CCF score** | **.65** | **6.63** | **< .001** |
| **Model 11d** | Constant | | 1.54 | .128 |
| $F(5, 66) = 19.77$, | Age (months) | .16 | 1.90 | .062 |
| $p < .001$, $R^2 = 60\%$ | VWM total | .03 | 0.37 | .715 |
| $\Delta R^2 = 6.4\%$ | Hayling overall scaled score | .03 | 0.31 | .760 |
| | **CCF score** | **.52** | **5.29** | **< .001** |
| | **Analogical reasoning total** | **.29** | **3.24** | **.002** |

**Table A.3.2.** Regression models for science and maths RT. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variable** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science and maths RT | | | | |
| **Model 12a** | **Constant** | | **7.35** | **< .001** |
| $F(2, 69) = 2.12$, | Age (months) | -.23 | -2.00 | .050 |
| $p = .128$, $R^2 = 5.8\%$ | VWM total | .08 | 0.66 | .515 |
| **Model 12b** | **Constant** | | **8.26** | **< .001** |
| $F(3, 68) = 4.42$, | Age (months) | -.18 | -1.60 | .115 |
| $p = .007$, $R^2 = 16\%$ | VWM total | .07 | 0.62 | .539 |
| $\Delta R^2 = 10.5\%$ | **Hayling overall scaled score** | **-.33** | **-2.92** | **.005** |

**Table A.3.3.** Regression models for science and maths accuracy by discipline. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science accuracy | | | | |
| **Model 13** | Constant | | 0.98 | .333 |
| $F(5, 66) = 13.21$, | **Age (months)** | **.22** | **2.44** | **.018** |
| $p < .001$, $R^2 = 50\%$ | VWM total | .02 | 0.20 | .840 |
| | Hayling overall scaled score | -.01 | -0.13 | .897 |
| | **CCF score** | **.47** | **4.27** | **< .001** |
| | **Analogical reasoning total** | **.25** | **2.47** | **.016** |
| Maths accuracy | | | | |
| **Model 14** | Constant | | 1.60 | .113 |
| $F(5, 66) = 14.88$, | Age (months) | .07 | 0.74 | .463 |
| $p < .001$, $R^2 = 53\%$ | VWM total | .04 | 0.41 | .683 |
| | Hayling overall scaled score | .06 | 0.65 | .521 |
| | **CCF score** | **.49** | **4.58** | **< .001** |
| | **Analogical reasoning total** | **.29** | **2.95** | **.004** |

**Table A.3.4.** Regression models for science and maths RT by discipline. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variables** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science RT | | | | |
| **Model 15** | **Constant** | | **7.00** | **< .001** |
| $F(3, 68) = 3.80$, | Age (months) | -.16 | -1.40 | .167 |
| $p = .014$, $R^2 = 14\%$ | VWM total | .06 | 0.52 | .607 |
| | **Hayling overall scaled score** | **-.32** | **-2.78** | **.007** |
| Maths RT | | | | |
| **Model 16** | **Constant** | | **8.46** | **< .001** |
| $F(3, 68) = 3.79$, | Age (months) | -.18 | -1.58 | .118 |
| $p = .014$, $R^2 = 14\%$ | VWM total | .07 | 0.64 | .552 |
| | **Hayling overall scaled score** | **-.30** | **-2.63** | **.010** |

**Table A.3.5.** Regression models for science and maths misconception accuracy. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variable** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science and maths misconception accuracy | | | | |
| **Model 17a** | **Constant** | | **-2.58** | **.012** |
| $F(3, 68) = 21.22$, | Age (months) | .12 | 1.36 | .178 |
| $p < .001$, $R^2 = 48\%$ | **Science and maths control accuracy** | **.63** | **6.66** | **< .001** |
| | VWM total | .08 | 0.89 | .377 |
| **Model 17b** | Constant | | -1.77 | .081 |
| $F(4, 67) = 23.21$, | Age (months) | .08 | 1.01 | .318 |
| $p < .001$, $R^2 = 58\%$ | **Science and maths control accuracy** | **.40** | **3.91** | **< .001** |
| $\Delta R^2 = 9.7\%$ | VWM total | .00 | .04 | .970 |
| | **CCF score** | **.41** | **3.94** | **< .001** |
| **Model 17c** | **Constant** | | **-2.43** | **.018** |
| $F(5, 66) = 21.68$, | Age (months) | .12 | 1.50 | .139 |
| $p < .001$, $R^2 = 62\%$ | **Science and maths control accuracy** | **.34** | **3.34** | **.001** |
| $\Delta R^2 = 4.1\%$ | VWM total | -.02 | -0.27 | .789 |
| | **CCF score** | **.34** | **3.29** | **.002** |
| | **Analogical reasoning total** | **.24** | **2.67** | **.010** |

**Table A.3.6.** Regression models for science and maths misconception RT. Significant predictors ($p < .05$) are highlighted in bold. β = standardised coefficients.

| **Dependent variable** | **Independent variables** | **β** | ***t*** | ***p*** |
|---|---|---|---|---|
| Science and maths misconception RT | | | | |
| **Model 18a** | **Constant** | | **4.37** | **< .001** |
| $F(3, 68) = 50.31$, | **Age (months)** | **-.17** | **-2.48** | **.016** |
| $p < .001$, $R^2 = 69\%$ | **Science and maths control RT** | **.79** | **11.48** | **< .001** |
| | VWM total | .03 | .44 | .661 |
| **Model 18b** | **Constant** | | **3.99** | **< .001** |
| $F(4, 67) = 40.77$, | **Age (months)** | **-.20** | **-2.91** | **.005** |
| $p < .001$, $R^2 = 71\%$ | **Science and maths control RT** | **.81** | **11.96** | **< .001** |
| $\Delta R^2 = 1.9\%$ | VWM total | -.03 | -0.37 | .717 |
| | **CCF score** | **.16** | **2.11** | **.038** |