Beyond number sense: Contributions of domain-general processes to the development of numeracy in early childhood

Rebecca Merkley

University College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2015

Table of Contents

ABSTRACT

Beyond number sense: Contributions of domain-general processes to the development of numeracy in early childhood
Rebecca Merkley
University College, University of Oxford
Thesis submitted for the degree of Doctor of Philosophy

A large proportion of recent research on the development of numerical cognition has focused on the foundational role of approximate number sense, yet number sense alone cannot fully explain how young children acquire numeracy skills. This thesis aims to investigate how other domain-specific processes and domain-general cognitive processes relate to numeracy in early childhood and whether they play a role in learning about mathematics. The experiments presented in Chapter 2 explored how domain-general processes relate to young children's attention to discrete number in non-symbolic representations through a correlational approach. Results supported the role for inhibitory control in selecting numerosity as the relevant stimulus dimension. In order to investigate causal relationships in emerging maths performance, Chapter 3 reports a cognitive training study aimed at contrasting transfer effects of domain-general and domain-specific training in pre-schoolers. Findings suggested caution in interpreting published transfer effects without the highest level of control. The latter chapters targeted learning mechanisms by tackling a specific process in mathematical cognition: acquiring the meaning of numerical symbols. Specifically, the experiments presented in Chapter 4 employed an artificial learning paradigm to test factors influencing adults' and children's formation of novel symbolic numerical representations. Congruency between discrete and continuous non-symbolic quantity influenced novel representations and numerical order information facilitated learning, especially in children. In order to explore symbolic representations of real numbers, Chapter 5 focuses on associations between different representational formats of real numbers in young children and how this relates to both domain-specific and domain-general factors. Children had stronger mappings between symbols and precise non-symbolic representations for numbers smaller than four, than between larger numbers and approximate non-symbolic representations. Taken together, results from the experiments presented in this thesis highlight the need to incorporate factors beyond number sense in theories of numeracy development.

Acknowledgements

Firstly, I would like to thank the Department of Experimental Psychology for being an inspiring environment in which to pursue doctoral studies. I have greatly enjoyed laughing and snacking with the members of the ABCD group and office D152: Andria Shimi, Stefano Belli, Hannah Hobson, Chii Fen Hiu, Lauren Heathcote, Brianna Doherty, Simone Haller, Pawel Matusz, and Eira Formicheva. I would also like to thank Emily Connally and Bea Krause for motivation during the writing phase. Furthermore, I have benefitted from Dorothy Bishop's supervision as well as from feedback from the Developmental Science group as a whole. I would also like to thank Ann Dowker and Lucy Cragg for useful comments on the thesis.

My research would not have been possible without the financial support of both the Clarendon Fund and Canada's Natural Sciences and Engineering Research Council. Additionally, I am indebted to the schools, nurseries, teachers, carers, parents, and children who participated in this research. Special thanks go to Luca and Leila, who were invaluable pilot participants as well as case studies of preschool cognitive development.

Science friends and colleagues based outside of Oxford, and particularly current and former members of the Numerical Cognition Lab at Western, have been great sources of inspiration and support. It was Daniel Ansari who encouraged me to apply to graduate programs when I finished my undergraduate degree with no plans for the future other than a desire to travel. I would also like to thank Stephanie Bugden and Eric Wilkey for insightful comments on drafts of chapters of this thesis.

I had the privilege of being a member of University College and have developed friendships with students from varied countries and fields of study who have expanded my worldview. Additionally, it was a pleasure playing with the ladies of the Foxes Women's Football Club (Cuppers Champions '15!) during my time at Oxford. I am also grateful to my friends outside of Oxford and the fact that whenever we are reunited it does not feel as though any time or distance has come between us. Years ago, I promised Katy Sluzar I would acknowledge her in scientific writing of some kind in exchange for reciprocal acknowledgment in a book she planned to write titled 'Dating is so 90's'. This seems like the opportune place to fulfil that promise. I would also like to thank here all those who supported me throughout any part of this process.

My family is a perpetual source of support and reassurance and I would not have pursued this degree if not for them. My mum inspired a love of learning in me from an early age, my dad has encouraged me to be more confident, and my sister, Vickie, believed in me.

Finally, I would like to thank Gaia Scerif for being the very best supervisor I could have hoped for as well as an incredible role model. Her innovative approach to research has guided my doctoral work and her passion for science is infectious. Gaia once told me that it is better to laugh than to worry and I am ceaselessly impressed by her unwaveringly positive attitude. I am also grateful to her for sharing her seemingly endless supply of tea and chocolate.

Publications Arising from this Thesis

Portions of this thesis appear in the following publications:

Merkley, R. & Scerif, G. (2015). Continuous visual properties of number influence the formation of novel symbolic representations. *Quarterly Journal of Experimental Psychology*.

Merkley, R., Shimi, A., & Scerif, G. (under review). Electrophysiological markers of newly acquired symbolic numerical representations: The role of magnitude and ordinal information. *ZDM Mathematics Education*.

CHAPTER 1: GENERAL INTRODUCTION

1.1. Foundations of Numeracy

There are three types of people: Those who can count and those who cannot. Preschoolers tend to be of the latter type, and therefore would likely fail to grasp the humour in that statement. Children learn the cardinality of the number words between the ages of 2 and 4-years (Wynn, 1992), and thus prior to the start of formal education. The pervasive view on how the meanings of numerical symbols are acquired is illustrated by this quote from Dehaene (2008, p.552), “when we learn number symbols, we simply attach their arbitrary shapes to the relevant non-symbolic quantity representations”. However, the cognitive mechanisms underlying this learning are not well understood and this process does not seem to be as simple as Dehaene posited. This thesis aims to address the development of numeracy skills, particularly learning numerical symbols, in early childhood and explore contributions from domain-specific as well as domain-general cognitive factors.

1.1.1. The approximate number system

In recent years, a large proportion of research into the foundations of numeracy has focused on the role of approximate number system (ANS) acuity, also termed ‘number sense’. The ANS is a pre-verbal system for numerical quantification and supposedly represents numbers as approximate analog magnitudes that activate imprecise, overlapping distributions along an internal ‘mental number line’ (see Dehaene, 2011 for review). This idea is supported by research in non-human primates showing that neurons selectively fire for numerosities but that there is some overlap in numerosities to which a neuron responds (see Nieder & Dehaene, 2009 for review). For example, a neuron that fires in response to seeing five objects, may also respond to four and six objects, but to no other numerosities beyond that. The width of the distributions, i.e. the selectivity with which neurons respond to a given numerosity, characterizes the precision of one's ANS. ANS precision is typically assessed using numerical magnitude comparison tasks in which participants are asked to choose the numerically larger of two dot arrays (non-symbolic comparison task) or numerically larger of two Arabic digits (symbolic comparison). Performance on tasks measuring the ANS is consistent with Weber's law, which states that the ability to discriminate between two stimuli is proportional to their magnitude (Moyer & Landauer, 1967). For example, numerical comparison task performance is ratio-dependent: Response times and error rates increase as the ratio between the to-be-compared numerosities increases. For example, it is quicker and easier to decide that 8 is larger than 4 (ratio of .5) than that 8 is larger than 6 (ratio of .75). These characteristic ratio effects are seen when comparing both symbolic and non-symbolic magnitudes (Dehaene, 2011), and this has been taken to suggest that number is represented in a format-independent way. Evidence from neuroimaging has also supported the idea that number representation in the brain is format-independent (e.g. Dehaene, Dehaene-Lambertz, & Cohen, 1998), but this remains controversial (e.g. Ansari, 2008; Cohen Kadosh & Walsh, 2009), a controversy to which I return, extensively, throughout the thesis. Based on these findings, it has been proposed that numerical symbols are grounded in ANS representations (see Piazza, 2010 for a review).

The ANS has been linked to numerical symbol acquisition as well as to mathematics achievement more broadly. Individual differences in 14-year olds' number sense, as measured by a non-symbolic magnitude comparison task, were found to be correlated with past scores on standardized measures of mathematics achievement as far back as the first year of formal education (Halberda, Mazzocco, & Feigenson, 2008). Furthermore, preschool children's ANS acuity predicted math achievement at 6 years of age (Mazzocco, Feigenson, & Halberda, 2011), and, infants' number sense predicted performance on a standardized measure of early numeracy at 3.5 years of age (Starr, Libertus, & Brannon, 2013). It has therefore been argued based on these correlational findings that pre-verbal number sense is the basis upon which mathematical thinking is built. This has led to the development and testing of interventions to increase ANS acuity, with the goal of transferring to improvements in mathematical achievement more generally (e.g. Hyde, Khanum, & Spelke, 2014; Park & Brannon, 2013; Wilson, Revkin, Cohen, Cohen, & Dehaene, 2006). However, a systematic review revealed that the evidence for the relationship between non-symbolic comparison performance and mathematics achievement was mixed, whereas there is more consistent evidence for a significant relationship between symbolic magnitude comparison and mathematics achievement (De Smedt, Noël, Gilmore, & Ansari, 2013). This suggests that the relationship between the ANS and mathematics achievement may not be as direct as previously thought, and therefore that training non-symbolic comparison may not be the most effective intervention route.

Two separate lines of evidence have recently further called for the revision of ANS theories. First, extracting numerosity from non-symbolic arrays has been shown to be less straightforward and less automatic as was originally proposed (Gebuis & Reynvoet, 2012a, 2012b). Specifically, the visual perceptual features associated with continuous quantity influence estimation and comparison of numerosities, regardless of how they are controlled for (Gebuis & Reynvoet, 2012a, 2012b; Szűcs, Nobes, Devine, Gabriel, & Gebuis, 2013). Second, the extent to which symbolic representations are grounded in corresponding non-symbolic quantities has been questioned (Lyons, Ansari, & Beilock, 2012). Lyons and colleagues (2012) proposed the symbolic estrangement hypothesis: Symbolic numerical processing does not entail automatically activating an approximate representation of the corresponding quantity, but rather is dependent on the ordinal relations between the symbols themselves. This hypothesis has been corroborated by recent neuroimaging findings emphasizing differences in the neural underpinnings of symbolic compared to non-symbolic representations of number (Bulthé, De Smedt, & Op de Beeck, 2014, 2015; Lyons, Ansari, & Beilock, 2014). Taken together, these lines of research highlight the need to consider factors beyond ANS acuity in the study of mathematical development.

1.1.2. Beyond the ANS

While the ANS has had a major influence on the field of numerical cognition, it is but one of many factors that have been recognized as playing a role in the development of mathematical competency. Knowledge of numerical symbols and counting principles is another domain-specific predictor of mathematics achievement (e.g. Göbel, Watson, Lervåg, & Hulme, 2014). Gelman and Gallistel (1978) identified five counting principles. 1) The one-to-one principle is the understanding that each object in an array should be assigned only one number word. 2) The stable-order principle specifies that number words must be assigned in order and not at random. 3) The cardinal principle refers to the understanding that the last number word used when counting a set indicated the cardinality, or number of objects in the set. 4) The abstraction principle highlights the fact that counting can be applied to any objects, actions, or sounds. 5) The order irrelevance principle describes the notion that the number of objects in a set is the same regardless of the order or direction in which they are counted. Children learn these principles between the ages of two and four years, over the preschool period and there are individual differences in the rate at which separable abilities, and in particular understanding of the order-irrelevance principle, develop (Dowker, 2008). Therefore, even prior to the start of formal education, mathematics abilities are multi-componential, i.e. composed of separable components.

In addition to cognitive skills that are specific to mathematics, there are other cognitive skills that are important for learning and achievement more generally that are especially related to mathematics achievement. Domain-general predictors of mathematics achievement include logical thinking ability (Piaget, 1952), intelligence, particularly spatial intelligence (e.g. Bull, Espy, Wiebe, Sheffield, & Nelson, 2011), linguistic factors (e.g. LeFevre et al., 2010), and executive functions (EFs), cognitive processes responsible for controlling information processing and action, (e.g. Bull & Scerif, 2001). As previous studies have revealed that EFs are related to mathematics achievement over and above measures of verbal intelligence (IQ) (e.g. Fuhs & McNeil, 2013) and crystallized knowledge (e.g. Bull et al., 2011), an increasing amount of research has focused on relationships between EFs and mathematics achievement. I expand on the evidence for these relationships in section 1.3 below, but first provide background on the definition of EFs in order to frame the interpretation of this body of work.

1.2. Structure of Executive Functions

Executive functions have been likened to a chief executive officer (CEO), as they are responsible for directing and monitoring all other cognitive processes (e.g. Goldberg, 2002). Furthermore, the idea of executive control, or a "central executive" is a crucial component of prominent models of memory (Baddeley, 2000) and attention (Posner & Petersen, 1990). There is some consensus that executive function consists of several domains, including complex reasoning and problem solving, working memory, attentional control, cognitive flexibility, self-monitoring, and regulation of cognition, emotion, and behaviour (e.g. Diamond & Lee, 2011). However, there have been many challenges for defining and measuring EFs. As such there currently is no universally accepted model. For the purpose of designing and interpreting the research presented in this thesis, I used Miyake and colleagues' (2000) model of EFs, which consists of three processes: updating, shifting, and inhibition. *Shifting* refers to flexibility of attention and ability to switch between tasks, *updating* is short for monitoring and updating information in working memory, and *inhibition* is defined as the ability to inhibit automatic responses. Miyake and colleagues (2000) administered a battery of executive function assessments to a large sample of adults and, using confirmatory factor analysis, demonstrated that these processes were distinguishable from each other but still moderately correlated.

My rationale for using the model by Miyake et al (2000) as a framework is manifold. This model of EFs has been tested across development using factor analysis, making it a useful starting point for an investigation of domain-general influences on the foundations of numeracy. Intriguingly, cross-sectional designs and results have been mixed, depending on the age group and tasks used to measure EFs. A study in 8 to 13-year-old children found results that fit with Miyake and colleagues' (2000) model (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003), but executive functioning in younger (3-5-year-old) children seems to be best explained by a single underlying factor (Hughes & Ensor, 2011; Wiebe et al., 2011). However, there is great improvement in EFs between the ages of three and five, and some evidence suggests that by the end of this period shifting, updating, and inhibition can be distinguished from each other (Garon, Bryson, & Smith, 2008). Furthermore, there is some divergence in developmental trajectories of these processes (e.g. Huizinga, Dolan, & Van der Molen, 2006). Taken together, these findings highlight how EFs change over development and therefore likely operate differently in interaction with other cognitive processes at different times over the course of development. With that in mind, I now turn to the relationships between EFs and mathematics achievement over preschool and primary school.

1.3. Relationships Between Executive Functions and Numeracy

An extensive body of research has linked EFs and mathematics both concurrently and longitudinally in school aged-children (Best, Miller, & Naglieri, 2011; Bull & Scerif, 2001; Lefevre et al., 2013; St Clair-Thompson & Gathercole, 2006; Van der Ven, Kroesbergen, Boom, & Leseman, 2012; Yeniad, Malda, Mesman, van IJzendoorn, & Pieper, 2013) as well as in preschoolers, (Bull et al., 2011; Bull, Espy, & Wiebe, 2008; Clark, Pritchard, & Woodward, 2010; Clark, Sheffield, Wiebe & Espy, 2013; Steele, Karmiloff-Smith, Cornish, & Scerif, 2012). In a recent systematic review of this literature, Cragg and Gilmore (2014) proposed a theoretical model of relationships between separable components of EFs and mathematical skills across development. The authors posited that working memory, or updating, may be important for acquiring new numerical facts as well as for holding numerical information in mind while executing maths procedures. Shifting may be necessary for switching between procedures when tackling complex maths problems as well as for directing attention to relevant information when learning new concepts.

The relationships between inhibition and different aspects of maths may be especially dependent on age, as inhibition may be useful for suppressing previously learned facts and strategies in favour of flexibly approaching more complex problems and acquiring new strategies. This model provides useful hypotheses that remain to be further empirically tested

While it is clear that both mathematics and EFs are multi-componential, surprisingly little research has incorporated both domain-specific and domain-general factors to investigate early numeracy development in a comprehensive fashion. Specifically, researchers have tended to include measures of multiple executive processes and a global measure of math achievement (e.g. Van der Ven et al., 2012), or multiple domain-specific measures and one executive measure (e.g. Xenidou-Dervou, De Smedt, van der Schoot, & van Lieshout, 2013). In one of the few studies modelling relationships between multiple domain-general and domain-specific contributors to the development of numeracy in young children, 4 and 5-year-old children completed assessments of linguistic skills, numeracy skills, and spatial attention, and completed assessments of mathematics achievement two years later (LeFevre et al., 2010). Results supported hypothesized relationships between linguistic skills and symbolic mathematics and early quantitative skills and numerical magnitude processes. Spatial attention was related to both symbolic and non-symbolic skills concurrently and longitudinally. These results again highlight the importance of looking at differential relationships across separate components of mathematics outcomes. In a related vein, a longitudinal study found that executive functions and classroom attention mediated the relationship between kindergarten number sense and mathematics outcomes a year later (Hassinger-Das, Jordan, Glutting, Irwin, & Dyson, 2014). This suggests that domain-general processes are predictive of mathematics achievement longitudinally even over and above domain-general number sense.

While correlational evidence, both concurrent and longitudinal, supports the importance of EFs for performing numerical operations, it remains unclear whether and how EFs matter for learning numerical skills. One study found that executive control predicted growth in arithmetic fluency in 8-10-year-old children (LeFevre et al., 2013), which suggests that executive control played a role in the acquisition of these skills. However, this finding does not shed any light on potential causal mechanisms. Similarly, Clark et al. (2013) found that preschool executive control at 3-years of age predicted performance on mathematics assessments at 5-years of age, but acknowledged that these results could not discern mechanisms. It is likely that executive control is necessary for effectively deploying knowledge when performing mathematics, but it could additionally be that executive control plays an essential role in the acquisition of basic numerical knowledge. As executive control directs and regulates attention and memory to information relevant to learning (Posner & Rothbart, 2007), it, theoretically, may be particularly relevant to educational achievement, but there is little empirical evidence supporting this hypothesis.

Testing the influence of executive control on *learning* requires experimental intervention, for example, through designing and evaluating a preschool curriculum (Diamond, Barnett, Thomas, & Munro, 2008). One type of intervention is cognitive training, which usually takes the form of a computerized regime designed to improve updating or another executive function. Cognitive training has been used to investigate causal mechanisms underlying relationships between EFs and numeracy by testing whether improvements in executive control transfer to mathematics outcomes (e.g. Goldin et al., 2014; Holmes, Gathercole, & Dunning, 2009; Kroesbergen, van't Noordende, & Kolkman, 2014; Witt, 2011), but the evidence for transfer is inconsistent (e.g. Melby-Lervåg & Hulme, 2012). I return to the use of cognitive training as an experimental tool to investigate causality in the relationships between EFs and early numeracy in Chapter 3.

Teacher's classroom observations support a role for EFs in mathematics education. A qualitative study of teachers' perceptions of the role of EFs in mathematics found that most teachers believed that executive function skills were important for learning maths despite the fact that only 20% of the participating teachers were familiar with the term executive functions (Gilmore & Cragg, 2014). Teachers rated the EF skills almost as highly as the maths specific skills. Number of years of teaching experience was positively correlated with the ratings of EFs, which suggests that it may take a few years for teachers to realize these skills are important for maths. Somewhat surprisingly, shifting and inhibition were rated higher than updating skills, which empirical studies have found may be most strongly related to maths success (e.g. Alloway & Alloway, 2010; Bull & Scerif, 2001; Bull et al., 2008; Bull & Lee, 2014). This highlights discrepancies between findings from controlled cognitive psychology studies and what can be gleaned from more qualitative applied research. In turn, it illustrates the importance of considering the context in which learning occurs.

To summarize, executive functions have been linked to multiple components of mathematical achievement primarily through evidence from correlational studies. Therefore, the underlying mechanisms are not well understood (Blair, Knipe, & Gamson, 2008) and future research is required to test specific hypotheses pertaining to the role that separable executive functions play in different aspects of mathematics achievement. As relationships between EFs and domain-specific numeracy skills are not stable across development, a more nuanced understanding of the mechanisms underlying these relationships requires disentangling processes that are important during specific periods in development. I return to this particular question and approach in Chapter 2. Furthermore, investigating whether executive control plays a role in the acquisition of basic numerical competencies necessitates an experimental approach isolating influences of executive control on learning itself.

1.4. The Role of Executive Functions in Emerging Numeracy in Preschool

The preschool stage is of particular interest as it precedes the start of formal education and it is during this time that the foundations of numeracy are laid.

Two related questions pertaining to this developmental stage are, I believe, of particular importance. How do children learn to attend to the abstract property of number when processing non-symbolic arrays? Additionally, how do children acquire the meaning of numerical symbols? Previous research into these questions has been primarily domain-specific and centred around the ANS. However, as discussed in section 1.1.1., the ANS alone cannot account for the processing of non-symbolic representations of number and the acquisition of numerical symbols. Considering the executive demands of these two problems in addition to the domain-specific ones could shed more light on the cognitive mechanisms underlying this learning. I expand on the existing theories and supporting evidence below, with a focus on hypothesized roles of executive control.

1.4.1. Attention to non-symbolic number

In contrast to the idea that humans are born with an intuitive sense of number (e.g. Butterworth, 2005; Starr et al., 2013), a systematic review concluded that infants' looking habits are swayed by continuous quantity in addition to discrete numerosity and, therefore, that the influence of this variable should be studied, rather than controlled (Cantrell & Smith, 2013). A previous review of the infant literature reached the same conclusion (Mix, Huttenlocher, & Levine, 2002), yet whether infants' quantification is truly based on number remains hotly debated. Similarly, Leibovich and Henik (2013) argued that the role of perceptual features of continuous quantity in non-symbolic magnitude processing should be considered, rather than controlled for, again, in the development of numerical competencies. They proposed an alternative developmental model of non-symbolic processing in which infants first make judgments based on continuous properties and then, with experience and education, eventually learn to make comparisons based on discrete number. Taken together, these alternative proposals emphasize how children must learn to attend to discrete number either in correlation to or in contrast with continuous quantity.

Non-symbolic stimuli in numerical comparison and estimation tasks for participants of all ages are usually controlled for various perceptual features, such as total surface area, in order to ensure that participants are responding to number (e.g. Halberda et al., 2008; Starr et al., 2013). However, controlling stimuli in this way often puts the size of individual dots in direct conflict with the number of dots in the array. Commonly used measures of numerical cognition therefore require inhibition of non-numerical dimensions (Clayton & Gilmore, 2014; Gilmore et al., 2013), a process that may be particularly challenging when the ability to select these representations is not as well established as it is later in childhood or adulthood. Indeed, preschoolers are influenced by perceptual variables that vary with numerosity when making magnitude judgments (Negen & Sarnecka, 2014; Rousselle & Noël, 2008; Rousselle, Palmers, & Noël, 2004). Specifically, children performed above chance on magnitude comparison when density and contour were controlled for, but not when area was controlled for and thus conflicted with number (Rousselle et al., 2004). Based on these findings, the authors argued that, under some conditions, young children base their magnitude judgments on stimulus dimensions other than discrete number, and suggested that this reflects immature numerical competence.

Adults are less susceptible to the conflict between continuous and discrete quantity on magnitude comparisons than children are (Szűcs et al., 2013) which suggests that having more experience with number may facilitate the inhibition of irrelevant cues from continuous quantity in favour of attending to discrete numerosity. Furthermore, a recent study found that preschoolers who had not yet acquired the cardinality principle failed to base their magnitude judgments on discrete numerosity when it conflicted with total surface area (Negen & Sarnecka, 2014). In other words, children who had not yet learned the cardinality principle, and therefore had immature knowledge of number words and counting, did not inhibit perceptual cues associated with continuous quantity in order to attend to discrete number. In a related vein, 2-4-year-old children who had not yet learned the cardinality principle failed to match sets of objects based on their numerosity, but were able to match based on colour or shape (Slusser & Sarnecka, 2012). This suggests that it is only once children understand what “number” is, as evidenced by knowledge of the cardinality principle, that they may use this to exert top-down attentional control in order to attend to the number of items in a set.

The set size of non-symbolic arrays also influences attention to non-symbolic number, in terms of whether comparisons are made on the basis of discrete number or continuous quantity (Cantrell, Kuwabara, & Smith, 2014).

Specifically, preschoolers were more likely to base magnitude judgments on discrete number for sets smaller than four than for larger sets. The authors suggested that small sets are more easily viewed as sets of individual objects, whereas larger sets may be seen as continuous wholes. These findings are convergent with other recent work suggesting that different dimensions of non-symbolic sets shift in saliency as a function of set size (e.g. Clayton & Gilmore, 2015). This points to a potential contribution of domain-general cognitive factors as constraints of visual attention may account for the discrepancy in magnitude judgments across small and large set sizes. In particular, four is the maximum number of objects that can be attended to in parallel (Trick & Pylylshyn, 1994), and enumerating small versus large sets of objects show distinctive neural correlates, which reflects differences in allocating attention across sets (Ansari, Lyons, van Eimeren, & Xu, 2007; Demeyere, Rotshtein, & Humphreys, 2014; Sathian et al., 1999). This process of rapid and exact enumeration of small sets is known as subitizing (Kaufman, Lord, Reese, Volkmann, 1949).

Of note, subitizing can be employed when one is instructed to enumerate. A complementary mechanism driving attention to non-symbolic number is spontaneous focusing on numerosity (SFON): Hannula and Lehtinen (2005) proposed that SFON, measured in the absence of instruction, is related to mathematical skills. In other words, while children can attend to numerosity when directed to, what may be of more importance for mathematical development is their tendency to pay attention to the numerical dimension of stimuli without being prompted to do so. SFON is typically assessed with an imitation task in which children are instructed to, for example, feed a puppet in the same way as the experimenter. They are considered as having focused on number if they fed the puppet the same number of items as the experimenter did. Results from a longitudinal study showed that SFON in kindergarten was indeed a significant predictor of mathematical, but not reading ability, two years later (Hannula, Lepola, & Lehtinen, 2010).

In sum, contrary to previous theories that the ANS quantifies non-symbolic representations based on discrete number, the extent to which numerosity is the attended-to dimension of non-symbolic arrays is dependent on factors including age, experience, set size, congruency, and task instructions. It therefore seems that children must learn to attend to the dimension of discrete numerosity in non-symbolic arrays. Executive control may play a role in this learning, as inhibitory control is necessary for inhibiting irrelevant cues associated with continuous quantity, if the cues are not congruent with the number of items. However, the direction of the relationship between attending to discrete number and learning the cardinality of numerical symbols remains unclear. It could be that once children learn the cardinality principle, they develop a better attentional template for number and use this to direct their attention (a top-down attentional control mechanism). Alternatively, it could be that certain properties of non-symbolic arrays increase the saliency of and thus facilitate attention to discrete number (a bottom-up attentional control mechanism), which in turn influences the development of understanding of numerical symbols. I explore manipulations geared to study factors influencing attention to non-symbolic number in Chapter 4.

1.4.2. Learning numerical symbols

It is thought that numerical symbols are mapped onto corresponding representations of approximate non-symbolic quantity (e.g. Dehaene, 2008). However, the evidence supporting this assumption is mixed and the way in which non-symbolic and symbolic representations become associated with one another over development is not well established. Adults' symbolic representations are not tightly linked to corresponding approximate representations (Lyons et al., 2012), but this finding does not rule out the possibility that the ANS is important for the acquisition of numerical symbols and therefore that they may be linked earlier in development. Children learn the cardinality of the number words 'one' to 'four' slowly and sequentially between the ages of 2 and 4-years (Wynn, 1992), but then learn the larger numbers more quickly. Thus, it is necessary to investigate children's associations between symbolic and non-symbolic representations of number from a very early age in order to glean insight into mechanisms underlying the acquisition of numerical symbols.

One study of number word knowledge in 2-5-year-old children found that children were only able to map number words to approximate non-symbolic quantities after learning the cardinality principle, and thus the exact meaning of number words (Le Corre & Carey, 2007). This suggests that the ANS cannot be involved in the acquisition process as number word knowledge is a prerequisite for attaching words to ANS representations. Specifically, Le Corre & Carey (2007) investigated children's numerical estimation abilities and found that only a subset of children showed scalar variability in their estimates of quantities larger than four (i.e. the amount of variance increased with the size of the estimate – a signature of the ANS). The children who did demonstrate this estimation ability had acquired the cardinality principle, but not all children who knew the principle showed scalar variability in their estimates, suggesting that this mapping occurs a few months following the acquisition of the cardinality principle. However, all participating children were able to estimate the numerosity of sets up to four, regardless of cardinality knowledge. Furthermore, the variability of their estimates of sets up to four was not scalar, suggesting that they did not rely on the ANS (Le Corre & Carey, 2007). As discussed in the previous section, small sets of up to four objects can be subitized (Kaufman et. al, 1949) and therefore represented precisely rather than approximately. This suggests that number words one to four are mapped onto exact rather than approximate non-symbolic representations of quantity, and thus that the ANS is not involved in this acquisition process. As young children learn the count sequence by rote before they understand the cardinality of each number word (Wynn, 1992), Carey (2009) proposed that the words act as placeholders. By this account, children attach words one to four to exact non-symbolic representations, and subsequently align larger number words with representations of non-symbolic quantities once they realize the common order. This suggests that ordinal information is also important in the acquisition process (e.g. Wiese, 2003) and highlights the need to consider factors beyond the ANS.

Altogether, it is as yet unclear how children acquire the meaning of numerical symbols, which happens prior to the onset of formal education and is an important predictor of subsequent math achievement (e.g. Göbel et al., 2014). On the one hand, if the assumption that numerical symbols are grounded in corresponding non-symbolic representations is correct, then it is necessary to study the influence of continuous quantity on acquisition as it has been established that non-symbolic processing is influenced by continuous quantity. On the other, if the ANS is not the basis of, or at least not the sole basis of symbol acquisition, then other relevant processes such as subitizing, numerical ordering and executive functions also require attention. This points to the need to explore multiple factors pertaining to symbol acquisition and the possible role of executive control in directing attention to numerical information most relevant to learning the meaning of numerical symbols.

1.5. Aims of the Thesis

The overarching goal of this thesis is to elucidate mechanisms underlying the observed relationships between EFs and numeracy skills in early childhood. Specifically, the question motivating my doctoral work is: Does executive control matter for *learning* the foundations of mathematics? This can be further broken down into pursuing the two learning problems elaborated in section 1.4: Learning to attend to the abstract property of number and learning the numerical meaning of seemingly arbitrary symbols.

Studying conceptual change and *learning* as opposed to studying performance *online*, i.e. as it happens, has proven challenging and complementary approaches were taken in order to investigate this question. Firstly, correlational designs were employed to study the interplay between domain-general and domain-specific numeracy processes in young children (Chapter 2). Secondly, an intensive environmental manipulation was used to assess malleability and transfer of these processes in preschoolers (Chapter 3). Thirdly, an artificial learning paradigm was employed in order to test the influence of various factors of interest on the formation of novel symbolic representations (Chapter 4). Finally, the mapping between distinct representational formats of number was investigated in young children who had recently acquired numerical symbols (Chapter 5). While these latter experiments, presented in chapters 4 and 5, were conducted on adults and children in primary school, they were designed with the goal of probing the problem of how young children learn the meaning of numerical symbols, which typically occurs during the preschool period. Overall, the data presented in this thesis provide novel contributions to understanding the role of factors, beyond number sense, in early learning about number.

In addition to gleaning insight into cognitive mechanisms, I aim to conduct research with potential applications for education. My thinking about whether and how executive control matters for learning about number has benefitted from conversations with teachers and carers, classroom observations, and anecdotes from parents of young children. Moreover, my work has been shaped by my participation in conferences of educational neuroscience, an emerging field with the aim of applying research from cognitive science to educational practice. As knowledge of numerical symbols has been shown to be a significant predictor of growth in subsequent mathematical skills (e.g. Göbel et al., 2014; Vanbinst, Ghesquière, & De Smedt, 2014), it is possible that gaining a better understanding of how this learning occurs could lead to developing ways of facilitating this process in preschool. However, making applications is complicated by the fact that controlled psychology experiments do not easily transfer to the chaotic environments of real-life preschool classrooms. The tension between theoretical and applied goals of research is a point to which I return when discussing the implications of my findings.

CHAPTER 2: INFLUENCES OF INHIBITORY CONTROL ON EARLY NUMERACY

2.1. Introduction

As outlined in Chapter 1, both executive control and numeracy are multi-componential, and it is therefore necessary to disentangle the nature of the observed relationships between executive control and mathematics more broadly by pinpointing specific processes that relate to one another across domains. One way to investigate these relationships is to target particular executive processes hypothesized to play a role in early maths achievement and use experimental manipulations to isolate executive demands of numerical tasks. The experiments described in this chapter focus on preschoolers' inhibitory control in particular because this has been postulated to relate to emerging mathematical abilities (e.g., Houdé, 2000) and, in particular, is important for non-symbolic magnitude comparison (Gilmore et al., 2013; Clayton & Gilmore, 2015). An Animal Size Stroop task used previously to measure inhibitory control in adults and 5-8-year old children was modified to be appropriate for 3-6-year-old children (Bryce, Szũcs, Soltész, & Whitebread, 2011; Szűcs, Soltész, Bryce, & Whitebread, 2009). In Experiment 1, we used this task to investigate relationships between inhibitory control and early numeracy skills, including non-symbolic magnitude comparison. To further explore these relationships, in Experiment 2, we manipulated the inhibitory demands of a non-symbolic comparison task and compared performance on this task and the Animal Size Stroop task across young children in preschool and the first two years of formal education. The development of inhibitory control in relation to non-symbolic comparison is of particular interest given the hypothesis elaborated in the Chapter 1 that inhibitory control plays a role in learning to attend to discrete numerosity in non-symbolic arrays.

2.1.1. The role of inhibitory control in early numeracy

Houdé (2000) argued that many of Piaget's conclusions that the errors young children make are a result of immature concept knowledge or reasoning skills, both in general and in the context of numerical concepts, could instead be interpreted as a result of failures of inhibition. For example, failure on a conservation of number task may reflect a failure to inhibit the idea that length equals number rather than a lack of ability to perform calculations of number. This suggests that cognitive development in the domain of number representation may mean learning to ignore competing task-irrelevant dimensions of the stimuli. Preschoolers from low-income families' performance on a magnitude comparison task was found to be related to mathematics achievement, but this effect was driven by trials that required inhibiting an irrelevant stimulus dimension (surface area) to select the larger numerosity (Fuhs & McNeil, 2013). This lends support to the hypothesis that the ability to ignore irrelevant perceptual information and focus on number may explain why inhibitory control relates to early numeracy. Furthermore, 3 to 5-year-olds performed better on a number matching task when given audio and visual information about number compared with when given information in just one modality, suggesting that highlighting the amodal properties of number, as in the audiovisual condition, benefitted them (Jordan & Baker, 2010). Gilmore and colleagues (2013) found complementary results to Fuhs & McNeil (2013) in older children and argued that the relationship between non-symbolic magnitude comparison performance and math achievement can be fully accounted for by individual differences in inhibitory control.

There is therefore a likely precise cognitive reason why inhibitory control measures and magnitude comparison relate to each other concurrently. As detailed in Chapter 1, the ANS is thought to allow for the representation of the approximate number of items in an array (Dehaene, 2011). In order to be sure that participants make judgments based on the number of stimuli in an array, and not on related features of continuous quantity, visual properties are typically controlled for in non-symbolic arrays in numerical tasks (e.g. Halberda et al., 2008). This essentially leads to magnitude comparison tasks that heavily rely on inhibitory control, as on some trials number and continuous quantity are congruent whereas on others they are in conflict (e.g. Clayton & Gilmore, 2015). For example, Clayton and Gilmore (2015) systematically manipulated the inhibitory demands of a non-symbolic comparison task in order to investigate which factors influenced 7-9-year-old children's judgments. Specifically, they contrasted comparisons when discrete and continuous quantity were congruent with comparisons when discrete and continuous quantity were incongruent. Results revealed that both ANS acuity and inhibitory control influenced accuracy on the task, with incongruent information associated with continuous quantity being more salient for larger numerosities than for smaller sets.

A similar non-symbolic numerical Stroop paradigm was used to test the interference of conflicting area information on number judgments and the reciprocal interference of number on area judgments in four-to-six-year-old children (Rousselle & Noël, 2008). Participants also performed a Day/Night Stroop task, an established measure of inhibition in young children (Gerstadt, Hong, & Diamond, 1994). Results showed significant effects of congruity in both the area and numerosity judgment tasks. The size of the congruity effect of number on the area task increased with age, which supports the hypothesis that discrete numerosity becomes a more salient feature of the environment later in childhood (Mix et al., 2002; Leibovich & Henik, 2013). Number and other stimulus dimensions, such as area, usually covary in the environment. Therefore, discriminating continuous quantity typically leads to the same results as discriminating numerosity, so it could be that young children do not learn to distinguish continuous and discrete quantity until relatively late in childhood (Mix et al., 2002). Interference effects on the magnitude judgment tasks were not significantly correlated with interference effects on the Day/Night Stroop, which the authors took to suggest that non-numerical and numerical magnitude processing develop independently of inhibition. However, five and six-year olds' accuracy on the Day-Night task was quite high, suggesting potential ceiling effects and therefore reduced individual differences on that measure. Even when it was first introduced, Gerstadt and colleagues (1994) found that the Day-Night Stroop was quite difficult for children three-and-a-half to four-and-a-half years old but that it was comparably easy for six and seven-year-olds. It could be that in the study by Rousselle and Noël (2008), this task was not a sensitive measure of inhibition in the majority of the sample and therefore was not useful for investigating the association between interference on magnitude comparisons and the development of inhibitory control.

2.1.2. Inhibitory control and number sense across development

In summary, then, a working hypothesis is that inhibitory tasks requiring interference control, the process of suppressing a stimulus, or dimension of a stimulus, that requires a competing response to that required of the task instructions (Nigg, 2000), relate to early numeracy because early numerical tasks demand the ability to inhibit related non-numerical dimensions of stimuli. Indeed, there is an increasing body of evidence supporting the hypothesis that, even in adults, specific tasks that are traditionally construed as measurements of “number sense” (e.g., judgments of non-symbolic magnitudes) are influenced by stimulus dimensions such as area and density, regardless of how these parameters are controlled (Gebuis & Reynvoet, 2012a; Szűcs, Nobes, Devine, Gabriel, & Gebuis, 2013). Gebuis and Reynvoet (2012) manipulated the average diameter of the dots, the convex hull (smallest contour around the dot area), and density (average diameter divided by convex hull) of dot arrays so that each feature was not correlated with increasing numerosity and had adults perform a numerical estimation task on these stimuli. The size of visual cues influenced estimates despite the fact that the cues separately did not correlate with numerosity, supporting the authors’ hypothesis that information about numerosity is obtained by combining multiple visual cues. Furthermore, adults are susceptible to interference from area on a numerical Stroop task (Hurewitz, Gelman, & Schnitzer, 2006). Therefore, numerical competence may reduce, but not eliminate entirely, the requirement to inhibit perceptual dimensions of stimuli when they conflict with numerosity.

The relationship between size and numerical magnitude was further illustrated by a study that used a size prime on a numerical magnitude judgment task (Gabay, Leibovich, Henik, & Gronau, 2013). Adults saw a prime animal image and the size of the image remained constant but the real-world size of the animal was either small or large. The prime was followed by an Arabic digit and participants had to judge whether the digit was small or large. Results showed that participants' judgments of numerical magnitude were influenced by the conceptual size of the prime and reaction time was faster when the size of the target numerical value was congruent with the size of the prime. The authors argued that this suggests continuous and discrete magnitudes are processed in highly similar ways. This is one possible explanation for why choosing the larger of two non-symbolic arrays is easier when discrete and continuous magnitudes converge, and why inhibitory control is necessary to resolve the conflict when they do not. In a related vein, a main effect of congruency of number on an area judgment task was found in adults' performance on a number/size Stroop paradigm (Nys and Content, 2012). The authors argued that adults do automatically extract numerosity from non-symbolic arrays and that this interferes with area judgments. An arithmetic measure was also included in the study and did not correlate with overall accuracy on the numerical comparison task but did correlate with the effect size of the interference of number on the area comparison task, suggesting that, in adults, math achievement is associated with the ability to extract information relevant to number in non-symbolic arrays.

The role of inhibitory control in mathematics achievement is of particular interest given the mixed evidence for the relationship between non-symbolic magnitude comparison and math achievement from primary school into adulthood (see De Smedt et al., 2013 for a review). While symbolic magnitude processing has consistently been shown to correlate with formal mathematics performance, conflicting findings have been reported concerning non-symbolic magnitude processing. One explanation for the varied results is that non-symbolic comparison tasks vary in terms of control of visual parameters, length of stimulus display, and even which outcome measure is used. ANS precision can be operationalized as overall accuracy on a comparison task, the Weber fraction estimate of performance, or ratio effects on accuracy or reaction time and these different indices of performance do not always correlate strongly with each other (Gilmore, Attridge, De Smedt, & Inglis, 2014; Price, Palmer, Battista, & Ansari, 2012). As non-symbolic tasks measure inhibitory control in addition to number sense (Clayton & Gilmore, 2015), it could be that the observed relationships between non-symbolic comparison and maths achievement are driven by inhibitory control (Fuhs & McNeil, 2013; Gilmore et al., 2013). A recent meta-analysis of the relationship between non-symbolic comparison and mathematics achievement found that this relationship was overall weak, but strongest for children younger than six (Fazio, Bailey, Thompson, & Siegler, 2014). In a related vein, younger children are more susceptible to congruity effects on non-symbolic comparison tasks than older children and adults are (Szűcs et al., 2013). This suggests that the inhibitory demands of this task are strongest for young children, and this in turn may explain why non-symbolic comparison and maths relate most strongly in this age group.

Relationships between inhibitory control and different components of mathematics achievement change across development (Gilmore, Keeble, Richardson, & Cragg, 2015). For example, inhibitory control was found to be most strongly related to conceptual mathematics knowledge in adults, but most strongly associated with procedural knowledge in 11-14-year old children (Gilmore et al., 2015). However, in a study of children's arithmetic strategy use, inhibitory control was found to be related to conceptual knowledge in 8-10-year-old children (Robinson & Dubé, 2013), further demonstrating the importance of the context in which inhibition is occurring. As young children have limited knowledge of numerical symbols, they have a more narrow range of maths skills across which to investigate relationships with inhibitory control, and for this reason research in preschoolers has focused on non-symbolic comparison.

2.1.3. Challenges in measuring inhibitory control in young children

Developing valid and reliable measures of inhibitory control across the preschool years is not straightforward. Many tasks designed to measure shifting and inhibition in young children, such as the Dimensional Change Card Sort (Zelazo, 2006) and Shape School (Espy, 1997), have shown that five-year-olds perform at ceiling on tasks that three-year-olds find challenging. Children experience rapid development in executive control during this time (Garon et al., 2008), which presents a challenge for ensuring tasks are at an appropriate level of difficulty across this age group. Additionally, the task impurity problem highlights that because tasks necessarily measure inhibition of something (e.g. irrelevant stimulus dimensions, a prepotent response) they always require processes in addition to inhibition (Burgess, 1997). This is especially pertinent in a developmental context, as development in non-executive cognitive skills may also be critical to performance. One way to address this issue is by using multiple measures of inhibitory control and calculating a composite measure. For example, in a study of preschool inhibitory control and mathematics achievement, Fuhs and McNeil (2013) measured inhibitory control in three different tasks: Head/Feet, Day/Night Stroop, and Knock/Tap. However, despite hypothesizing that math achievement relies on being able to filter out irrelevant information at the level of the stimulus, two of the three included tasks relied primarily on inhibition of a gross motor response, whereas interference requirements were less heavy. Specifically, resolving conflict between discrete and continuous quantity in non-symbolic magnitude arrays requires inhibiting perceptual features associated with continuous quantity in order to attend to the number of items in the set (Clayton & Gilmore, 2015), or interference control.

2.2. Experiment 1. Relationships between inhibitory control and mathematics in preschool

The current study was designed to test the hypothesis that inhibitory control, specifically interference control, is relevant to early numeracy skills. An Animal Size Stroop task (Bryce et al., 2011; Szűcs et al., 2009) was modified to be appropriate for preschoolers (Thompson, 2011, unpublished undergraduate dissertation). Participants had to decide which of two animals was larger in real life and the size of the images was manipulated so that, on incongruent trials, the size of the animal image was in conflict with its size in the real world. Results from a study using an object/size stroop task on which participants were instructed to pick the larger of two pictures of familiar objects showed that adults experienced conflict effects when the larger image was of an object that was smaller in real life even though real-world size was irrelevant (Konkle & Oliva, 2012). This suggests that we automatically access the real-world size of an object when looking at an image of it and further supports the hypothesis that conflict effects are expected on the Animal Size Stroop. A different version of an Animal Stroop was previously developed by Wright, Waterman, Prescott, and Eaton (2003) in which children were required to name the animal stimuli, including incongruent trials where stimuli had heads and bodies of different animals. However in this experiment, size congruency was of interest as this stimulus dimension may be especially relevant to early maths skills. Furthermore, a task similar to the current one is also part of the Cognitive Assessment System (Naglieri & Das, 1997), although participants are required to give verbal responses.

The goal of the current research was to use this measure to elucidate specific relationships between inhibitory control and early numeracy skills in young children. Based on the hypothesis that mathematics achievement involves ignoring irrelevant stimulus dimensions, we predicted performance on the Animal Size Stroop to be associated with math achievement, as measured by a standardised measure of mathematics in young children. We also expected Animal Size Stroop performance to be associated, more precisely, with non-symbolic magnitude comparison, given the inhibitory demands of that task.

2.2.1. Method

Participants. Seventy children recruited from local nurseries and schools participated. Informed consent was obtained from parents. Twelve children were excluded for missing data for two or more measures. Eight of the twelve excluded children were three –year-olds, which suggests the tasks may be more difficult for the youngest children, but the excluded children did not differ from the rest of the sample in terms of gender, verbal IQ or visual-spatial IQ. The final sample used in the analyses included fifty-eight children ranging in age from 36 to 72 months (*M* = 51.97, *SD* = 10.98), of which twenty-nine were female.

Materials. All computer games were presented with E-Prime 1.0 software on an Elo AccuTouch 17" touchscreen monitor. The stimuli in the Animal Stroop task was taken from a set of coloured Snodgrass and Vanderwart pictures (Rossion & Pourtois, 2004). In the Animal Size Stroop task, the colours of the cow image were modified to increase familiarity, and an image of a ladybird was added.

***Animal Size Stroop task*.** The stimuli were chosen based on results from a pilot task in which 3-4 year old children played a zoo sorting game and sorted 17 animal pictures (7.5 by 4.5 cm) into a small or large cage based on their real life size. Four large animals (elephant, horse, cow, and lion) and four small animals (frog, ladybird, mouse, and rabbit) that were sorted correctly by the majority of children were selected for inclusion in the Stroop task. Participating children also completed the sorting task as part of the experimental protocol in order to assess their knowledge of animals' sizes. Children who failed to correctly sort at least six of the eight animals were excluded from the study, as the size congruency manipulation on the Animal Size Stroop task was dependent on animal size knowledge.

In the Animal Size Stroop task, children were instructed to choose which of two animals was the largest in real life, whilst ignoring the differing sizes of the animal images on the computer screen. The two animals were displayed on the screen with a smaller image (6 by 4 cm) on one side and a larger image (15 by 10 cm) on the other (see Figure 2.1). The response area was identical for each picture (17 x 14 cm) and stimuli were displayed for up to 5000ms. Correct responses were reinforced with a smiley face displayed for 100ms, whereas incorrect responses were followed by a blank display. Congruency, response side, and animal size were counterbalanced and trials were displayed in random order.

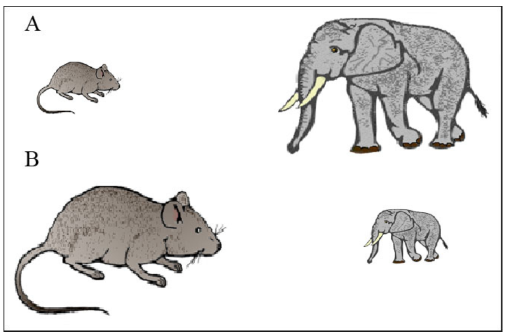

Figure 2.1. Example stimuli. (A) Congruent condition. (B) Incongruent condition.

Test of Early Math Achievement – 3. Early numeracy skills, including counting, cardinality, symbolic number knowledge, magnitude comparison, and arithmetic, were assessed with the Test of Early Math Achievement – 3 (TEMA-3; Ginsburg & Baroody, 2003). The TEMA-3 has 72 items and testing was discontinued when children reach a ceiling of 5 consecutive failed items. Entry point varies by age, but a basal of 5 consecutive correct items must be reached. The raw score was the number of items correct below ceiling and all items below the basal were counted as correct.

Magnitude comparison task. A computerized non-symbolic magnitude comparison task was used to measure number sense. Children had to pick the larger of two dot arrays and were instructed to help Bob the Builder judge the hole digging competition. In order to test whether set size influenced magnitude comparison, one block of trials required small comparisons (1-3), and the other large comparisons (10-39) (see Figure 2.2). The ratio between each comparison was small (.33), medium (.5), or large (.67). Order of block presentation was counterbalanced and there were 24 trials per block. The dots were presented for a maximum of 1200ms to discourage counting. A blank screen followed dot presentation for an additional 1800ms, which allowed for a total response window of 3000ms. Contour, area, density, and brightness of the stimuli were controlled for and the area of individual dots was randomised across items and numerosity. It is important to note that this task was designed in 2009, and, therefore, these stimulus parameters were chosen in order to be consistent with the literature at the time.

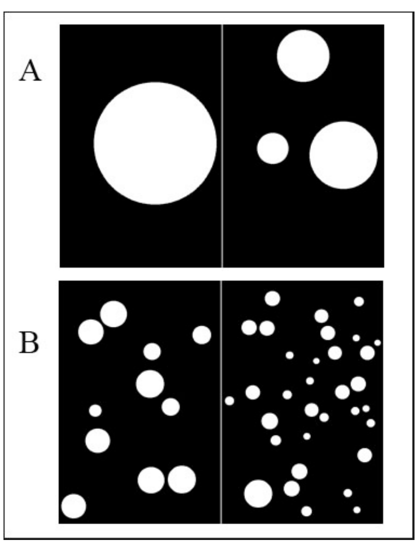

Figure 2.2. Example stimuli. (A) Small number condition. (B) Large number condition.

British Ability Scales – II. The Pattern Construction Subscale of the British Ability Scales-II (PC-Subscale, BAS-II; Elliot, Smith, & McCulloch, 1996) was used to assess nonverbal IQ. Children used coloured foam squares and patterned cubes to copy patterns presented in a book. Scores were based on response time and accuracy.

British Picture Vocabulary Scale – III. Verbal IQ was measured with the British Picture Vocabulary Scale (BPVS-III; Dunn et al., 2009), a measure of receptive vocabulary. Children were shown four pictures and instructed to point to the picture that corresponded to the word said by the experimenter. The raw score was the total number of items correct and the task was administered until a child made 8 or more errors on one set of 12 items.

Procedure. Each child was tested in a quiet area for approximately 1.5 hours over the smallest number of possible sessions and given regular breaks. The data included in this experiment are a subset from a larger battery of tasks. All sessions took place with an experimenter at the school or nursery and children were allowed to choose a sticker at the end of each session.

Analysis. Raw scores of standardised measures were used in the analyses as not all included tasks were standardised. Overall accuracy was used as the dependent measure in the analyses as well as accuracy separated by condition as this method was used in previous studies (Fuhs and McNeil, 2013; Gilmore et al., 2013). Furthermore, Gilmore and colleagues (2014) showed that overall accuracy was the most reliable measure of ANS acuity on magnitude comparison tasks in children. Analyses on reaction time data are not reported because low accuracy led to a limited number of valid trials contributing to these analyses.

2.2.2. Results

Descriptive statistics. Statistical analyses were performed with SPSS (release version 20.0.0). Descriptive statistics separated into age categories are reported in Table 2.1. One child had just turned 6 years old in the month they were tested but they have been included with the other 5-year-olds. Age in months was significantly correlated with all measures and was therefore included as a covariate in all analyses. There were no significant gender differences on any of the measures.

Table 2.1. Descriptive statistics separated by age.

		3yo		*4yo*		*5yo*	
Task	Measure	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N
Animal	Accuracy (Congruent)	86.06% (13.91%)	22	89.64% (13.33%)	12	94.59% (8.79%)	17
Size	Accuracy (Incongruent)	72.27% (32.39%)	22	66.72% (27.38%)	12	89.76% (18.77%)	17
Stroop	Median RT (Congruent)	1471.3 (405.69)	22	1243.04 (236.61)	12	1097.41 (219.56)	17
	Median RT (Incongruent)	1709.84 (301.99)	20	1472.67 (309.44)	12	1231.35 (221.2)	17
Magnitude	Accuracy (Small)	48.24% (22.86%)	22	48.18 (18.52%)	13	78.53% (21.16%)	17
Comparison	Accuracy (Large)	51.46% (14.26%)	23	43.31% (16.12%)	12	68.41% (22.62%)	17
	Median RT (Small)	1502.73 (514.62)	22	1156.62 (483.59)	13	1245.03 (413.06)	17
	Median RT (Large)	1274.23 (482.05)	23	1045.21 (316.65)	12	1087.15 (385.78)	17
TEMA	Raw Score	9.69 (5.61)	26	15.07 (4.83)	14	28.11 (7.32)	18
BAS	Raw Score	7.28 (5.95)	25	12.35 (8.15)	14	25 (6.78)	18
BPVS	Raw Score	39.92 (8.59)	25	51.85 (11.46)	13	66.44 (8.19)	18

Note: TEMA = Test of Early Mathematics Achievement – III; BAS = British Abilities Scale; BPVS = British Picture Vocabulary Scale

Animal Size Stroop. Two children failed to complete the task and five children sorted less than 75% of the animals correctly in the sorting task and were excluded from the analysis. Therefore, data from 51 participants were included in the analysis. Kolmogorov-Smirnov tests revealed that neither incongruent nor congruent accuracy scores were normally distributed. In order to deal with these violations, but still test the hypotheses set out in the introduction, statistically significant effects were also tested using the appropriate equivalent non-parametric statistics. A repeated measures analysis of variance (ANOVA) was run on the accuracy data with Congruency as a within-subjects factor and Age in years as a between subjects factor. Results revealed a significant main effect of congruency, $F(1,48) = 15.19$, $p < .001$, $\eta^2_p = .24$, driven by higher accuracy on congruent ($M = 90.1\%$, $SE = 1.8\%$) than incongruent trials ($M = 76.3\%$, $SE = 3.9\%$). There was also a main effect of Age, $F(1,49) = 3.4$, $p = .042$, $\eta^2_p = .12$. Bonferroni-corrected post hoc tests showed that 5-year-olds' accuracy ($M = 92.2\%$, $SE = 4.2\%$) showed a trend towards being significantly higher than 3-year-olds' accuracy ($M = 79.2\%$, $SE = 3.7\%$), $p = .07$, and there was no significant differences between 4-year-olds' accuracy ($M = 78.2\%$, $SE = 5\%$), $p = .11$. In terms of non-parametric statistics, Wilcoxon's Signed Ranks test confirmed a significant difference between accuracy on congruent and incongruent conditions, $Z = -3.54$, $p < .001$. A Kruskal-Wallis test provided converging results for a significant effect of Age, $X^2(2) = 8.42$, $p = .015$.

Magnitude comparison. Performance on this task was quite low (see Table 2.1), in part because of missed responses due to the speeded aspect of the task. Overall mean accuracy on large comparisons was 55.12% (*SD* = 20.05) and a one-sample t-test showed that this was not significantly different from chance performance, 50%, $t(51) = 1.842$, $p = .071$. Only 12 children had mean accuracies greater than 50% for both large and small comparisons and therefore using an accuracy cut off would exclude the majority of the data set. I return to poor accuracy in the discussion section for Experiment 1, and in the justification for methodological improvements in Experiment 2.

Kolmogorov-Smirnov tests showed that accuracy was not normally distributed in any condition except in the small comparison, large ratio condition. With caution (and better replication in Experiment 2), a repeated measures analysis of variance (ANOVA) was run with comparison Size and Ratio as within-subjects factors and Age in years as a between-subjects factor. Results revealed a significant main effect of Age, $F(2,46) = 14.46$, $p < .001$, $\eta^2_p = .386$. Bonferroni-corrected pairwise comparisons showed that 5-year-olds' accuracy (*M* = 73.5%, *SE* = 3.8%) was significantly higher than both 4-year-olds' (*M* = 45.8%, *SE* = 4.5%) and 3-year-olds' (*M* = 50%, *SE* = 3.5%) accuracy (*ps* < .001), and there was no significant difference between 3 and 4-year-olds, $p = 1$. A Kruskal-Wallis test provided also revealed a significant effect of Age, $X^2(2) = 9.6$, $p = .008$. There was a significant interaction between Size, Ratio, and Age $F(4,92) = 2.87$, $p = .027$, $\eta^2_p = .111$. An analysis of simple main effects revealed this was driven by a significant main effect of Size on medium ratio trials for 5-year-olds, *F*(1,46) = 7.51, *p* = .009, η^2_p = .14. This effect is difficult to interpret as it was only seen on medium ratio trials, yet there were no significant main effects of ratio. Wilcoxon Signed Ranks test also failed to reach significance when contrasting accuracy between small and large comparisons, *Z* = -.911, *p* = .362.

Predicting early mathematics achievement. Bivariate and partial correlations between measures are reported in Table 2.2. Non-parametric bivariate correlations were also run, given the previously reported violations of parametric assumptions, and parametric statistics interpreted with caution throughout. When age was partialled out, TEMA score was significantly correlated with BAS score, *r*(54) = .463, *p* < .001, magnitude comparison accuracy, *r(*52) = .394, *p* = .003, and also with accuracy on the congruent and incongruent conditions of the Animal Size Stroop, congruent condition *r*(47) = .379, *p* = .007, incongruent condition, *r*(47) = .358, *p* = .011.

Hierarchical linear regression models were run in order to test whether inhibitory control, as measured by performance on the Animal Size Stroop, predicted TEMA score. Age alone accounted for approximately 71% of the variance in TEMA raw scores (adjusted R^2 = 0.705), *t* = 11.72, *p* < .001 (see Table 2.3). Overall accuracy on the Animal Size Stroop was a significant predictor of TEMA score above and beyond the variance that could be accounted for by age and BAS raw score, *t* = 2.73, *p* = .009, and the model had predictive validity (*F*(3,45) = 68.05, *p* < .001, adjusted R^2 = .802. Unexpectedly, magnitude comparison did account for some variance in TEMA scores over and above age and BAS score. However, magnitude comparison accuracy was not a significant predictor of TEMA score over and above Stroop accuracy. It should be noted that this sample size afforded limited statistical power for multiple regression with four predictors and results should be interpreted with caution.

Table 2.2. Spearman's bivariate correlations (above diagonal) and Pearson's partial correlations controlling for age in months (below diagonal)

Measure	1.	2.	3.	4.	5.	6.	9.
1. Age	-	.739**	.792**	.832**	.351*	.38*	.594**
2. BAS		-	.702**	.768**	.256	.43**	.524*
3. BPVS		.254*	-	.718**	.321*	.289*	.419**
4. TEMA		.463**	.215	-	.504**	.52**	.652**
5. AS Acc Congruent		.048	.131	.379**	-	.535**	.369*
6. AS Acc Incongruent		.178	.033	.358*	.404**	-	.529*
9. MC Acc		.154	-.139	.394*	.103	.204	-

Note:$^*p < .05$; $^{**}p < .01$. BAS = British Abilities Scale; BPVS = British Picture Vocabulary Scale TEMA = Test of Early Mathematics Achievement – III; AS = Animal Size Stroop; MC = Magnitude Comparison

Table 2.3. Regression models predicting TEMA scores

	Variable	ß	Adjusted R^2
Model 1	1. Age (months)	.843**	.721
Model 2	1. Age	.543**	.785
	BAS	.402**	
Model 3	1. Age	.511**	.801
	2. BAS	.343**	
	3. Stroop Accuracy	.197*	
Model 4	1. Age	.429**	.811
	2. BAS	.368**	
	3. MC Accuracy	.201*	
Model 5	1. Age	.397**	.821
	2. BAS	.329**	
	3. Stroop Accuracy	.166*	
	4. MC Accuracy	.163	

*Note: *p < .05; **p < .01.* BAS = British Abilities Scale; MC = Magnitude Comparison

2.2.3. Discussion

The significant effects of congruency on Animal Size Stroop accuracy suggested that this task was an effective measure of inhibitory control in preschoolers. Though there were main effects of age, with older children performing better than younger children, the older children also experienced conflict effects, suggesting that the task effectively measures inhibition in this specific age-range (3- to 5-year-olds). A limitation of the Animal Size Stroop is that it requires knowledge of real world animal sizes and preschoolers' knowledge varies depending on their experience, even for the animals that were selected to be highly familiar to the youngest children in our target age range. Accordingly, some three-year-olds were excluded from the current study because they lacked sufficient animal size knowledge. Individual differences in young children's animal size knowledge may be associated with variability in caregiver instruction and a potential alternative that cannot be ruled out by the current study is that children who have greater animal size knowledge have also received more instruction in other domains, including number. As children who did not meet the minimum knowledge requirements were excluded, and significant congruency effects were observed, participating children did have enough animal size knowledge to experience interference from incongruent size information on the task.

Mean accuracy on the non-symbolic magnitude comparison task was below chance level for both 3 and 4-year-olds, and 5-year-olds performed significantly better than the younger children. This finding is convergent with those of previous research showing that young children had difficulty choosing the larger of two non-symbolic arrays when perceptual features such as surface area were in conflict with numerosity (Rousselle et al., 2004; Sarnecka & Negen, 2014). This lends further support to the hypothesis that young children only learn to distinguish discrete and continuous quantity through experience with number (Mix et al., 2002; Rousselle & Noël, 2008; Leibovich & Henik, 2013; Negen & Sarnecka, 2014). Five-year-olds were in reception year, and therefore had more formal instruction about number. Alternatively, the timed nature of the task led to many missed reponses, suggesting that the youngest children's low accuracy could also be due to immature processing speed.

As predicted, performance on the Animal Size Stroop task was correlated with performance on a standardized measure of mathematics achievement. This relationship remained significant when age and visual-spatial IQ were accounted for, and suggested that the ability to inhibit stimulus dimensions specifically associated with number is related to early numeracy skills beyond simply judging the larger of two non-symbolic arrays. The TEMA includes measures of counting and cardinality and does include a non-symbolic comparison item, but the stimuli are not controlled for area and so do not create conflict. Fuhs and McNeil (2013) found a similar relationship between their inhibitory control measure and TEMA score when accounting for verbal IQ as measured by receptive vocabulary. Furthermore, they found that when they entered inhibitory control into their regression models, ANS acuity was no longer a significant predictor of TEMA score, and that without the inhibitory control measure, ANS acuity was a significant predictor, but this was driven by incongruent comparison trials. In the current study, accuracy on the non-symbolic comparison task was also a significant predictor of TEMA score, somewhat surprisingly considering that it did not seem to be an effective measure of number sense. When Animal Size Stroop accuracy was entered in to the model, magnitude comparison was no longer a significant predictor of TEMA score. This again suggests that the relationship between non-symbolic comparison and mathematics achievement is driven by inhibitory control, rather than number sense, as number and area were incongruent on the task.

We therefore manipulated the inhibitory demands of a non-symbolic comparison task to further explore this hypothesis in Experiment 2.

2.3. Experiment 2. Preschool magnitude comparison: Number sense or inhibitory control?

Performance on the magnitude comparison task in Experiment 1 was very low and main effects of ratio did not reach significance, which suggests that it was not a valid measure of number sense. The magnitude comparison task in Experiment 2 was therefore designed to be easier for young children than the previous one was, by manipulating stimulus parameters and allowing for slower responses. A previous study in preschoolers used stimuli in which the overall surface area was controlled for, but the size of individual objects in an array was homogeneous, and found that magnitude comparison performance was significantly above chance level (Wagner & Johnson, 2011). In this experiment, we also kept the size of individual dots constant within arrays in order to make numerosity more salient than other visual features. Additionally, we aimed to test more precisely whether the relationship between inhibitory control and magnitude comparison is driven by congruence of stimulus dimensions and added a manipulation of surface area similar to that of previous number/area Stroop tasks used with young children (Fuhs & McNeil, 2013; Rousselle & Noël, 2008; Rousselle et al., 2004). We hypothesised that greater inhibitory control would be required under conditions of conflict between number and other stimulus dimensions. We predicted that Animal Size Stroop performance would most strongly predict magnitude comparison abilities under conditions of conflict between number and dot size (when area across distinct numerosities was equated, “incongruent trials” henceforth), compared to the situation in which number and area were allowed to covary (“congruent trials” henceforth), providing redundant information.

2.3.1. Method

Participants. Sixty-seven children between the ages of 3 and 6 recruited from local nurseries and schools participated. Informed consent was obtained from parents. Five children were excluded due to insufficient animal knowledge, four children were excluded for failing to complete tasks, nine were excluded for not following task instructions, and one was excluded because of a diagnosis of a learning disorder. All of the excluded children were 5 years old or younger. The final sample used in the analyses included forty-eight children ranging in age from 39 to 80 months ($M = 63$, $SD = 13$), of which thirty-one were female.

Materials. All computer games were presented with E-Prime 2.0 software. Children in nurseries played on an Elo AccuTouch 17” touchscreen monitor and children in schools played on a 13.3” Toshiba laptop and used buttons on the keyboard to make responses.

Magnitude comparison. Children were asked to choose who had more, a hippo character or a moose character and to indicate the side with more by either touching the monitor or pushing a button. There was a 500ms fixation screen followed by a 2500ms stimulus display screen and then an unlimited response window. Stimulus display was timed in order to prevent counting but children were able to respond afterwards in order to avoid missed trials due to slow processing speed. Quantities ranged from 1 to 9 and pairs were chosen from ratios .25, .5 and .75, and we would expect magnitude comparisons to be more accurate and faster for pairs of quantities that yielded a small ratio (e.g., 1 and 4). Pairs were randomly presented and the side of the screen on which the largest quantity appeared was counterbalanced. In one block, all of the items were the same size, so that the array with the larger number of items also had a greater total surface area (see Figure 2.3A, congruent). In the other, the total surface area of the stimuli was equated across arrays so that the items were smaller in size in the array with the larger number of items (see Figure 2.3B, incongruent). We predicted that a greater degree of inhibitory control would be needed to compare numerosity in the incongruent compared to the congruent trials. Presentation of these two conditions was blocked here to facilitate explaining, especially to the youngest children, that their task throughout was to report which side of the display contained “more objects”, regardless of their visual characteristics. Block order was counterbalanced between participants. Both dots and real circular objects were used to retain task engagement in younger children, but as the particular stimuli did not affect accuracy or speed, analyses were carried out collapsing across stimulus types.

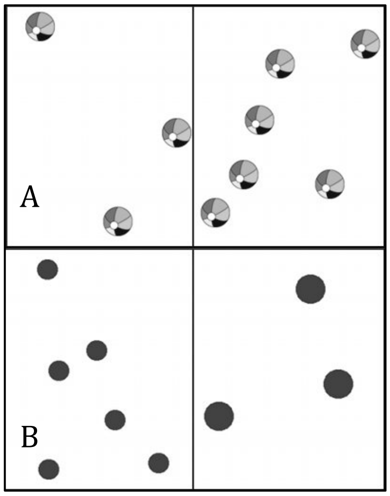

Figure 2.3. Example stimuli. (A) Congruent condition. (B) Incongruent condition.

Procedure. The data included in this experiment are a subset from a larger battery of tasks. Each child was seen for two sessions lasting approximately one hour in total. All sessions took place with an experimenter at the school or nursery and children were allowed to choose a sticker at the end of each session. The Animal Stroop task was administered in the same way as in Experiment 1 but the animal knowledge test was only administered to children who were in nursery.

Analysis. As in Experiment 1, accuracy was used as the dependent measure for both Animal Size Stroop, and non-symbolic magnitude comparison. Accuracy has consistently been used as the dependent measure for non-symbolic comparison (e.g. Clayton & Gilmore, 2015; Gilmore et al., 2015), whereas reaction time has been used as the dependent measure of an Animal Size Stroop task for older children (Gilmore et al., 2015). However, in the current study, reaction times would be more difficult to interpret given the untimed nature of the task and the large discrepancy in performance between younger and older children. Unlike in Experiment 1, age was treated as a between-subjects factor and three and four-year-old children were categorized as 'younger children', whereas five and six-year-old children were considered 'older children'. Equal numbers were tested in each age group, but more younger children than older children were excluded from the analysis.

2.3.2. Results

Animal Size Stroop. Accuracy data on the Animal Size Stroop were analysed with an ANOVA with Congruency as a within subjects factor and Age as a between subjects factor. Results revealed significant main effects of Age, $F(1,46) = 16.78$, $p < .001$, $\eta^2_p = .27$ and Congruency, $F(1,46) = 20.26$, $p < .001$, $\eta^2_p = .31$. There was also a significant interaction between Age and Congruency $F(1,46) = 8.15$, $p = .006$, $\eta^2_p = .15$ (see Figure 2.4). An analysis of simple main effects revealed that, for congruent trials, older children (M = 97.4%, SE = 1.8%) were significantly more accurate than younger children (M = 91%, SE = 2.1%), $F(1,46) = 5.49$, $p = .024$, $\eta^2_p = .11$, and this was also true for incongruent trials (M_{older} = 92%, SE = 4.2%; $M_{younger}$ = 67%, SE = 5%), $F(1,46) = 14.5$, $p < .001$, $\eta^2_p = .24$. Furthermore, accuracy was significantly higher for congruent than incongruent trials in younger children, $F(1,46) = 23.19$ $p < .001$, $\eta^2_p = .34$ $p < .001$, but the difference did not reach significance in older children. Kolmogorov-Smirnov tests revealed that neither incongruent nor congruent accuracy scores were normally distributed and Box's M was significant, suggesting caution when interpreting parametric statistics. A Mann-Whitney test provided converging results indicating a significant age difference, *Z* = -3.49, $p < .001$. Wilcoxon Signed Ranks tests confirmed non-parametrically a significant difference between accuracy on congruent and incongruent conditions for younger children, Z = -3.06, $p = .002$, and, in contrast to parametric statistics, revealed a significant difference between accuracy on congruent and incongruent conditions for older children as well, $Z = -2.09$, $p = .037$. Reaction time data are reported in Table 2.4.

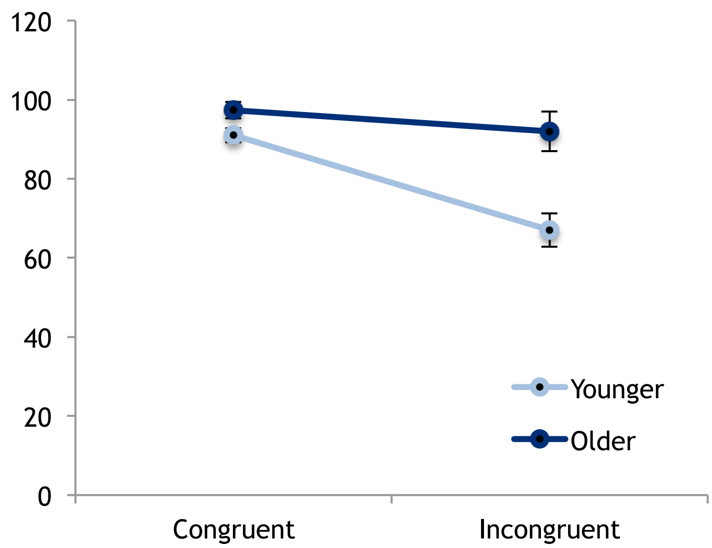

Figure 2.4. Animal Stroop accuracy separated by Age and Congruency. Error bars represent standard error.

Table 2.4. Median reaction times in groups, defined by splitting the sample by median age. Congruent and incongruent Conditions are reported separately and standard deviations are in parentheses.

	Younger (N = 20)	Older (N = 28)
Age in months	39 (7)	62 (6)
AS Congruent	1535.52 (440.11)	1188.12 (365.53)
AS Incongruent	2055.21 (907.93)	1283.38 (355.05)
MC Congruent	1717.3 (522.83)	1140.02 (332.8)
MC Incongruent	1854.6 (762)	1192.89 (365.25)

Magnitude comparison. The effect of block order did not reach significance, $F < 1$. Accuracy data on the magnitude comparison task was analysed with an ANOVA with Congruency and Ratio (small, medium, or large) as within subjects factors and Age as a between subjects factor. The Greenhouse-Geisser correction was used for the effects of Ratio as the assumption of sphericity was violated. Results revealed expected significant main effects of Congruency, $F(1,46) = 4.7$ $p = .035$, $\eta^2_p = .09$, driven by higher accuracy for congruent (*M* = 82.8%, *SE* = 1.9%) than incongruent trials (*M* = 78.6%, *SE* = 2.1%), and Ratio, $F(2,92) = 51.81$, $p < .001$, $\eta^2_p = .53$. Accuracy was higher for small ratio trials (*M* = 89.3%, *SE* = 1.5%) than medium ratio trials (*M* = 82.7%, *SE* = 2.2%), and higher for medium compared to large ratio trials (*M* = 70%, *SE* = 2.4%), *ps* < .001. There was also a main effect of Age, $F(1,46) = 19.4$, $p < .001$, $\eta^2_p = .3$, driven by better performance by older children (*M* = 88.2%, *SE* = 2.2%) than by younger children (*M* = 73.1%, *SE* = 2.6%) (see Figure 2.5). None of the interaction effects reached significance. Kolmogorov-Smirnov tests revealed that neither incongruent nor congruent accuracy scores were normally distributed and Box's M was significant, suggesting caution when interpreting parametric statistics. A Mann-Whitney test confirmed non-parametrically that there was a significant age difference, $Z = -3.88$, $p < .001$. However, a Wilcoxon Signed Ranks Test showed that the difference between accuracy on congruent and incongruent trials failed to reach significance, $Z = -1.43$, $p = .153$. A Pearson's bivariate correlation revealed a moderate negative correlation between age in months and the difference between accuracy on congruent trials compared to incongruent trials (i.e. conflict score) $r(46) = -.31$, $p = .034$, suggesting that younger children were more sensitive to the congruency between number and area than older children were.

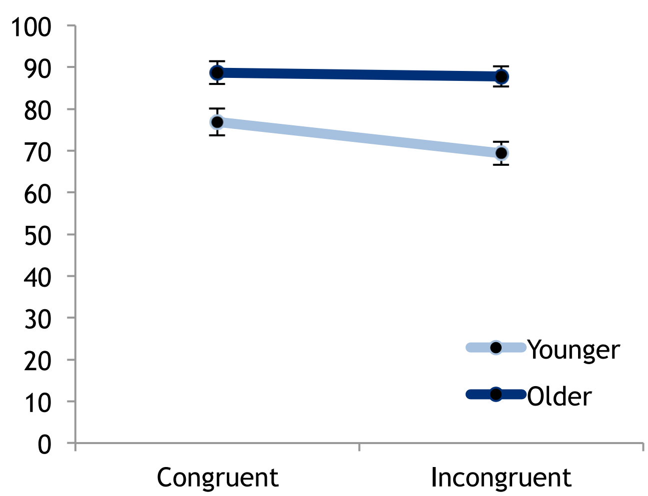

Figure 2.5. Magnitude comparison accuracy plotted by Age and Congruency. Error bars represent standard error.

Relationships between inhibitory control and number sense. Animal Size Stroop conflict scores, i.e., the difference in accuracy between conditions on the Animal Size Stroop, were significantly and negatively correlated with accuracy on incongruent trials in Magnitude Comparison, $r(45) = -.351$, $p = .015$, even when accuracy on the congruent condition of the magnitude comparison was partialled out. In other words, larger susceptibility to interference on the Animal Size Stroop was associated with lower accuracy on the incongruent magnitude comparison trials. However, the reverse did not hold: Accuracy conflict scores on the Animal Stroop task were not significantly correlated with accuracy on congruent magnitude comparison trials when accuracy on incongruent magnitude comparison trials was partialled out, $r(45) = .062$, $p = .678$, suggesting a specific relationship between conflict scores on the Animal Size Stroop and the higher conflict condition in the magnitude comparison task.

2.3.3. Discussion

Ratio effects reached significance on this modified magnitude comparison task, contrary to what was seen with the previous version of the task in Experiment 1. This suggests that children were basing their judgments on the number of stimuli and not solely on interfering perceptual information. Similarly to what was found in previous studies (Fuhs and McNeil, 2013; Gilmore et al., 2013), performance on magnitude comparison trials when size was in conflict with number was correlated with a measure of inhibitory control, but performance on trials when size and number were correlated was not. Specifically, performance on the incongruent trials was associated with the ability to inhibit irrelevant information about stimuli size. These results support the argument that non-symbolic comparison tasks are inhibitory control tasks when visual features are manipulated (Gilmore et al., 2013; Clayton & Gilmore, 2015). There was no significant difference in overall accuracy between the incongruent and congruent trials on the magnitude comparison task according to more conservative non-parametric tests. In previous studies (Fuhs & McNeil, 2013; Rousselle & Noël, 2008), significant differences in accuracy were found between conditions, but additional manipulations were built into those experiments that pitted continuous quantity and number even further in conflict. In the current study, the difference between conditions did negatively correlate with age, suggesting that children become less susceptible to conflicting size information with age and formal education. This finding supports the hypothesis that children become better at discerning discrete quantity with age and experience (Mix et al., 2002; Leibovich & Henik, 2013; Rousselle & Noël, 2008; Negen & Sarnecka, 2014). The difference between congruent and incongruent conditions on the Animal Size Stroop was significantly correlated with incongruent trials on the magnitude comparison task, suggesting that inhibitory control was required for those trials across the entire age range included in the study.

The current results suggest that age differences in both Animal Size Stroop and magnitude comparison performance are due to age-related improvements in inhibitory control, but, alternatively, they could also be due to improvements in other factors, such as processing speed or number knowledge. The interaction between congruency and age on the Animal Size Stroop indicated that older children were less susceptible to size interference than younger children were and, furthermore, that the task may be too easy for older children. Fewer animals were included in this version of the task than in Szűcs and colleagues' (2009), and there were more practice trials in order to ensure three-year-olds would be able to complete the task. However, Szűcs and colleagues (2009) reported congruency effects in older children and adults; therefore the difficulty of the task can be adapted for use with older children. Similarly, Gilmore and colleagues (2015) also found significant congruency effects on reaction time on an Animal Size Stroop task in older children and adults. It is likely that their version was speeded, but this information was not possible to glean from the paper. The discrepancy in task performance across younger and older children in the current study highlights the challenge in ensuring the difficulty of executive control tasks is appropriate across young children (e.g. Rousselle & Noël, 2008) discussed above (section 2.1.3.).

In summary, Experiment 2 broadly improved on the methodological weaknesses of Experiment 1, and overall provided support for our hypothesis that, even in very young children, performance on a measure tapping interference control (Animal Size Stroop) relates to performance on a measure traditionally thought to index number sense (non-symbolic number comparison), but is increasingly better construed as a task requiring a high degree of inhibitory control.

2.4. General Discussion

2.4.1. Summary of findings

Evidence was provided for a relationship between inhibitory control, as measured by the Animal Size Stroop task, and both non-symbolic magnitude comparison (Experiment 2) and a standardized measure of math achievement (Experiment 1). The role of non-symbolic magnitude comparison skills in the foundations of mathematical achievement has become highly debated (De Smedt et al., 2013) and these results provide additional support for the argument that measures of ANS precision are significantly influenced by visual parameters of non-symbolic stimuli (Gilmore et al., 2013; Szűcs et al., 2013), especially for children with limited cardinality knowledge (Negen & Sarnecka, 2014). Perhaps tellingly, the youngest children in Experiment 1 failed to perform above chance (i.e., to choose systematically the most numerous displays) when a commonly used format of the magnitude comparison task was used. In Experiment 2, magnitude comparison performance was facilitated, and a robust relationship across our interference control task and magnitude comparison performance was replicated, in a sample extending to children as young as three-years of age.

2.4.2. Limitations

A limitation of Experiment 2 was that a measure of mathematics achievement was not included, and so differential relationships between congruency conditions on the magnitude comparison task and math achievement could not be assessed. Based on the correlations between magnitude comparison and Animal Size Stroop tasks, models similar to Fuhs and McNeil (2013) and Gilmore and colleagues (2013) would be expected, with inhibitory control skills mediating fully or at least moderating the predictive role of magnitude comparison for math achievement. A further limitation is that the correlational nature of the study did not allow for an investigation of the precise mechanisms underlying the relationship between executive control and mathematics. There are at least two potential mechanisms for the established relationship between executive control and math achievement: the first, that executive processes in general may act as a gateway in the acquisition of math skills, even prior to formal math instruction; the second, that at each point in developmental time performing mathematical skills required deploying executive control. In order to explore the hypothesis that inhibitory control and other executive processes are important for learning about number, studies should be developed that assess learning, rather than simply measuring what children know. This could be done by evaluating different ways of teaching children about number in which executive control demands are manipulated or by using a longitudinal design with many time points in order to obtain a more dynamic view of children’s learning trajectories. Perhaps the ability to inhibit irrelevant information about size and other perceptual information in non-symbolic arrays is important for helping children to learn the meaning of numbers and that, for example, “three” applies to all sets of three regardless of perceptual characteristics, something that very young children struggle with (Huang, Spelke, & Snedeker, 2010; Slusser & Sarnecka, 2012). This hypothesis remains to be explored empirically.

2.4.3. Implications

Gaining greater insights into the relationship between inhibitory control and math achievement could lead to applications for math curriculum design in early childhood. The results presented here highlight the importance of inhibitory control for extracting information about number from non-symbolic arrays when faced with distracting size cues and suggest that this should be taken into consideration when assessing number skills in young children. It's possible that inhibitory control completely accounts for the relationship between non-symbolic comparison and magnitude comparison, specifically in younger children (Fazio et al., 2012). Alternatively, symbolic number skills have been found to be stronger predictors of maths achievement than non-symbolic skills in children over the age of five (e.g. Lyons et al., 2014; Göbel et al., 2014). As young children have limited symbolic knowledge, evaluating relationships between symbolic knowledge and maths achievement in this age group is less straightforward. However, recent evidence suggests that cardinality knowledge mediates the relationships between ANS acuity and maths achievement, even in preschoolers (Chu, vanMarle, & Geary, 2015). It remains unclear as yet how non-symbolic numerical representations may be important for the acquisition of symbolic knowledge. Clarifying the mechanisms underlying these relationships and the hypothesized role of inhibitory control could be useful for informing methods of teaching about symbolic number prior to the start of formal education.

In conclusion the two experiments presented here provide further evidence that non-symbolic magnitude comparison tasks require inhibition of irrelevant stimuli dimensions associated with continuous quantity (Gilmore et al., 2013; Szűcs et al., 2013; Clayton & Gilmore, 2015). This finding isolates an inhibitory process related to this maths skill and therefore highlights the need to test relationships between specific inhibitory processes and particular mathematical operations. In broader terms, these results also emphasize that researchers should consider the overlapping nature of executive demands with other developing skills when designing inhibitory control assessments for preschoolers. In addition to relating to magnitude comparison, performance on the Animal Size Stroop was also correlated with a more general measure of maths. The specific mechanisms underlying the relationship between executive control and general math achievement remain to be explored further. In order to investigate the hypothesis that inhibitory control and other executive processes are important for learning about math, future work should investigate how children learn, in addition to assessing what they know, and the role played by executive functions in the process of learning itself.

CHAPTER 3: PITTING NUMBER SENSE TRAINING AGAINST DOMAIN-GENERAL TRAINING

3.1. Introduction

As detailed in Chapter 2, correlational approaches are ultimately limited unless they are complemented by experimental manipulations. These are necessary for probing mechanisms underlying the relationships between executive control and mathematics outcomes, as well as for testing the hypothesis that executive control plays a role in learning mathematics. This chapter investigates one experimental manipulation aimed at modifying cognitive processes: training executive functions. In Experiment 3, we pitted domain-general EF training against a domain-specific numeracy intervention in order to test the hypothesis that improvements in executive function would transfer to early numeracy. I first review the controversial evidence for the efficacy of training EFs in terms of improving both the trained construct and transferring to educationally-relevant outcomes.

3.1.1. Theoretical rationale for training executive functions

Executive processes are correlated with and predictive of cognitive abilities required for succeeding in school and for this reason cognitive training regimes purporting to improve executive functions have recently been a focus of both empirical research (see Klingberg, 2010 for a review) and commercially available products (e.g. Cogmed). As executive functions are thought to be separable but related processes (Miyake et al., 2000), one would hypothesize that improvement in one process, such as updating, should be associated with improvements in related functions. In addition to the behavioural correlations between executive functions, neuroimaging evidence suggests that working memory and attentional control share neural circuits in parietal and prefrontal cortex (Klingberg, 2010), and training-induced plasticity of these circuits is a potential mechanistic explanation for hypothesized transfer effects from training one executive process to other related functions. Therefore, beyond the applied potential of improving educational achievement, training paradigms offer the opportunity to test whether manipulating executive processes affects domain-specific numeracy skills.

Training studies can have theoretical or applied purposes but it can be challenging to accomplish both (Jolles and Crone, 2012; Wass, Scerif, & Johnson, 2012). Specifically, training could be used with the goal of improving outcomes in populations with learning disorders, or could be used to investigate causal relationships between cognitive processes. The majority of existing studies have been motivated by application rather than theory and have aimed to improve educational outcomes in children who struggle with learning, such as children with attention deficit hyperactivity disorder (ADHD) (Klingberg, Forssberg, & Westerberg, 2002). For example, training on a mixed regime of tasks targeting working memory (e.g. Klingberg et al., 2002) led to improvements on the trained tasks, as well as an untrained visuo-spatial working memory task and a test of nonverbal reasoning in children with ADHD. However, when tested with more rigorous control groups, findings revealed no difference in parental and teacher ratings of behaviour across groups, suggesting that previously reported effects could be placebo effects (Chacko et al., 2014). Furthermore, despite the fact that the trained tasks should load onto the same construct in theory, the variety of trained tasks makes it difficult to determine what exactly is responsible for the observed improvements. For this reason, results from studies such as this cannot address theories about the causal mechanisms of executive processes in improvements in other skills. Furthermore, even evaluating the practical effectiveness of cognitive training programs is complicated by the variety of training programs, durations, and outcome measures used across studies, and it is to this literature that I now turn.

3.1.2. Review of executive function training studies

Methods and results of cognitive training studies in typically developing children aged 3-5-years old are summarized in Table 3.1. Although, as detailed later, there have been a number of meta-analyses and systematic reviews of cognitive training studies across the lifespan, this review focuses on the preschool age range, as this is the target age of Experiment 3. A number of overarching observations emerge: First, training updating, or working memory, has received the most attention, but executive attention, inhibitory control, and task switching programs have also been tested. As illustrated by Table 3.1, the degree to which a training program shows near transfer effects, defined as improvements in untrained tasks that measure the trained construct, and far transfer effects, defined as improvements in untrained tasks that measure constructs other than the one that was trained, varies greatly across studies. A training regime could be considered effective if it leads to near transfer alone, but for applied purposes should ideally also have far transfer effects. Cogmed, a commercially available working memory training program, stemmed from a study of working memory training in children with ADHD (Klingberg, et al., 2002) and has since been tested in comparison to a variety of control groups, including active control training programs targeting inhibitory control (Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009) or reasoning (Bergman Nutley et al., 2011) and has consistently shown near transfer effects to untrained measures of updating. However, evidence for far transfer effects of Cogmed, as well as other EF training, has been more mixed (see Table 3.1; Shipstead, Redick, & Engle, 2012). For example, one study in preschoolers showed that Cogmed working memory training improvements did not transfer to a measure of fluid intelligence (i.e., the ability to solve novel problems) (Bergman Nutley et al., 2011), whereas an executive attention training regime did claim to show far transfer to fluid intelligence in preschoolers (Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005).

Table 3.1. Training executive functions in typically developing preschoolers

Study	*n*	Age	Training Tasks	Training Duration	Control Group	Near Transfer	Far Transfer
Bergman Nutley et al., 2011	112	4 year-olds	Cogmed Non verbal reasoning training (NVR) (developed by Cogmed) Combined group (CB)	25 sessions, 15 min each, 5 weeks	Same tasks as CB group but stayed at lowest difficulty level	*WM group:* **Odd One Out*** Word Span *NVR group:* **Fluid intelligence latent variable*** *CB group:* **Fluid intelligence latent variable*** **Odd One Out*** Word Span	*WM group:* Fluid intelligence latent variable *NVR group:* Odd One Out Word Span
Kloo & Perner, 2003	74	3-5 year-olds	Task-switching (card-sorting) False-belief training	2 sessions, 15 min each, 2 weeks	Number-conservation tasks and relative-clauses	*Task-switching:* **DCCS*** *False-belief*: False-belief	*Task-switching:* **False-belief*** *False-belief:* DCCS
Kroesbergen et al., 2014	51	5 year-olds	Working memory training (domain-general) Working memory training with numerical information (domain-specific)	8 30min sessions, 4 weeks	Passive control	AWMA – Word Recall Backwards **AWMA - Odd One Out*** 	*Domain-general:* Early Numeracy Test - Revised **Magnitude Comparison*** *Domain-specific:* **Early Numeracy Test - Revised*** **Magnitude Comparison***

Passolunghi & Costa, 2014	48	5 year-olds	Working memory training (domain-general) Early numeracy training (domain-specific)	10 60min sessions, 5 weeks	Passive control	*Domain-general:* **Visuo-spatial working memory*** **Verbal working memory*** *Domain-specific:* **Early Numeracy Test***	*Domain-general:* **Early Numeracy Test*** *Domain-specific:* Visuo-spatial working memory Verbal working memory
Rueda et al., 2005	73	4 year-olds and 6 year-olds	Adaptive executive attention training	Five sessions over 2-3 weeks	Watched videos	Child Attention Network Test	K-BIT -Vocabulary **K-BIT - Matrices*** **K-BIT - IQ*** (four-year-olds only)
Thorell et al., 2009	65	4-5 year-olds	Cogmed Adaptive inhibition training (developed by Cogmed)	25 sessions, 15 min each, 5 weeks	Active control (commercial computer games) and Passive control	*WM group:* WAIS-R-NI - **Span board task*** **Word span task*** *Inhibition group:* Day-Night Stroop Go/No-go task	*WM group:* **NEPSY - Continuous Performance Test*** Day-Night Stroop **Go/No-go task*** WPPSI-R - Block Design *Inhibition group:* NEPSY - Continuous Performance Test WAIS-R-NI - Span board task Word span task WPPSI-R - Block Design

*indicates significant increase in performance at post-training compared to control group
Note: K-BIT = Kaufman Brief Intelligence Test; WAIS = Wechsler Abbreviated Intelligence Scale; NEPSY = A Developmental Neuropsychological Assessment; WPPSI = Wechsler Preschool and Primary Scale of Intelligence; DCCS = Dimensional Change Card Sort; AWMA = Automated Working Memory Assessment

The initial finding that training working memory in adults led to improvements not only in working memory span but also in fluid intelligence, a construct that had been previously believed to be static over the lifespan (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008), has been influential in the training literature. However, far transfer effects, particularly to fluid intelligence, have proven difficult to replicate (Chooi & Thompson, 2012; Thompson et al., 2013). A meta-analysis and systematic reviews showed that far transfer effects of working memory training are not reliable (Melby-Lervåg & Hulme, 2012; Shipstead et al., 2012a; Shipstead, Hicks, & Engle, 2012b). Based on a meta-analysis of twenty-three published training studies, Melby-Lervåg and Hulme (2012) concluded that working memory training did reliably lead to improvements in working memory skills but did not transfer to other skills, such as reasoning skills and inhibitory processes. Training studies have also been criticized for not having adequate control groups, not having reliable and valid measures of near and far transfer, and not assessing maintenance of training effects (Melby-Lervåg & Hulme, 2012; Shipstead et al., 2012a; Shipstead et al., 2012b).

Can individual differences predict training effectiveness? Evaluations of training effectiveness are based on group averages and there are likely large individual differences in improvements seen with training. Identifying factors that could predict training gains would be useful for determining whether, when, and to whom training could benefit. Given the inconsistent results of cognitive training studies, researchers have begun to investigate whether certain factors can predict improvement from training. For example, recent reviews of cognitive training in children argued that training may be more efficacious when administered earlier in the lifespan based on evidence that functional neural networks are less specialized at that time and may be more plastic (Jolles and Crone, 2012; Wass, et al., 2012). Brain maturation and the development of cognitive skills are influenced by interactions between genes and environment (Heckman, 2006), which supports the idea that intervening earlier, for example by training, may have a greater impact on subsequent development. Melby-Lervåg and Hulme (2012) included age as a moderator variable in their meta-analysis and found that training effects were larger for younger children compared to older children. Similarly, Wass and colleagues (2012) conducted a systematic review of transfer effects of working memory and mixed attention training studies and found a significant, yet weak, negative correlation between the age of participants and the size of the reported transfer effects, which suggested that training was more likely to have transfer effects in younger children.

Motivational factors may also account for individual differences in training effects. A training study in adults explored relationships between individual differences in motivation and theories of intelligence and working memory training effects (Jaeggi, Buschkuehl, Shah, & Jonides, 2014). Results showed that participants who believed that intelligence was malleable, as indicated by the Theories of Cognitive Abilities measure, were more likely to show improvements on a measure of nonverbal reasoning following working memory training than participants who believed that intelligence is fixed. Self-perceived cognitive failures were also assessed and participants who volunteered to complete the training were more likely to indicate cognitive deficits than participants who completed only the pre-test assessments. Furthermore, participants who completed the full training regime tended to have higher pre-test scores, suggesting that those who are highly intelligent and also believe that their cognitive skills need improvement may be most motivated to complete cognitive training.

It could also be that participants with the greatest cognitive deficits to start with stand to benefit the most from training. In a training study in typically developing children (aged 8-11), a significant negative correlation between fluid reasoning score before testing and the extent of the gain in reasoning score between pre and post training was found, which suggests that children who started with lower scores were likely to have higher gains (Mackey, Hill, Stone, & Bunge, 2011). The authors did not find the same correlation with processing speed scores, the skill the control group was trained on. Additionally, children who had the lowest math ability prior to domain-specific number sense training showed the largest improvements (Wilson, Dehaene, Dubois, & Fayol, 2009). Whether training is more effective for children who have a deficit in the trained skill to begin with is something that should be investigated further. From a practical standpoint, training should be targeted towards children with difficulties, as there is a clear need to intervene in these groups, whereas there is less incentive for improving cognitive function in typically developing children who are already achieving satisfactory academic outcomes

3.1.3. Far Transfer from Executive Function Training to Numeracy

Fewer studies have investigated transfer from cognitive training of executive processes to educationally relevant mathematics outcomes, than near transfer, or transfer to intelligence. Given the controversy surrounding far transfer of EF training discussed above, it is necessary to critically evaluate the evidence for far transfer from EF training to numeracy. In response to mounting criticism of working memory training, Gathercole, Dunning, & Holmes (2012) explained that conducting training evaluation studies with the most rigorous methods for testing the efficacy of a training program requires a substantial amount of time and money. They argued that preliminary studies are necessary in order to determine if it is worth investing the required resources into investigating a training intervention, but they should not be taken as definitive evidence that training reliably leads to near transfer effects and therefore could be considered effective. An ideal training intervention study should: a) randomly assign participants to groups and have assessors blind to their assignment, b) have a sample size that is sufficiently large to afford high statistical power, c) include a group that does not receive any training as well as an active control group that does some form of intervention that is not hypothesized to lead to the same improvements as the experimental training regime, and d) use multiple measures of the construct being trained as well as constructs to which far transfer is hypothesized (Gathercole et al., 2012; Rabipour & Raz, 2012;

Shipstead et al., 2012b). With these guidelines in mind, I turn to evaluating previous EF training studies that explored far transfer to numeracy. Furthermore, these guidelines influenced the design of Experiment 3, a point to which I return below.

A study in 6-year-olds from a low SES area showed that children who played adaptive games targeting working memory, planning, and inhibition improved at school measures of language and math compared to an active control group who played less cognitively demanding computer games (Goldin et al., 2014). However, this effect was moderated by school attendance, such that the participants who had missed more days of school showed the most benefits from the training. This suggests that the training helped these particular children catch up with their peers who attended more days of school and, therefore, implies that far transfer effects are specific to children who are struggling in school, rather than generalizing across all children in this age range. Furthermore, the study lacked a passive control group, and therefore it is not possible to rule out the possibility that improvements were due to test-retest effects or change over time, rather than to the training itself. In a study evaluating a different computerized working memory training program, eight weeks of training did not transfer to standardized mathematics assessments in a sample of typically developing 5-8-year old children (St Clair-Thompson, Stevens, Hunt, & Bolder, 2010), highlighting discrepancies across training regimes and populations. Witt (2011) found that 9-10-year old children who completed a working memory training intervention did show greater improvements in mental arithmetic relative to a passive control group. However, the training program consisted of a variety of games, including practicing a backwards digit recall task, inhibiting distractors, and verbal rehearsal. This is problematic because it is not possible to determine which of the trained tasks led to the observed improvements, especially because there was no active control comparison group.

Two published studies are particularly relevant to Experiment 3 as they contrasted domain-general updating training with domain-specific numeracy training in children in kindergarten (Kroesbergen et al., 2014) and preschool (Passolunghi & Costa, 2014). Kroesbergen and colleagues (2014) compared training effects across a group of low-performing 5-6-year-old children who completed domain-general working memory training, a group who completed working memory training with number-specific stimuli, and a passive control group. Children in both training groups showed improvements in post-training measures of working memory performance as well as on a standardised early numeracy skills assessment. Only the domain-specific training group showed significant improvements on a non-symbolic magnitude comparison task. Correlations showed that children’s improvement on a measure of visual-spatial working memory was significantly correlated with improvement on the early numeracy measure, which supports the hypothesis that working memory is causally related to mathematics achievement. Importantly, the domain-specific training study was not strictly domain-specific as it also involved working memory. In contrast, Passolunghi & Costa (2014) trained one group of preschoolers on paper-and-pencil tasks targeting working and short-term memory, and another on tasks targeting early numeracy skills, number line estimation, mapping between numerals and non-symbolic quantities, and counting. Results revealed that both intervention groups showed significant improvements on a measure of early numeracy, when compared to a passive control group, but only the domain-general group showed a significant increase in working memory skills. This provides additional support for a causal relationship between working memory, or updating, and mathematics achievement. However, both of these studies targeted updating exclusively and neither assessed transfer to other executive processes, therefore further work is needed to clarify mechanisms underlying relationships between executive control and mathematics achievement in preschool more broadly.

To summarize, given that one published study lacked an active control group (Witt, 2011), another lacked a passive control group (Goldin et al., 2014), and two others did not include measures of untrained executive processes (Kroesbergen, 2014; Passolunghi & Costa, 2014), more research is necessary in order to clarify whether training executive functions reliably leads to improvements to related executive processes as well as educationally-relevant mathematics outcomes. It is also important to note that studies on low-achieving samples (Goldin et al., 2014; Kroesbergen et al., 2012), have applied goals, namely improving outcomes in these children, and therefore are less appropriate for addressing theoretical questions pertaining to mechanisms underlying transfer. Indeed, the training regimes in each of these previous studies included multiple games targeting aspects of working memory. Furthermore, in contrast with the emphasis on preschoolers for training studies assessing near transfer (e.g., Thorell et al., 2009), preschoolers have more rarely been targeted or contrasted with older children in the context of testing far transfer: speculatively, researchers have suggested that training and transfer effects may be larger for younger individuals (Melby- Lervåg & Hulme, 2012; Wass et al., 2012), but further empirical evidence for the role of age in influencing training outcomes is still limited.

3.1.4. Domain-specific numeracy training interventions for young children

Interventions targeting specific mathematical abilities in order to improve maths achievement are ideal active controls when assessing far transfer from EF training to maths outcomes. For example, the Number Race software was designed as an intervention for children with dyscalculia, a specific deficit in mathematics in the absent of intellectual impairment. The intervention was based on the hypothesis that a core deficit in number sense underlies the difficulties in mathematics that occur in children with dyscalculia (Wilson et al., 2006). The primary task of the software is a numerical comparison task in which children select the side that has more items. Results of a study of the effectiveness of the software on 7-9-year-old children with dyscalculia showed that children performed significantly better on assessments of subtraction, numerical comparison, and subitizing, but not on addition, after using the software for ten hours in total over a five week period. However, only nine children were included in the final sample for the study and there was no control group, therefore this study alone is not convincing evidence for the effectiveness of the software. A further study was conducted on a larger sample of 4-6-year-old children in low socioeconomic status areas (Wilson, et al., 2009). Participants completed both the Number Race training and training with a commercially available educational software targeting reading ability in counter-balanced order. Children showed significant improvement on symbolic, but not non-symbolic, number comparison, following completion of the Number Race training, which the authors took to suggest that the training improved *access* to number sense. In other words, the training seemed to improve the ability to link numerical symbols to their associated approximate non-symbolic quantities, rather than the ability to process non-symbolic representations themselves. Thus, the mechanisms underlying even domain-specific interventions remain poorly understood, as training hypothesized to improve ANS precision did not lead to improvements on non-symbolic comparison as expected.

3.2. Experiment 3. Mechanism-focused cognitive training in preschoolers

The aim of Experiment 3 was to use cognitive training, both domain-general switching training and domain-specific ANS training, in typically developing preschoolers in order to manipulate these cognitive processes and explore causal mechanisms. Specifically, the goal of the study was to investigate mechanisms underlying correlational relationships between executive control and emerging numeracy skills, rather than to achieve the more practical goal of improving educational outcomes. A systematic review of relationships between executive processes and mathematics revealed updating to be the most reliable predictor of math achievement (Bull & Lee, 2014) and updating has been the focus of previous training studies aiming to improve numeracy (e.g. Kroesbergen et al., 2014).

However, training of other executive processes may also transfer to mathematics: there is evidence that task switching training leads to near transfer effects in older children (e.g. Karbach & Kray, 2009). Moreover, children learn to select discrete number as the key dimension in non-symbolic arrays, while ignoring other dimensions, over the first few years of life (e.g. Leibovich & Henik, 2013). Therefore improvements in directing attention to different stimulus dimensions, to select the task-relevant one, could relate causally to non-symbolic comparison skills or acquisition. Here, we therefore opted to train the ability to select stimulus dimensions and switch across them, to test our hypothesis of transfer to mathematical operations that, like non-symbolic number comparison, have an element of selection.

An active control group completed a number sense training protocol in order to directly compare effects of domain-specific versus domain-general cognitive training. If the correlational relationships between executive processes and mathematics are driven by a causal mechanism leading executive control to constrain maths performance, then training switching should transfer to early numeracy skills, but training number sense should not transfer to task switching. A further important element was that the target age group spanned preschool and early primary school, and we hypothesized that age would moderate training effects, specifically so that larger gains would be seen for younger compared to older children.

3.2.1. Method

Participants. Participants were recruited through local nurseries and primary schools. Children whose parents gave written consent for them to participate were included in the study. Sixty children between the ages of 3 years and 5.5 years participated in the study (M = 4.2, SD = 0.61), thirty-two of which were female. All of the children spoke English and 31 children were exposed to an additional language at home.

Assessments.

British Ability Scales – II. The Pattern Construction Subscale of the British Ability Scales-II (PC-Subscale, BAS-II; Elliot, Smith, & McCulloch, 1996) was used to assess non-verbal IQ and was described in Chapter 2 (section 2.2.1.).

British Picture Vocabulary Scale – III. Verbal IQ was measured with the British Picture Vocabulary Scale (BPVS-III; Dunn et al., 2009), a measure of receptive vocabulary. This assessment was also described in Chapter 2 (section 2.2.1.).

Test of Early Math Achievement – 3. The TEMA-3 was used to assess early maths skills and was previously described in Chapter 2 (section 2.2.1.).

Automated Working Memory Assessment. The Dot Matrix, Digit Recall, Backwards Recall, and Odd One Out subtests of the Automated Working Memory Assessment (AWMA; Alloway, 2007) were used to assess visuo-spatial short-term memory, verbal short-term memory, verbal working memory, and visuo-spatial working memory respectively. In the Dot Matrix test, children had to remember the order of locations of a red dot in a grid, the Digit Recall test was a digit span task and Backwards Recall was a backwards span task. The Odd One Out test required children to identify which of three shapes was different and remember its location. Memory span increased in each block and testing was discontinued when 3 items were failed in one block. Raw scores were based on the number of correct trials.

Leiter International Performance Scale – Revised. The Attention Sustained subtest of the Leiter International Performance Scale – Revised (Leiter-R; Roid & Miller, 1997) was used to assess sustained attention. Children completed timed visual search tasks in which they had to select target items amidst distractor items and draw a line through them. The dependent measure was the adjusted raw score, which was calculated by subtracting the number of errors made from the number of correct marks made within the time limit set.

Spatial conflict task. A spatial conflict task was used as a measure of inhibitory control. The display consisted of two response images (animals), one in each bottom corner of the screen, and a target image whose position varied. In the neutral condition, the target was displayed in the centre of the screen, in the congruent condition the target was displayed on the same side as the matching response, and in the incongruent condition the target was displayed on the opposite side of the response. Children were instructed to touch the response image that matched the target. Each trial consisted of a fixation screen presented for 500ms followed by the stimulus display for up to 3000ms. Feedback was given in the form of an animation for correct responses and a blank screen for incorrect responses. Three blocks of 16 neutral trials were presented in alternating order with three blocks of 24 congruent and incongruent trials, presented randomly in equal numbers. A conflict score was calculated by subtracting mean reaction time on the congruent condition from mean reaction time on the incongruent condition.

Flexible Item Sorting Task. Task switching was assessed using a modified version of the Flexible Item Sorting Task (FIST; Jacques & Zelazo, 2001). This task was chosen based on a review of tasks used to assess switching in children and adults (Cragg & Chevalier, 2012). Stimuli varied along four dimensions: colour, shape, size, and number. For the 3 practice trials, children were shown four cards, consisting of two pairs of exact matches, and instructed to select two cards that are alike in one way and then to select two cards that are alike in a different way. For the 12 test trials children were shown sets of three cards and given the same instructions. The three cards were all the same on two dimensions but varied along the other two dimensions. Children had to select two cards that were the same in one dimension and then two that were the same in a different dimension. For example, on one trial, the top card had 3 large blue squares, the middle had 3 large green circles, and the bottom had 3 large green squares, so cards one and three were the same shape and cards two and three were the same colour (see

Figure 3.1). Switch accuracy was operationalized as the proportion of correct second pair selections, when the first pair choice had also been correct.

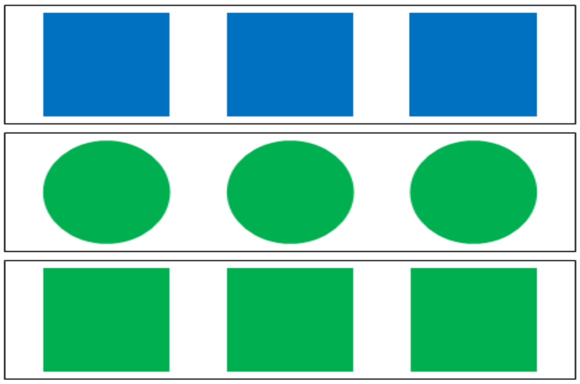

Figure 3.1. Example FIST stimuli.

Training Programs.

Task switching training. The task-switching training program was a Finding Nemo themed game based on the Dimensional Change Card Sort (Zelazo, 2006). Children alternated between playing “Marlon’s Game” in which they had to choose which of two bivalent stimuli was the same colour as the test stimuli and “Dory’s Game” in which they had to choose which of two bivalent stimuli was the same shape as the test stimuli (see Figure 3.2). The first session was a practice session based on manipulations that scaffolded focusing on the relevant dimensions (Caroll & Cragg, 2012). The next session consisted of 3 levels of four blocks of six trials. The rule switched between each block. At the end of each level, children were rewarded with a new character and each level they completed brought them closer to Nemo. After two sessions like this, children who achieved 80% accuracy on switch trials progressed to the next difficulty level. In the more difficult version, the number of switches was manipulated between levels. The first level had 10 blocks of two trials, the next had 4 blocks of 8 trials, and the next had 6 blocks of 4 trials. After three sessions like this, children who achieved 80% accuracy on switch trials progressed to the next difficulty level. In that version, a new rule was introduced. In “Mr Ray’s Game”, children had to choose which of two bivalent stimuli was the same orientation as the test stimuli. In the first part of this level, 4 blocks of 6 trials were presented alternating between the orientation and colour rules, the following part alternated between the shape and orientation rules, and the third alternated between all three rules.

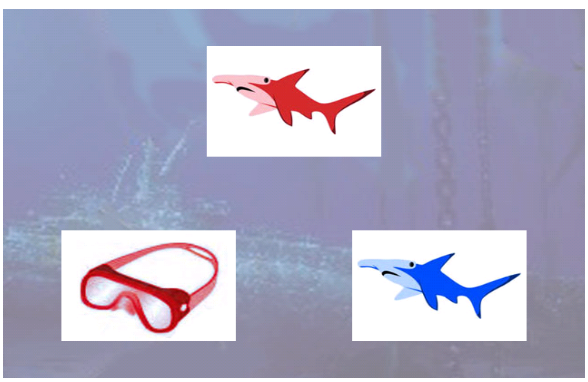

Figure 3.2. Example training trial. During Marlon’s game blocks, the correct response was the red mask and during Dory’s game blocks, the correct response was the blue shark.

Number sense training. Children in the number sense training group played The Number Race software developed by Wilson and colleagues (2006), an adaptive numerical comparison task incentivized by racing against a computerized character. Children were shown two quantities and instructed to choose the side they wanted (see Figure 3.3). They then moved forward along a game board the number of spaces corresponding to the quantity they chose. At the age range in the current study, the majority of comparisons were non-symbolic, but symbolic numbers were gradually introduced if children succeeded consistently at the non-symbolic comparisons. Only magnitudes from one to nine were included. Verbal labels often accompany the visual stimuli as well, so that in many trials information about number could be obtained from the non-symbolic array, the number word, or the symbolic number. If the child won the race, they were allowed to free a character.

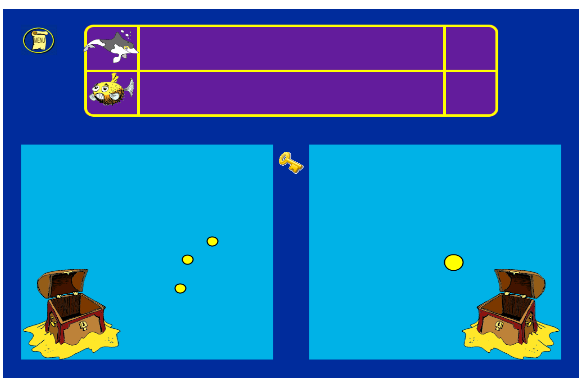

Figure 3.3. Example training trial. Non-symbolic comparison.

Procedure. The first thirty-nine participating children were assigned to a training group using stratified random assignment (by age and number of languages spoken) in order to have balanced groups. Twenty participants were in the number sense training group and 19 participants were in the task switching training group. Participants completed eight 15 minute training sessions over the course of six weeks. They also completed two 30-40 minute assessment sessions over the three weeks prior to the start of training and the three weeks following the completion of training. Participants in the passive control group completed only the assessment sessions. Computerized tasks were presented on a 17" Elo AccuTouch touchscreen monitor and E-Prime 2.0 software was used to display stimuli (Psychology Software Tools, Pittsburgh, PA). All sessions took place with an experimenter at the school or nurseries and children were allowed to choose a sticker at the end of each session.

Statistical design. Power analyses were run prior to the start of the study and indicated that sixty participants per group were required in order to reliably detect a medium effect size, such as the effects reported in a similar study (Thorell et al., 2009). However, training 120 children was not feasible and, importantly, previous studies have found effects with similar sample sizes, which suggested that it has previously been possible to detect effects with the current sample size, should they exist (e.g. Kroesbergen et al., 2014; Witt, 2011; Passolunghi & Costa, 2014).

3.2.2. Results

Correlational relationships pre-training. Relationships between measures from the pre-training assessment were investigated in order to determine which executive processes were associated with math achievement prior to training. One participant who had only recently turned 3 was excluded from the start of the study for failure to comply with the testing instructions. There were also additional missing data points from children who refused to finish certain tasks either due to boredom or fatigue. Descriptive statistics are reported in Table 3.2. Kolmogorov-Smirnov tests were run to test assumptions of normality and the only measure that violated this assumption was Backwards Digit Recall.

Table 3.2. Descriptive statistics

Task	Measure	N	Mean (SD)
TEMA	Standard Score	59	100.58 (13.11)
BAS	T Score	59	53.69 (9.08)
BPVS	Standard Score	59	99.09 (11.96)
Spatial Conflict	Accuracy Conflict	56	5.28% (14.99%)
	Median RT Conflict	56	136.53 (197.99)
Visual Search	Adjusted Raw Score	59	33.46 (13.65)
AWMA (Raw Scores)	Digit Recall	59	14.24 (7.04)
	Dot Matrix	57	10.23 (4.49)
	Odd-One-Out	57	7.73 (2.92)
	Backwards Recall	57	1.58 (2.92)
FIST	Switch Accuracy	57	32.43% (25.4%)

Note: TEMA = Test of Early Mathematics Achievement – III; BAS = British Abilities Scale; BPVS = British Picture Vocabulary Scale; AWMA = Automated Working Memory Assessment; FIST = Flexible Item Sorting Task

Many children struggled with the Backwards Recall task, with the majority obtaining a score of 0 (N = 42) and so this measure was excluded from further analyses. The spatial conflict task did seem to measure executive attention as conflict effects were seen on both accuracy, $t(55) = 2.64$, $p = .011$, and reaction time, $t(55) = 5.16$, $p < .001$.

Age in months was significantly correlated with all measures except for spatial conflict accuracy and reaction time conflict scores (see Table 3.3). When age was partialled out, raw TEMA score was significantly positively correlated with verbal and nonverbal IQ measures, visuo-spatial short-term memory, and visual search, and significantly negatively correlated with spatial conflict accuracy conflict score. In other words, better early numeracy was associated with better performance on sustained attention and inhibition measures, but not with updating or switching measures. FIST switch accuracy was only significantly positively correlated with visuo-spatial short-term memory when age was partialled out. Measures of different executive processes were not significantly correlated with each other when age was accounted for.

Table 3.3. Bivariate correlations (above diagonal) and partial correlations controlling for age (below diagonal)

Measure	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Age	.67**	.61**	.60**	-.01	-.08	.61**	.51**	.43**	.36**	.39**
2. TEMA	-	.75**	.7**	-.24	-.05	.62**	.45**	.59**	.39**	.39**
3. BAS	.57**	-	.54**	-.12	.03	.65**	.23	.52**	.35**	.32*
4. BPVS	.5**	.26*	-	-.03	.04	.6**	.45**	.59**	.39**	.37**
5. SC Acc Conflict	-.31*	-.14	-.03	-	.28*	-.15	.06	-.32*	-.05	-.2
6. SC RT Conflict	.01	.1	.11	.28*	-	.05	-.06	-.08	.14	-.06
7. Visual Search	.37*	.44**	.37**	-.18	.13	-	.26*	.56**	.35**	.38**
8. Digit Recall	.18	-.11	.2	.08	-.02	-.07	-	.3*	.29*	.37**
9. Dot Matrix	.46**	.36**	.46**	-.34**	-.05	.46**	.1	-	.45**	.42**
10. Odd-One-Out	.23	.17	.23	-.05	.18	.18	.13	.35**	-	.32*
11. FIST	.19	.11	.18	-.21	-.03	.19	.21	.3*	.21	-

*Note: *p < .05; **p < .01.* TEMA = Test of Early Mathematics Achievement – III; BAS = British Abilities Scale; BPVS = British Picture Vocabulary Scale

One-way ANOVAs were run on all measures with Group as a between subjects factor in order to ensure that the groups did not differ prior to training. A main effect of Group reached significance only for Digit Recall, $F(2,56) = 7.61$, $p = 0.01$, $\eta^2_p = .214$. Post hoc tests revealed that the passive control group had a higher mean Digit Recall score ($M = 18.45$, $SD = 7.35$) than the numeracy training group did ($M = 10.7$, $SD = 5.93$), $p = .001$, but the switching training group's mean score ($M = 13.53$, $SD = 5.61$) did not differ from the other training group, p = .512, or the passive control group, p = .057. Importantly, the groups did not differ in terms of IQ or in the near transfer measures, TEMA or FIST.

Training results. The task switching training had three levels and proceeding to the next level was contingent on an accuracy cut-off of 80%. Figure 3.4 shows the number of children who reached each level. A Spearman's correlation revealed that age was not correlated with level reached, $r(18)$ = .034, p = .892. Information on training gains was not available from the Number Race software.

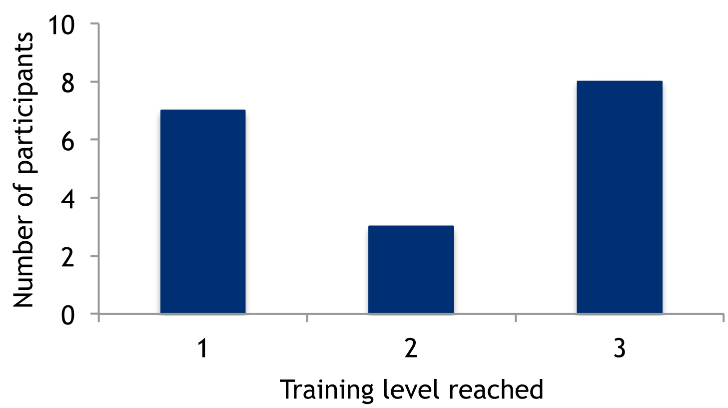

Figure 3.4. Number of children who reached each level of switching training

Transfer effects. Three participants from each training group failed to complete post-assessments due to noncompliance or absence from nursery and so there were 17 children in the number sense training group for these analyses and 16 in the switching training group. A one-way ANOVA was performed on the gain in TEMA raw score with Group as a between subjects factor and Age in months as a covariate. Results failed to reach significance, indicating that there were no significant group or age differences (see Figure 3.5). A one-way ANOVA on gain in FIST switch accuracy revealed the same pattern of results. One sample t-tests revealed that both mean TEMA gain (M = 2.79, SD = 3.1) and mean FIST gain (M = 9.2%, SD = 23.8%) were significantly different from 0, $t(51) = 6.5$, $p < .001$, and $t(48) = 2.7$, $p = .010$, respectively, indicating that on average participants' performance on these outcome measures improved between pre-assessment and post-assessment.

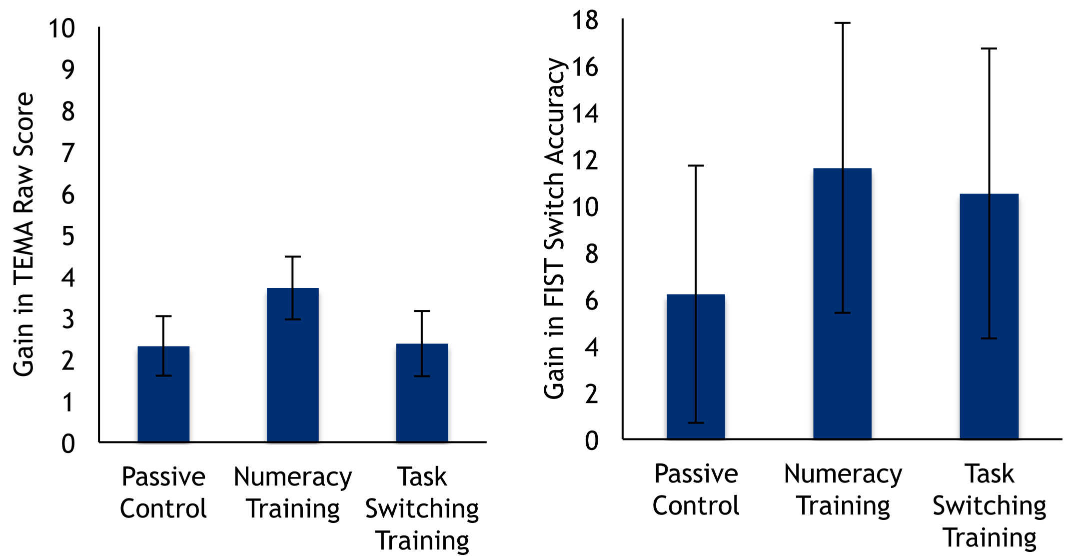

Figure 3.5. Group mean gain scores for TEMA and FIST. Error bars show standard error of the mean.

Individual differences in training gains. Training gain scores were highly variable across participants and so correlations were run between gain scores and measures that were significantly correlated with TEMA or FIST score at pre-assessment to investigate potential predictors of improvement. TEMA gain scores and FIST gain scores were not significantly correlated with each other, $r(48) = .04$, $p = .789$, suggesting that improvements were not attributable to a common underlying factor. Bivariate correlations on the whole sample revealed that gain in FIST switch accuracy was significantly negatively correlated with switch accuracy pre-assessment score, $r(49) = -.406$, $p = .004$, and weakly positively correlated with BPVS raw score, $r(49) = .283$, $p = .048$. Gain in TEMA score was not significantly correlated with any other measure. Correlations were then run separately across groups in order to disentangle training effects from test-retest effects (see Table 2.4). Independent samples t-tests were run to test whether there were differences between children in the switching training who reached level 3 compared to those who reached the first level. Results revealed no significant differences in either TEMA gain $t(12) = .63$, $p = .54$, or FIST switch accuracy $t(11) = .82$, $p = .44$. Independent samples t-tests were also run to see whether there were baseline differences between participants who reached different levels of training. Participants who reached level 3 had significantly higher verbal IQ than those who reached level 1, $t(13) = 3.72$, $p = .004$. Participants who reached level 3 also showed a trend towards having higher baseline scores on TEMA, $t(13) = 3.14$, $p = .012$, visuo-spatial short-term memory, $t(13) = 2.93$, $p = .012$, and sustained attention, $t(13) = 2.63$, $p = .021$, but these did not survive a Bonferroni correction for multiple comparisons ($p < .008$). There was no significant difference at baseline between these sub-groups on switch accuracy $t(12) = 1.88$, $p = .09$ or non-verbal IQ, $t(12) = 1.95$, $p = .07$.

Table 3.4. Correlations between baseline assessments and training outcomes.

	Numeracy	Training (N = 17)	Switching	Training (N = 16)	Passive	Control (N = 19)
Pre-Assessment	TEMA	FIST	TEMA	FIST	TEMA	FIST
Age (months)	.21	.35	.3	-.05	-.25	-.1
TEMA Raw	.27	.22	.43	.33	-.03	.09
FIST Switch	-.04	-.35	.03	-.51†	.3	-.45†
Visual Search	.25	.31	.28	.27	.4	-.43†
Dot Matrix	.42	.27	.21	.36	.04	-.16
BAS Raw	.04	.33	.59*	.38	-.17	-.11
BPVS Raw	-.04	.26	.44	.34	.14	-.01

Note: *$p < .05$; **$p < .001$; †*trend towards significance* ($p < .07$)

3.3. Discussion

3.3.1. Summary of findings

The aim of this study was to manipulate young children’s EF skills, namely switching, through training in order to explore how these changes would affect early numeracy. However, no significant effects of training were found either on closely related untrained measures, or on the measures to which far transfer was hypothesized. Participants showed improvements, on average, on both switching and early numeracy outcome measures between pre- and post-assessment, regardless of whether they completed switching training, numeracy training, or did nothing in the intervening time. These results highlight the need to have adequate control groups for training studies as, had a passive control group not been included, results would instead have suggested that the training programs did have significant near and far transfer effects. The fact that the participants in the passive control group did not differ from the training groups in terms of change in performance between pre- and post- assessments indicates instead that improvements are likely attributable to test re-test effects or developmental time rather than the training regimes. The lack of correlations between improvement on the switching training itself and improvements on the outcome measures, as well as the fact that the significant negative correlation between baseline performance and gain score on the FIST was seen in both the training and passive control group, further support the idea that the training programs were not responsible for the observed increases in performance. These results contribute to a growing body of evidence suggesting that training executive processes does not generalize to related cognitive skills (Melby-Lervåg & Hulme, 2012; Shipstead et al., 2012a; Shipstead et al., 2012b). The current findings contradict those of previous similar studies in preschoolers (Kroesbergen et al., 2014; Passolunghi & Costa, 2014) that did show significant far transfer effects from EF training to early numeracy. However, the EF training regimes used in those studies targeted updating, whereas switching was trained in the current study. An alternative explanation is that updating may be causally related to mathematics achievement, whereas switching is not.

We also hypothesized that there would be individual differences in training gains and, in particular, that younger children would be more likely to show training-related improvements. There were no significant correlations between age in months and training gains, which fails to support this hypothesis. Children in the switching group who reached level 3 of the training had significantly higher verbal IQ scores at baseline compared to children who did not progress past level 1, suggesting that children who had higher verbal IQ at baseline were more likely to progress further in the switching training. However, there were no significant differences in improvements on the two outcome measures between participants who reached the first level of switching training and participants who reached the third level, which indicates that these improvements were not associated with the degree of difficulty reached and furthermore that these improvements were not due to the training at all.

The battery of different executive assessments used before and following training allowed for the investigation of relationships across these measures. Surprisingly, different executive measures were not significantly correlated with each other once age was partialled out, which is contrary to what has been previously shown in this age group (e.g. Garon et al., 2008; Steele et al., 2012; Wiebe et al., 2011). However, the current sample size was smaller than those in previous work and did not afford the statistical power to perform more sophisticated analyses such as factor analysis. Early numeracy was positively correlated with sustained attention and inhibition over and above age, but not with updating or switching. Additionally, early numeracy prior to training was significantly positively correlated with both early numeracy and switching performance post-training, which suggests that the two constructs may be related longitudinally. However, the lack of a significant correlation between improvements on the switching task and the measure of early numeracy suggests instead that these two skills developed independently of one another. Additionally, FIST and TEMA performance were significantly correlated with each other six months following training, but were not prior to or immediately after training (see Appendix A). These findings highlight the importance of longitudinal designs for elucidating the developmental trajectories of executive processes and mathematics as concurrent patterns may not always be stable across different time points.

3.3.2. Limitations

It is certainly possible that the null training effects could be attributed to a lack of statistical power. However, similar previous studies have found significant effects with comparable sample sizes (Kroesbergen et al., 2014; Passolunghi & Costa, 2014; Witt, 2011). A recent meta-analysis of several hundred published psychology papers calculated the correlation between sample sizes and effect sizes, which should in theory not be related, and found a significant negative correlation between the two, implying that effect sizes may be overestimated in studies with smaller samples (Kühberger, Fritz, &

Scherndl, 2014). They also found that many more published studies reported p-values just below the critical .05 threshold than values just above that cut-off which illustrates that studies with statistically significant results are more likely to be published. Given this publication bias, it is also possible that training studies that have found null results are less likely to be published.

The initial design plan was to train 4- and 5-year-old children only, but, due to challenges with recruitment, the sample ended up including some 3-year-olds as well. Age accounted for a large proportion of the variance in many measures and this may explain the lack of significant correlations between executive measures when age was partialled out. However, age, somewhat surprisingly, did not account for any of the individual differences in training gains. Having children between these ages complete such a large battery of assessments as well as multiple training sessions proved challenging as, on some days, cognitive training was seen as a less appealing option for a 3-year-old compared to, for example, blowing bubbles at the water table with one's friends. There were also limitations of individual tasks used in the assessment battery. In particular, the Backwards Recall subtest of the AWMA was too difficult for this age range and therefore was not a reliable measure of individual differences as the majority of children performed at floor. The AWMA is normed for kids as young as 4 and it is therefore not surprising that the tasks may not be appropriate for 3-year-olds. If the initial plan had been to assess children as young as 3, alternative measures would have been considered.

In addition to reliability concerns with certain tasks, in order to avoid the task impurity problem (Burgess, 1997), each construct should ideally have been assessed using multiple measures. However, this unfortunately was not feasible due to time constraints as the assessment battery was rather lengthy for preschoolers as is. Due to this limitation, one should be cautious in interpreting these findings as it is not possible to know for certain whether each task was a valid measure of the construct it was meant to capture (e.g. whether the spatial conflict task measured inhibitory control). It would be preferable to be able to show that performance on each task was correlated with tasks thought to measure the same construct, but not related to tasks thought to measure different constructs.

Finally, at the time the study was conducted, there was already mounting evidence that cognitive training was not as effective as promising initial results purported it to be. Therefore, it seemed plausible that this study could have null training effects, and this should perhaps have been a deterrent from undertaking such a study. The study was run in spite of this, due to the exciting possibility of addressing theoretical underpinnings of the correlations between executive processes and maths achievement, should training effects be replicated in this sample. Furthermore, more rigorous controls were employed in comparison to previous studies, such as the inclusion of both an active and passive control group, and therefore these findings extend the literature on the effectiveness, or lack thereof of cognitive training.

3.3.3. Implications

Potential implications for education. Computerized training programs have been criticized for lacking ecological validity and neglecting the social aspect of learning (Howard-Jones, 2014). Many children may not be motivated or engaged by sitting in isolation at a computer and playing somewhat repetitive games. An alternative option would be to instead design curricula that engage entire classrooms and have a social component (e.g. Tools of the Mind, Diamond, et al., 2008). Additionally, if executive function training were to be designed with more applied goals, it could be useful to incorporate domain-specific material relevant to the desired outcome into training regimes that have already been shown to have near transfer to executive processes. For example, Kroesbergen and colleagues (2014) incorporated number stimuli in their working memory training and found that the group that completed this training showed greater improvement on a non-symbolic magnitude comparison task than a group that completed domain-general working memory training did. This idea could be extended even further to incorporate arithmetic and other domain-specific skills. However, there is tension between theoretical and applied goals of training studies as designing more educationally relevant training programs would likely lack the degree of experimental control necessary to thoroughly address cognitive psychology theories. If instead one aims to design educational interventions, it could be beneficial to collaborate with educators in order to generate a solution that suits classroom needs. This approach could also be more practical to target children which special learning needs or children at risk as these children may stand to benefit most from intervention.

The lack of significant training effects seen here suggests that intervening in typically developing children may not be useful or necessary. Wilson and colleagues (2009) tested the Number Race training on 5-year-olds from an area associated with low socioeconomic status and did find significant training effects in that sample. Similarly, working memory training was found to be associated with improvements in early numeracy in a sample of low achieving 5-6-year-olds (Kroesbergen et al., 2014). In other words, training may be useful for helping children who are struggling to ‘catch up’ to their typically developing peers but may not boost cognitive skills in children who do not have any learning difficulties.

Implications for theory. The null training effects meant that addressing hypotheses about the mechanism of the relationships between executive processes and math achievement was not possible. However, considering results from the many previously published training studies, as well as unpublished null training findings communicated informally, it is tempting to conclude that there is no convincing evidence that cognitive training of executive processes reliably transfers to other executive processes and, more distally, to academic outcomes. If training one executive process does not show these transfer effects, then what does that imply for theories of executive control? More research is needed to clarify the mechanisms underlying the robust correlations across executive processes as well as between executive control and academic achievement. Based on this correlational evidence, some have proposed that the mechanism through which executive functions influence other cognitive processes is by directing attention and thereby determining which information is filtered into and maintained in working memory (e.g. Posner & Rothbart, 2007). However, the lack of far transfer effects from training attention does not support this notion. Specifically, if efficiently directing attention to relevant information leads to better educational outcomes, one would expect that training attentional control would transfer to academic outcomes. Alternatively, the lack of observed training and transfer effects could be attributed to using measures that are not sensitive enough to training-related changes. Furthermore, in this young age group in particular, it has been speculated based on the correlational evidence that executive control may be important not just for math performance but for learning about maths (Clarke et al., 2013). As training-induced changes are evidence of learning, it was hoped that the current results could shed some light on this hypothesis, but the null effects instead highlight the need to generate better ways of investigating the potential role of executive control in learning.

The null effects of the number sense training contradict previous work showing training-related improvements with the Number Race software. However, these previous studies were conducted in children with dyscalculia (Wilson, Revkin, Cohen, Cohen, & Dehaene, 2006), or children from low socioeconomic background (Wilson et al., 2009), suggesting that, as discussed above, training may be more beneficial for children who are struggling with mathematics. Additionally, while correlational evidence has linked ANS acuity and maths achievement (e.g. Starr et al., 2013), this evidence is inconsistent (De Smedt et al., 2013) and it has not been established whether ANS acuity is causally related to mathematics achievement. Inconsistent ANS training results challenge the assumption that the ANS is the foundation of mathematical thinking as, if number sense truly does underlie mathematics, training it should transfer to mathematics achievement more broadly.

CHAPTER 4: USING ARTIFICIAL LEARNING TO INVESTIGATE ACQUISITION OF SYMBOLIC REPRESENTATIONS

4.1. Introduction

The experiments discussed in this chapter aimed to investigate influences of domain-general and domain-specific factors on learning processes, which so far had eluded either a correlational approach (Chapter 2) or an experimental intervention approach (Chapter 3). Here, I capitalize instead on variations of an artificial learning paradigm as it afforded the opportunity to tackle a specific learning problem. As discussed in Chapter 1, one learning challenge that preschoolers are faced with is acquiring the meaning of numerical symbols (words and digits).

Adults' learning of artificial symbols was explored in Experiments 4[1], 5, and 6[2], as there may be parallels with how children acquire the meaning of real numerical symbols. The stimuli were manipulated to emulate the conditions under which children might learn about numbers. In particular, if, as elaborated in Chapter 1, the ANS and non-symbolic representations do play a role in the acquisition process (e.g. Piazza, 2010), then visual parameters of continuous quantity may influence the formation of symbolic representations, since they influence performance on numerical tasks involving non-symbolic stimuli (e.g. Gebuis & Reynvoet, 2012a). Alternatively, or as well, if the ANS is not the (only) foundation of symbolic acquisition, then the influence of other factors, such as information on their ordinal structure may also play an important role. Artificial symbol learning paradigms allow

[1] This experiment has been published in the *Quarterly Journal of Experimental*
[2] This experiment has been submitted for publication in *ZDM Mathematics Education*

for contrasting different ways of teaching about novel symbols, in order to test which methods lead to more efficient learning.

However, and crucially, adults already have well-established representations of real numerical symbols, which likely influences the way in which they approach the novel learning challenge. Therefore, in Experiment 7, the artificial learning paradigm was modified to be appropriate for 6-year-old children, with the aim of investigating their ability to form novel symbolic representations for large non-symbolic quantities.

4.1.1. Artificial symbol learning in adults

Previous abstract symbol learning studies have shown that adults are able to attach arbitrary symbols to non-symbolic representations of number (Lyons & Ansari, 2009; Lyons & Beilock, 2009; Zhao et al., 2012). For example, Lyons & Ansari (2009) trained participants to associate abstract symbols with the numerical magnitudes of arrays of dots. Performance on a post-training magnitude comparison task as well as on a task in which participants were asked to put the symbols in order from smallest to largest indicated that they had successfully learned to associate the new symbols with magnitudes and also inferred the ordinal properties of the symbols. Artificial learning has also been used to contrast adults' performance when given information on the ordinality versus the magnitude of abstract symbols (Zhao et al., 2012). Participants in the order condition were trained on the spatial order of the symbols whereas participants in the magnitude condition were trained to associate symbols with non-symbolic arrays. After completing the training, participants performed a magnitude comparison task in which two symbols were presented but the magnitude group was instructed to indicate which symbol was larger whereas the order group was instructed to indicate which symbol belonged on the right. Both groups showed distance effects and there was no group difference in error rates on the task, but the magnitude group did perform significantly faster overall. Furthermore, electrophysiological data demonstrated similarities and differences in the neural processes recruited when participants performed the comparison task and an order-priming task. The authors interpreted these findings as suggesting that learning the magnitude of symbols does not necessarily also involve acquiring information about the order. However, the extent to which numerical magnitude and order are dissociable remains unclear, and this is a point I return to in relation to Experiment 6 below.

4.2. Experiment 4. Effect of congruency between discrete and continuous non-symbolic quantity on the formation of novel symbolic representations

In this experiment, we used an artificial learning paradigm to investigate the process of mapping symbols to non-symbolic quantities and to test whether this process could be facilitated by the congruency between total surface area and numerosity of non-symbolic arrays. As elaborated in Chapter 1 and Chapter 2, adults are always influenced by the visual cues associated with continuous quantity, such as total surface area, when making numerical judgments (e.g. Gebuis & Reynvoet, 2012a), and susceptible to conflict between continuous and discrete dimensions of non-symbolic stimuli (e.g. Gilmore et al., 2015; Szűcs, et al., 2013). Therefore, this conflict may affect the process of mapping between symbols and non-symbolic arrays. This also suggests a role for executive attention processes in selecting the relevant dimension of non-symbolic stimuli during the learning task. We hypothesized that participants would form stronger associations between symbols and quantities when the total surface area of the non-symbolic arrays was congruent with numerosity, than when it was incongruent.

A second manipulation was incorporated for the purpose of testing the effect of multisensory information and labels on learning associations between symbols and quantities. Numbers can be represented in three different formats (words, numerals, and non-symbolic arrays), and we included all three in the current study to parallel the information about number that children have access to when learning about numerical symbols. Adults show better perceptual category learning when stimuli were accompanied by redundant verbal labels than when they were not, suggesting that labels allowed participants to represent an ambiguous perceptual division with a more concrete verbal distinction (Lupyan, Rakison, & McClelland, 2007). Furthermore, a study in 3 to 5-year old children found that they performed better on a numerical matching task when given both auditory and visual number stimuli than when given stimuli in just one modality, suggesting that the multisensory information drew their attention to number as the relevant dimension of the stimuli (Jordan & Baker, 2010). Based on these findings, we hypothesized that participants would form stronger associations between symbols and quantities when given auditory labels in addition to visual symbols.

4.2.1. Method

Participants. Forty undergraduate students participated for course credit. They ranged in age from 18 to 37 years (M = 20.57, SD = 2.99) and 11 were male. Written consent was obtained from all participants prior to participation.

Materials. Experimental tasks were presented on desktop computers using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) and all participants wore headphones. Responses were made with keyboard button presses.

Symbolic stimuli. In order to ensure that participants had no previous experience with the symbols, six symbols were selected from a fictional alphabet from the fantasy world of Tékumel (Omniglot, 2013) (see Figure 4.1). The symbols were associated with the numerosities 26, 32, 40, 50, 62, and 75, respectively. In the audiovisual condition, symbols were given monosyllabic nonword names (Seidenberg, Plaut, Petersen, McClelland & McRae, 1994): bip, roo, plap, drow, vime, sny.

Figure 4.1. Symbolic stimuli

Non-symbolic stimuli. The dots were generated using a script written in Python (Price et al., 2012). The image dimensions were 353 x 384 pixels. For 26, 40, and 62, the dots had a total surface area of 6100 pixels (“incongruent condition”, because overall surface area was equated across stimuli, but average dot diameter and numerosity were inversely related). While total surface area was controlled for in these arrays, other visual properties, such as total perimeter and density, were not and it is possible that, for some arrays, some combination of cues was positively correlated with numerosity. However, in a rigorous study controlling for various visual properties of non-symbolic arrays, Gebuis and Reynvoet (2012b) showed that this is unavoidable. Total surface area has been shown to be the most salient visual cue in young children (Rousselle et al., 2004) and so we chose to manipulate congruency on this dimension in particular in the contrast between incongruent and congruent stimuli. Indeed, for 32, 50, and 75, total surface area was correlated with numerosity and the areas were 4880 pixels, 7591 pixels, and 11386 pixels, respectively ("congruent condition"). This manipulation was carried out within-subjects, i.e., intermixing congruent and incongruent stimuli, to obtain data that would be most easily comparable with the published literature (e.g. Gilmore et al., 2013; Szűcs et al., 2013). Spatial locations of the dots varied randomly and each array was only displayed once in the experiment (see Figure 4.2).

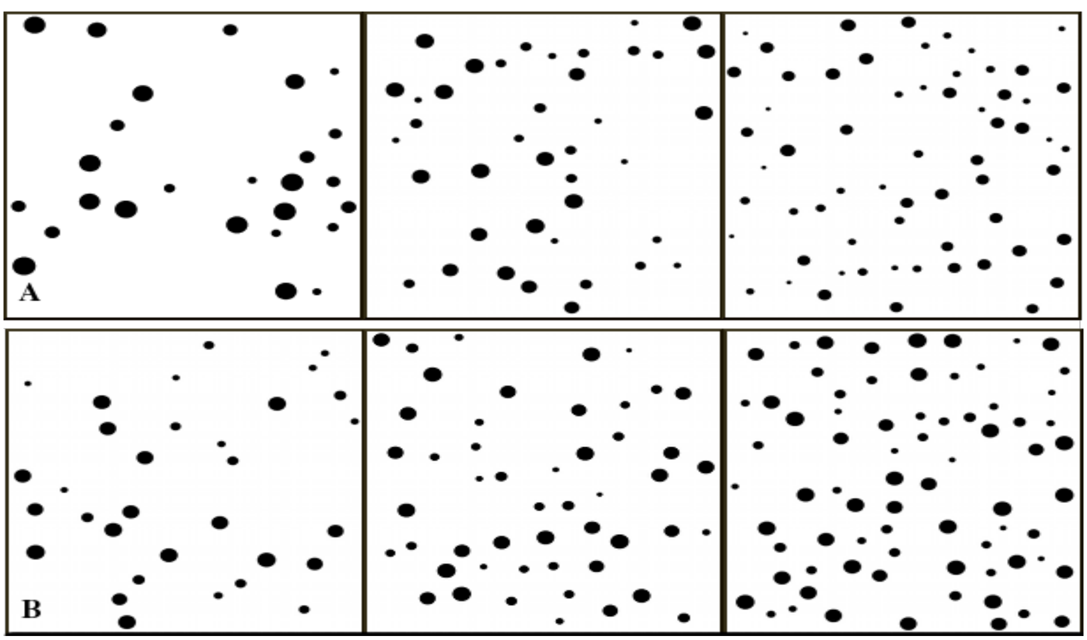

Figure 4.2. Example non-symbolic arrays. A) Incongruent (from left to right: 26, 40, 62). B) Congruent (from left to right: 32, 50, 75)

Procedure. Participants were randomly assigned to the audiovisual or visual only condition. They completed the training session, followed by a magnitude comparison task and an ordinality test.

Training. Participants were first exposed to the symbols in isolation and instructed to try to remember them. They were then informed that the goal of the training was to learn to associate numerical quantities with the symbols. They then saw 264 randomly presented trials. Each consisted of a 1000ms fixation screen followed by the display of a symbol for 1000ms, and subsequently both the symbol and a corresponding dot array were shown for an additional 500ms, requiring no response (see Figure 4.3). Twenty-four check trials were displayed pseudo-randomly throughout the passive viewing trials in an effort to keep participants' attention sustained. On these trials, they were asked to indicate whether a symbol was associated with a certain quantity, and a symbol and dot array were displayed simultaneously for 1500ms. The response window was open until a response was received and feedback was given. No labels were provided on the check trials. The only difference between the visual and audiovisual conditions was that in the audiovisual condition, an auditory label always accompanied the display of a symbol on training trials.

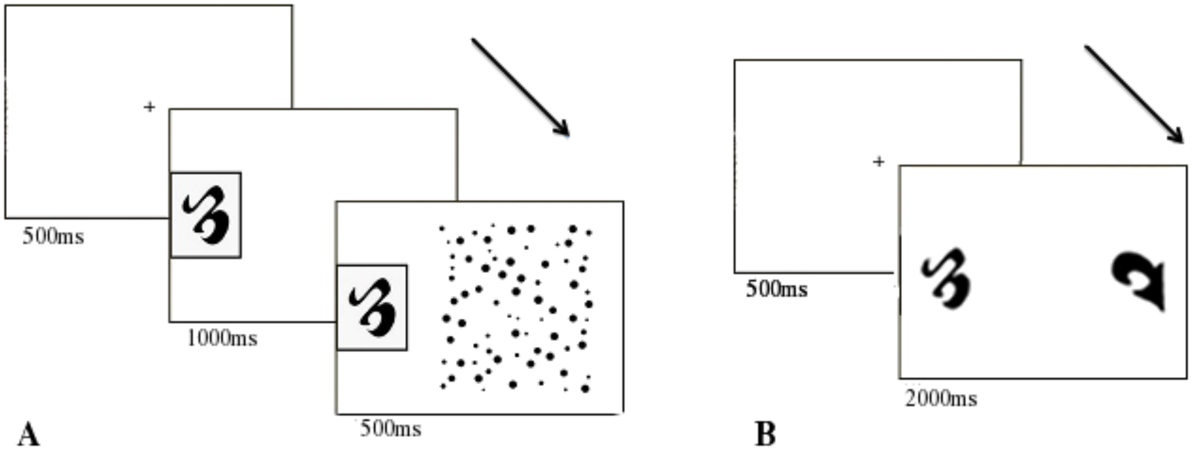

Figure 4.3. A) Timeline of a training trial. B) Timeline of a magnitude comparison trial.

Magnitude comparison task. Participants completed a symbolic magnitude comparison task in order to assess whether they had formed associations between the symbols and non-symbolic arrays that resemble those found for numerals (i.e. ratio effects). In the absence of non-symbolic arrays, they were instructed to choose which of two symbols was associated with the larger quantity. Trials consisted of a 500ms fixation screen followed by two symbols displayed for 2000ms and participants had to respond during that time. Each possible pairing of symbols was presented with display side of the screen counterbalanced, for a total of 30 pairs. These pairs were each presented four times for a total of 120 trials presented in 2 blocks of 60 trials with a break in between. The ratio between symbols was considered small if it was less than 0.53, medium if it was between 0.53 and 0.7 and large if it was greater than 0.7.

Pairs for which the symbol associated with the larger quantity was also associated with the greater total surface area during training were labeled as congruent because surface area and numerosity pointed comparison responses in the same direction. The remaining pairs were labeled as incongruent because of the contrast between total surface area and numerosity. More specifically, pairs that had both been associated with congruent arrays were considered congruent comparisons and pairs that had both been associated with incongruent arrays were considered incongruent comparisons. For the pairs that were made up of one symbol from each category, if the numerically larger of the two had been associated with more total surface area, it was considered a congruent comparison, and if the numerically larger of the two had been associated with less total surface area, it was considered an incongruent comparison, again because area and number indexed a different comparison response. Of note, the non-symbolic arrays were not available during the magnitude comparison task.

Ordinality test. Participants were given a pen and paper ordinality test. The six symbols appeared on the page with a corresponding letter and participants were asked to order the symbols from smallest to largest by writing the letters in the blanks provided.

4.2.2. Results

Magnitude comparison. One participant was excluded from this analysis due to missing data. Pairs for which the symbol associated with the larger quantity was also associated with the greater total surface area during training were labeled as congruent because surface area and numerosity pointed comparisons in the same direction and the other pairs were labeled as incongruent because of the contrast between total surface area and numerosity. Of note, the non-symbolic arrays were not available during the magnitude comparison task. A mixed factorial ANOVA was performed with Congruency and Ratio (small, medium, or large) as within subjects factors and Condition (visual or audiovisual) as a between subjects factor. The assumption of sphericity was violated for the area by ratio interaction as well as for ratio and so the Greenhouse-Geisser correction was used. Results revealed a significant main effect of congruency, $F(1,37) = 10.3$, $p = .003$, $\eta^2_p = .218$, a significant main effect of ratio, $F(2,58.02) = 70.18$, $p < .001$, $\eta^2_p = .655$, and a significant interaction between the two, $F(2,55.19) = 14.93$, $p < .001$, $\eta^2_p = .288$ (see Figure 4.4). An analysis of simple main effects showed that this interaction was driven by participants performing significantly better on large ratio congruent trials than on large ratio incongruent trials, $p < .001$, but not significantly differently across congruent and incongruent trials for the medium and small ratio trials. There were also significant differences between all levels of ratio in both congruency conditions, with participants performing better on small ratio trials than on medium ratio trials and better on medium ratio trials than on large ratio trials. No other main or interaction effects reached statistical significance.

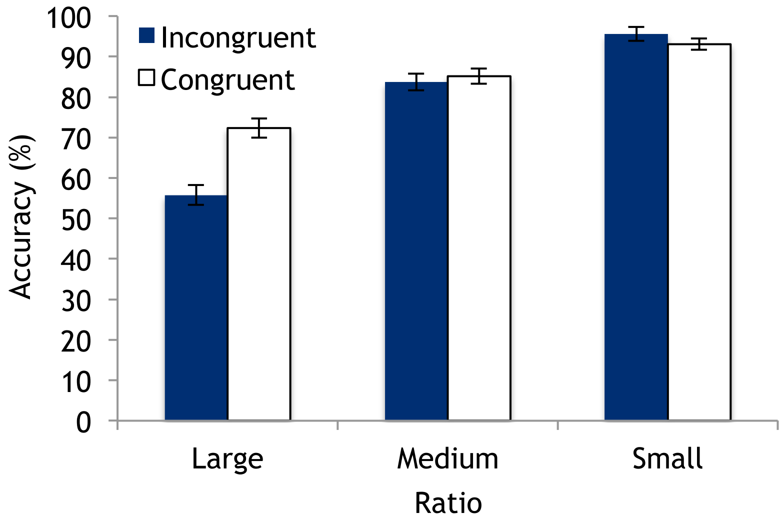

Figure 4.4. Magnitude comparison accuracy for different ratio and congruency trials. Error bars indicate standard errors.

A mixed factorial ANOVA was performed on mean reaction times for correct responses with Congruency (congruent or incongruent) and Ratio (small, medium, or large) as within subjects factors and Condition (visual or audiovisual) as a between subjects factor. Outliers above and below two standard deviations from the mean were excluded for each participant for each condition. Results revealed a significant main effect of Congruency, $F(1,32) = 13.01$, $p = .001$, $\eta^2_p = .289$, a significant main effect of Ratio, $F(2,48.7) = 14.31$, $p < .001$, $\eta^2_p = .309$, and a significant main effect of Condition, $F(1,32) = 7.03$, $p = .012$, $\eta^2_p = .18$. The interaction between Congruency and Ratio showed a trend towards significance, $F(1,58.35) = 2.49$, $p = .096$. Participants responded significantly faster on congruent trials (*M* = 773.56, *SE* = 24.31) than on incongruent trials (*M* = 819.52, *SE* = 28.2), and significantly faster on small ratio trials (*M* = 757.44, *SE* = 27.01) than on medium (*M* = 806.62, *SE* = 24.69), $p <$

.001, and large ratio trials (*M* = 825.57, *SE* = 24.69), $p < .001$. Mean overall reaction time was significantly faster in the visual condition (*M* = 728.83, *SE* = 35.04) than in the audiovisual condition (*M* = 864.25, *SE* = 37.17).

Ordinality. Two participants were excluded from this analysis due to missing data. The proportion of participants who correctly placed each symbol on the ordinality test is reported in Figure 4.5. Each participant received a score of 1 for correctly ordering each symbol or 0 for an incorrect response. The mode for overall accuracy on the ordinality test was 67%, but accuracy was highest at the two extremes of the ordinal sequence, 26 and 75.

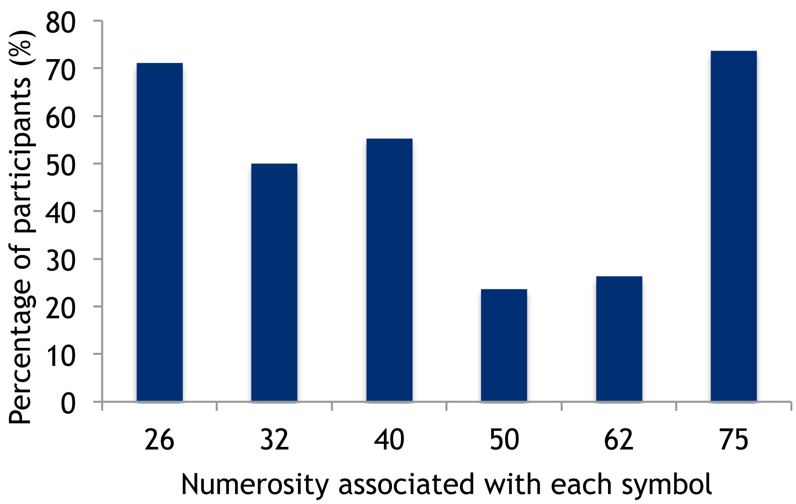

Figure 4.5. Percentage of participants who placed each symbol in the correct order.

The number of participants who swapped adjacent symbols was analyzed in order to investigate patterns of errors. Friedman's test revealed a significant difference across swap errors, $x^2(2) = 19.3$, $p < .001$. At the two extremes, participants were unlikely to swap 25 and 32 (N = 3), and 62 and 75 (N = 2). Wilcoxon's signed rank tests revealed that more participants swapped 50 and 62 (N = 17) than swapped 32 and 40 (N = 8), Z = -2.5, $p = .013$, or 40 and 50 (N = 1), Z = -3.77, $p < .001$. The symbol for 50 had been associated with non-symbolic arrays that had a greater overall surface area than arrays that had been associated with the symbol for 62 during training, whereas for the other two pairs (50 and 40; 32 and 40), the symbol associated with the larger number had also been associated with more surface area during training.

Training and individual differences. In order to investigate whether the visual and audiovisual groups differed in their performance on the training check trials and whether participants improved over the course of training, a mixed factorial ANOVA on check trial reaction time, excluding errors, was run with Condition as a between subjects variable and Time as a within subjects factor. The first twelve check trials were grouped into time one and the second half into time two. There was no effect of Condition, $F(1,37) < 1$,, but there was a significant main effect of Time $F(1,37) = 7.05$, $p = .012$, $\eta^2_p = .16$. Participants were significantly faster on the second half of check trials ($M = 1598.9$, $SD = 142.64$) than on the first half ($M = 1826.81$, $SD = 118.65$). The same ANOVA was run on check trial accuracy. The effect of Condition again failed to reach significance, and the effect of Time showed a trend towards significance, $F(1,37) = 3.3$, $p = .077$. Accuracy in the first half of trials ($M = 56.38\%$, $SD = 16.51\%$) tended to be lower than accuracy in the second half of trials ($M = 63.1\%$, $SD = 16.34\%$).

Accuracy on the first half of the check trials was not correlated with any of the post-training tasks, nor was it correlated with accuracy on the second half of the check trials. Accuracy on the second half of the check trials was significantly correlated with overall accuracy on the post-training magnitude comparison task $r(38) = .404$, $p < .012$, but not on the ordinality task. Overall accuracy on the ordinality test was positively correlated with overall accuracy on the magnitude comparison task, $r(37) = .615$, $p < .001$.

4.2.3. Discussion

Congruency between stimulus dimensions influenced symbolic comparison for trials in which the ratio between the associated numerosities was large, as indexed by higher accuracy for trials in which area and numerosity were congruent than for the incongruent trials, but for large ratio trials in particular. This pattern suggests that participants were reliant on visual cues associated with the continuous quantity dimension of non-symbolic arrays when making the most difficult symbolic comparisons. Crucially, the visual cues influenced performance on the symbolic comparison task, during which participants were not given any information about continuous quantity, suggesting that their symbolic representations of number included this information. Furthermore, in explicit ordinality judgments, symbols for which total surface area was incongruent with ordinal placement resulted in a greater number of ordinal errors.

These findings provide additional support for the argument that the effects of continuous visual parameters on discrete numerical operations should be investigated rather than controlled for (e.g. Cantrell & Smith, 2013; Szűcs et al., 2013). According to Leibovich and Henik’s (2013) model, adults should be able to rely solely on discrete quantity to make non-symbolic magnitude judgments, but here we found that they were influenced by visual cues from continuous quantity. Previous studies on the reciprocal interference of discrete and continuous quantity during magnitude comparison have found mixed results. One study reported that adults are equally influenced by congruence whether making judgments on number or area (Hurewitz, Gelman, & Schnitzer, 2006) and another found that adults are more susceptible to conflict from discrete number when making area judgments than the reverse (Nys & Content, 2012), which led the authors to conclude that discrete number is more salient in adults. The current results are convergent with the former finding and demonstrate that even adults are susceptible to interference from continuous quantity cues. As children are even more susceptible to this interference (Szűcs et al., 2013), this suggests that the congruence between continuous and discrete quantity is likely to have a larger influence on the formation of symbolic numerical representations in children. Furthermore, this suggests that inhibitory control may play a role in symbol acquisition as it is required for resolving conflict between continuous and discrete dimensions of non-symbolic quantity.

Accuracy on the second half of the check trials was correlated with accuracy on the symbolic comparison task, indicating that participants did learn to associate the symbols with the arrays over the course of training. Crucially, ratio effects, which are the signature of the ANS, were seen on the symbolic comparison task, suggesting that participants treated the newly learned numerical symbols in the same way they would Arabic digits. This is in line with what Lyons and Ansari (2009) found using a similar paradigm. Accuracy on the ordinality test was variable and overall performance was not very high, but participants were more likely to correctly place the symbols at the extremes than other symbols, suggesting that they were able to infer order based on the training phase. This is convergent with results of previous similar studies (Lyons & Ansari, 2009; Lyons & Beilock, 2009), but contrary to what Zhao and colleagues (2012) found using a different test of ordinality. Intriguingly, in the current experiment, participants' errors in ordinality judgments again suggested that continuous dimensions associated with non-symbolic stimuli during training affected their ordinal inferences. Many participants switched the order of the symbols for 50 and 62, as 50 had more total surface area. Interestingly, the symbol for 32 also had less area than the one for 26, but participants did not seem to be influenced by the incongruence in this case as most still ordered those symbols correctly. It could be that the symbols on the extremes of the range are treated differently and that being confident in the order of the symbols at either end of the range prevented participants from paying as much attention to the associated non-symbolic magnitudes.

There were no significant differences in magnitude comparison accuracy between the visual only and audiovisual conditions. This counters predictions originating from the benefits of labels on adults' category learning (Lupyan et al., 2007) and of multisensory information on children's performance on a numerical matching task (Jordan & Baker, 2010). It could be that these arbitrary labels did not facilitate the formation of symbolic representations because they did not provide any numerical information. Some participants in the visual only condition reported generating their own labels as mnemonic devices for recognizing the abstract symbols, and this may have worked in the same way as the non-word labels did.

Differences across reaction time and accuracy measures were seen not only with the group differences but also with regards to the interaction between congruency and ratio, which reached significance for accuracy but trended towards significance for reaction time. A possible interpretation is that reaction time is indicative of the speed of access to a representation whereas accuracy measures the precision of that representation (Prinzmetal, McCool, & Park, 2005). This fits with the idea that participants in the visual only group were able to access the newly formed representations more quickly than those in the audiovisual group, but the two groups did not differ in terms of the precision of these representations. It could also be that participants had more precise representations for congruent trials, but that they did not differ as much in terms of how quickly they were able to access these representations. This is further supported by the fact that many participants swapped 50 and 62 on the ordinality test, suggesting that they may have had imprecise representations for those symbols in particular.

Of note, even participants who performed very well on the post-training tasks reported that they did not think they had learned to associate the symbols with numerical magnitudes. A limitation of the experiment is that we did not assess for how long participants maintained the newly learned symbol associations. Testing participants on the symbolic comparison task some time following the training would allow us to see how robust the new numerical representations were and further investigate whether attaching symbols to non-symbolic representations is a viable learning strategy for forming symbolic representations.

4.3. Experiment 5. The influence of ordinality on the formation of novel symbolic representations

The null effects of providing auditory labels for the symbols in Experiment 5 led to further questions about how numerical representations can be formed from multiple sources of information. Wiese (2007) argued that language is the basis of numerical thinking as words provide information on ordinal and nominal properties of number in addition to cardinality. In Experiment 4, the auditory labels did not provide any numerical information, but simply provided a name for the associated symbol. Children learn the count sequence by rote before they understand the cardinality of each number word (Wynn, 1992) and it has been proposed that the words act as placeholders (Carey, 2009). By this account, children align the words with representations of non-symbolic quantities once they realize the common order, suggesting an important complementary role for ordinal information (e.g. Wiese, 2003). Furthermore, numerical ordering ability has been shown to mediate the relationship between magnitude comparison and math achievement (Lyons & Beilock, 2011), which suggests that the ordinal property of numbers links the ANS with symbolic numbers.

The goal of Experiment 5 was to test whether giving participants explicit information about ordinality could facilitate artificial symbol learning. In particular, we aimed to explore whether participants could learn an artificial count sequence and how this would influence the formation of numerical representations. Unlike previous artificial learning studies in which participants were taught the spatial order of new symbols (Zhao et al, 2012), or learned the order based on feedback on a comparison task (Tzelgov, Yehene, Kotler, & Alon, 2000), here information about order was provided aurally. This was designed to be analogous to how children first learn the count sequence.

4.3.1. Method

Participants. Thirty-three undergraduate students participated for course credit. They were between the ages of 18 and 24 (M = 20.2, SD = .98), and 11 were male. Written consent was obtained from all participants prior to participation.

Materials. As in experiment 4, experimental tasks were presented on desktop computers using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) and all participants wore headphones. Responses were made with keyboard button presses.

Symbolic stimuli. The symbolic stimuli described in Experiment 1 were used with the addition of one symbol (see Figure 4.6) and a corresponding nonword label: flun.

Figure 4.6. The additional symbol

Nonsymbolic stimuli. The dots were generated using the same script as in experiment 4. The image dimensions were 353 x 384 pixels. For 26, 40, 50, and 75 the dots had a total surface area of 6100 pixels (“incongruent condition”). For 32, 62, and 93, the total surface area was correlated with the numerosity and the areas were

3795 pixels, 7591 pixels, and 10844 pixels, respectively ("congruent condition"). Spatial locations of the dots varied randomly and each array was only displayed once in the experiment.

Procedure. Participants were randomly assigned to the order or control condition and completed the training session followed by a magnitude comparison task and an ordinality test.

Training. Participants were first exposed to the symbols paired with the associated labels and instructed to try to remember them. In this phase, both groups saw the symbols in the correct numerical order (sny, drow, bip, flun, roo, plap, vime) but the order group was told this was similar to the count sequence whereas the control group was not. They were then informed that the goal of the training was to learn to associate numerical meaning with the symbols. They then saw 252 randomly presented trials. The timing of the trials was the same as in Experiment 4. The order group saw the trials in pseudo-random order, as they always saw trials in the same sequence as the "count" sequence. Twenty-eight check trials were displayed pseudo-randomly throughout the passive viewing trials in an effort to keep participants' attention sustained. On these trials, participants in the control group were asked to indicate whether a symbol was associated with a certain quantity, and a symbol and dot array were displayed simultaneously for 1500ms. Participants in the order group were shown 7 boxes in a horizontal line with a symbol in one box for 1500ms and asked to indicate whether the symbol was in the correct ordinal position (see Figure 4.7). In both conditions the response window was open until a response was received and feedback was given. The symbols and labels were paired throughout training in both conditions.

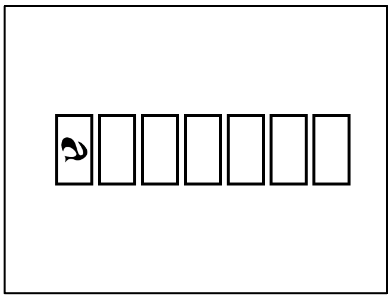

Figure 4.7. Example check trial for the order training.

Magnitude comparison task. This was the same as in experiment 4 except the ratio between symbols was considered small if it was less than 0.54, and large if it was greater than 0.64, in order to have more even numbers of trials in each condition. The task comprised of two blocks of 84 trials with a break in between.

Ordinality test. Participants were given a pen and paper ordinality test like the one used in experiment 4, but with the additional symbol. They were also asked to write the non-words in order from smallest to largest. Eighteen of the participants completed the ordinality test a second time, three weeks following the initial testing session.

4.3.2. Results

Magnitude comparison. Accuracy data on the magnitude comparison task was analysed with a mixed factorial ANOVA with Ratio (small or large) and Congruence as within subjects variables and Condition (order or control) as a between subjects variable. Results revealed a main effect of Ratio, $F(1,31) = 47.27$, $p < .001$, $\eta^2_p = .6$, which reflected higher accuracy on small ratio (*M* = 74.6%, *SE* = 2.8%) compared to large ratio (*M* = 64%, *SE* = 2.2%) trials (see Figure 4.8). There were also main effects of Congruence, $F(1,31) = 12.94$, $p = .001$, $\eta^2_p = .3$, with higher accuracy on congruent (*M* = 74.8%, *SE* = 3.2%) than incongruent (*M* = 63.8%, *SE* = 2.5%) comparisons, and Condition $F(1,31) = 7.5$, $p = .01$, $\eta^2_p = .2$, with performance being better in the order condition (*M* = 75.9%, *SE* = 3.4%) than the control condition (*M* = 62.8%, *SE* = 3.3%). Levene's test for equality of variances revealed that the assumption of homogeneity of variance was violated for all conditions except for congruent large ratio trials. Kolmogorov-Smirnov tests revealed that the assumption of normality was also violated for accuracy on small ratio congruent trials and large ratio incongruent trials, suggesting caution when interpreting parametric statistics. Non-parametric tests provided converging results for higher mean accuracy in the order condition compared to the control condition, $Z = -2.4$, $p = .017$. Accuracy was significantly higher on congruent than incongruent trials for both large, $Z = -4.05$, $p < .001$, and small ratio trials, $Z = -2.45$, $p = .014$.

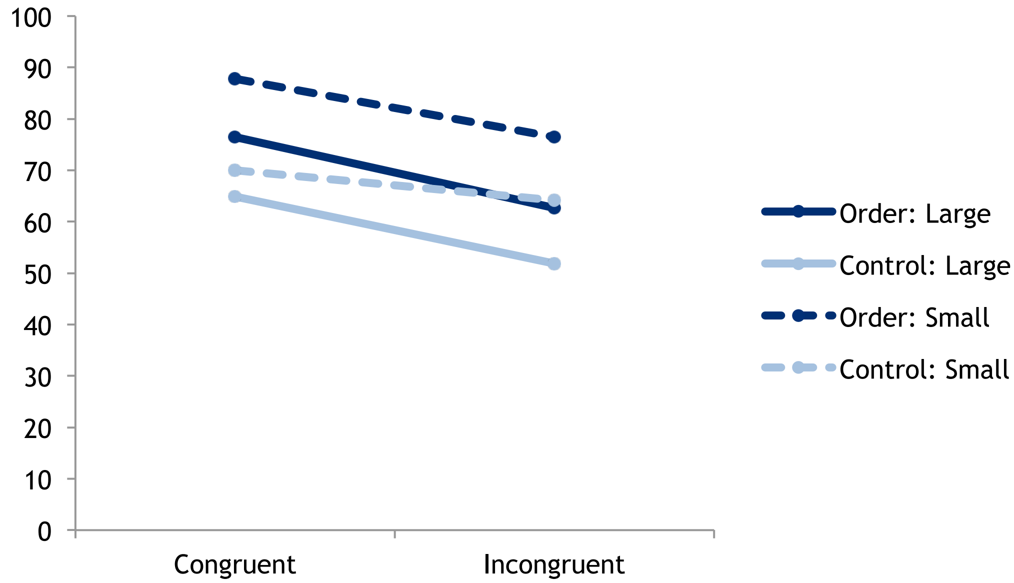

Figure 4.8. Magnitude comparison accuracy separated by ratio and condition.

The same ANOVA was run on reaction time data for correct responses, and results revealed a slightly different pattern of results. There were again significant main effects of Congruency, $F(1,31) = 26.67$, $p < .001$, $\eta^2_p = .46$, and Ratio, $F(1,31) = 6.84$, $p = .014$, $\eta^2_p = .18$, but the main effect of Condition did not reach significance, $F(1,31) = 3.04$, $p = .091$. Furthermore, there were significant interactions between Congruency and Ratio, $F(1,31) = 7.5$, $p = .01$, $\eta^2_p = .2$, as well as Ratio and Group, $F(1,31) = 7.5$, $p = .01$, $\eta^2_p = .2$. Pairwise comparisons revealed that, in the order condition, response times were significantly faster on small ratio trials ($M = 802.3$, $SE = 47.44$) than large ratio trials ($M = 879.9$, $SE = 54.67$), $p < .001$, but there was no significant difference between small ($M = 720.74$, $SE = 46.03$) and large ($M = 717.63$, $SE = 53.04$), $p = .877$, in the control condition. For large ratio trials, response times were significantly shorter in the control condition, than in the order condition, $p =$

.041, but there was no significant difference for small ratio trials, $p = .227$. For congruent trials, response times were significantly shorter on small ratio ($M = 695.39$, $SE = 27.54$) than large ratio ($M = 771.1$, $SE = 34.53$), $p < .001$, but there was no significant difference between small ($M = 826.43$, $SE = 42.4$) and large ratio ($M = 827.64$, $SE = 43.52$) on incongruent trials, $p = .961$. There were significant effects of congruence for both small and large ratio trials, $p < .001$. From inspecting Figure 4.9, it appears that these interaction effects may all be driven by the reaction times on large ratio trials in the control condition. Accuracy on large ratio incongruent trials in this condition was 51.88%, suggesting that not all participants performed above chance, which suggests caution when interpreting reaction times.

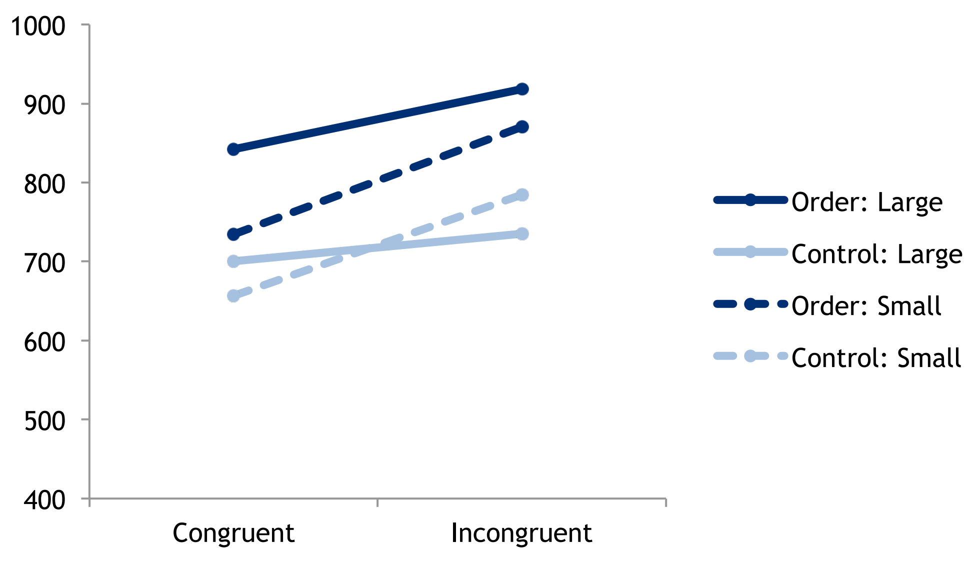

Figure 4.9. Magnitude comparison reaction time separated by ratio and condition.

Ordinality. Two participants were excluded due to missing data. Independent samples t-tests revealed that participants in the order condition were much more accurate than those in the control group at both ordering the symbols, $t(25.18) = 3.67$, $p = .001$, and the labels, $t(29) = 5.69$, $p < .001$ (see Figure 4.10). A Mann-Whitney test showed that significantly more participants in the order condition (N = 11) than in the control condition (N = 4) accurately mapped all of the symbols to their corresponding labels, $Z = -2.31$, $p = .021$. Performance on the ordinality test three weeks following the experimental was overall quite low and highly variable for both symbols (M = 43.65%, SD = 38.1%) and labels, (M = 38.1%, SD = 33.23%). Condition differences did not reach significance for this follow up ordinality test.

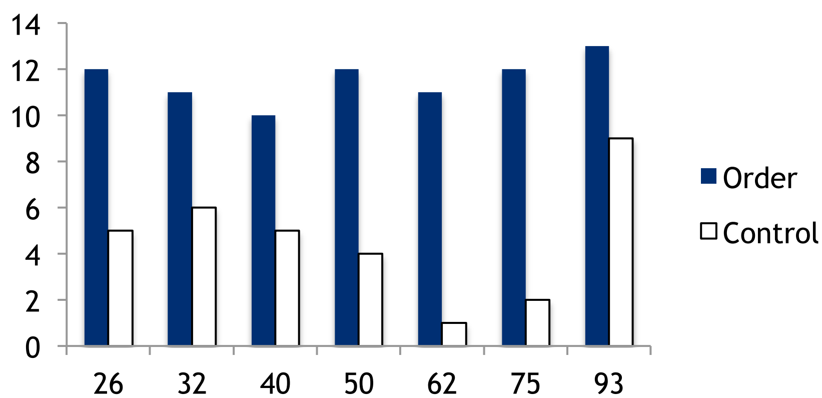

Figure 4.10. Number of participants who placed each symbol in the correct order

Correlations. Bivariate correlations between accuracy on training check trials and accuracy on post-training assessments are reported in Table 4.1. Magnitude comparison accuracy was also significantly positively correlated with accuracy both on symbol ordering, $r(31) = .66$, $p < .001$, and label ordering, $r(31) = .53$, $p = .002$.

Table 4.1. Bivariate correlations between training accuracy and post-training assessments.

	Training Accuracy
Magnitude Comparison Accuracy	.6**
Symbol Order Accuracy	.78**
Label Order Accuracy	.75**
Symbol Order Accuracy (Follow up)	.51*
Label Order Accuracy (Follow up)	.42

4.3.3. Discussion

The results of Experiment 5 supported our hypothesis that information about the order of numerical symbols would facilitate the formation of symbolic representations. Participants given aural cues about order were more accurate at the post-training comparison task than participants who were not. This experiment also provided converging evidence with the finding of Experiment 4 that congruency between continuous and discrete quantity influenced symbolic representations. This effect was seen even in the group that was given the exact order of the symbols. Unlike in the previous experiment, there was a significant main effect of congruency, rather than an interaction between ratio and congruency. Here, participants were more accurate overall on congruent comparisons compared to incongruent comparisons. It could be that the addition of a seventh symbol made the task more difficult and therefore adults were more susceptible to congruency than in the previous experiment.

As in Experiment 4, discrepancies were again seen between analyses on accuracy compared to reaction time data. In particular, the difference between learning conditions did not reach significance on reaction time. Accuracy data suggested that performance was not above chance for some comparison types and therefore reaction time may not be a reliable measure. As discussed above, reaction time is thought to measure speed of access, whereas accuracy is thought to measure precision of representations (Prinzmetal et al., 2005), therefore if participants were not performing above chance and likely guessing under some conditions, reaction times were not tapping into speed of access. Furthermore, accuracy was low on a follow up ordinality test three weeks following the training, suggesting that participants did not retain the novel symbolic representations long term.

4.4. Experiment 6. Electrophysiological indices of artificial learning

Results of Experiment 5 suggested that participants' symbolic representations were influenced by non-symbolic magnitude, even when they were taught the exact order of symbols. In Experiment 6, we therefore wanted to test whether adults could form symbolic representations on the basis of ordinality information alone. Furthermore, we used electroencephalography (EEG) in order to investigate whether electrophysiological indices of comparing the novel symbols were comparable to those observed previously when comparing Arabic numerals.

Evidence from electrophysiological studies has shown ratio effects in scalp-recorded event related potentials (ERPs) associated with numerical comparison tasks

(Dehaene, 1996; Libertus, Woldorff, & Brannon, 2007; Temple & Posner, 1998). Specifically, the amplitude of ERP components approximately 200ms after stimulus onset in electrodes over inferior parietal and occipital cortex was significantly larger for large ratio comparisons than for small ratio comparisons for both adults and 5-year-old children (Temple & Posner, 1998), a finding interpreted as suggesting that children's symbolic representations were similar to adults'. Furthermore, these electrophysiological effects were seen regardless of whether adults were comparing Arabic numerals or non-symbolic dot arrays (Libertus et al., 2007). The fact that these behavioural and neural ratio effects are seen independent of representational format has been taken as evidence that symbols are tightly linked to corresponding approximate quantities. However, Lyons and colleagues (2012) proposed that processing numerical symbols does not entail automatically activating an approximate representation of the corresponding quantity, but rather that symbolic processing is dependent on the ordinal relations between the symbols. More recently, the close coupling between non-symbolic and symbolic representations of number has been questioned even in younger children (e.g. Sasanguie, Defever, Maertens, & Reynvoet, 2013), a point I return to in Chapter 5. Given this potential for symbolic estrangement from very early in childhood, it remains unclear whether approximate magnitude information plays a functional role in the acquisition of novel symbols. Can numerical representations be formed in the absence of information about numerical magnitude?

Behavioural and neural ratio effects are also seen when participants are instructed to judge whether numerical symbols are in the correct order (Turconi, Campbell, & Seron, 2006), as all as when comparing non-numerical ordered stimuli, such as letters (e.g. Szűcs & Csépe, 2004; Turconi, Jemel, Rossion, & Seron, 2004). Converging evidence from an electrophysiological investigation showed differences in the timing and topography of distance effects of ERPs across three tasks: numerical magnitude comparison, numerical order comparison, and letter order comparison (Turconi et al., 2004). Specifically, for letters, there were effects of distance over bilateral parietal electrodes for early and later components, whereas, for numbers, there was a left lateralized effect of distance at parietal electrodes for magnitude comparison, and a slightly later bilateral effect for order judgments. For all three tasks, there was an effect of distance on the P3 component at frontal and parietal electrodes. In a similar study, judgments of numerical and alphabetical order revealed similar behavioural effects of distance, as well as distance effects irrespective of stimuli type on the latencies of the P2p and N2p components, but stimulus-specific effects on the amplitude of the right parietal N2p (Szűcs & Csépe, 2004). Taken together, this suggests that, even when behavioural effects indicate similarities in the effects of magnitude and order on numerical representations, investigating temporal dynamics can differentiate judgments of numerical magnitude from those of order. Of note, ERP differences across letter and numerical comparisons are difficult to interpret, because of the basic perceptual differences across letter and number stimuli. Furthermore, as the studies above investigated judgments involving numerals or letters, for which adults have well-established symbolic knowledge, those data do not provide insights into the role of magnitude and order information in the acquisition process.

In Experiment 6, we used artificial symbol learning in order to test differential influences of order and magnitude information on adults' formation of symbolic representations. Information about order was again provided by auditory verbal cues in order to be analogous with how children learn the count sequence. We used ERPs to test: a) whether processing of the newly learned symbols shows electrophysiological markers that are similar to real numbers; and b) whether electrophysiological ratio effects are modulated by whether participants' symbolic representations were formed on the basis of magnitude or ordinal information. We hypothesized that, if the distance effects observed on numerical magnitude comparison tasks are underpinned by overlapping approximate representations of quantity, then participants in the magnitude learning condition will exhibit these effects both on behaviour and in their ERPs when performing a comparison task with the learned symbols. In contrast, participants in the order learning condition will not show these markers of ratio effects, as they would have no sense of the approximate quantity represented by the symbols. Additionally, we hypothesized that participants in the order condition would perform better overall on the tasks with the learned symbols, as they would have a more precise understanding of the relations between the symbols than those in the magnitude condition.

4.4.1. Method

Participants. Thirty-two adults participated (M_{age} = 20, SD = 3), twenty-eight of which were female. University of Oxford undergraduate students participated for course credit and a few others from the community participated for monetary compensation. Two participants were excluded due to technical difficulties with EEG recording and one was excluded for failing to perform above chance on the post-training assessment.

Materials. Experimental tasks were presented on a computer using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Responses were made with keyboard button presses of ‘s’ and ‘l’ keys.

Stimuli. Five symbolic stimuli and five non-word labels were used (see Figure 4.11). Non-symbolic arrays of gold coins, as opposed to black dots, were generated using the same Python script as in the previous two experiments. The average individual coin size was held constant at 187 pixels so that total surface area increased as numerosity increased and therefore provided converging magnitude cues.

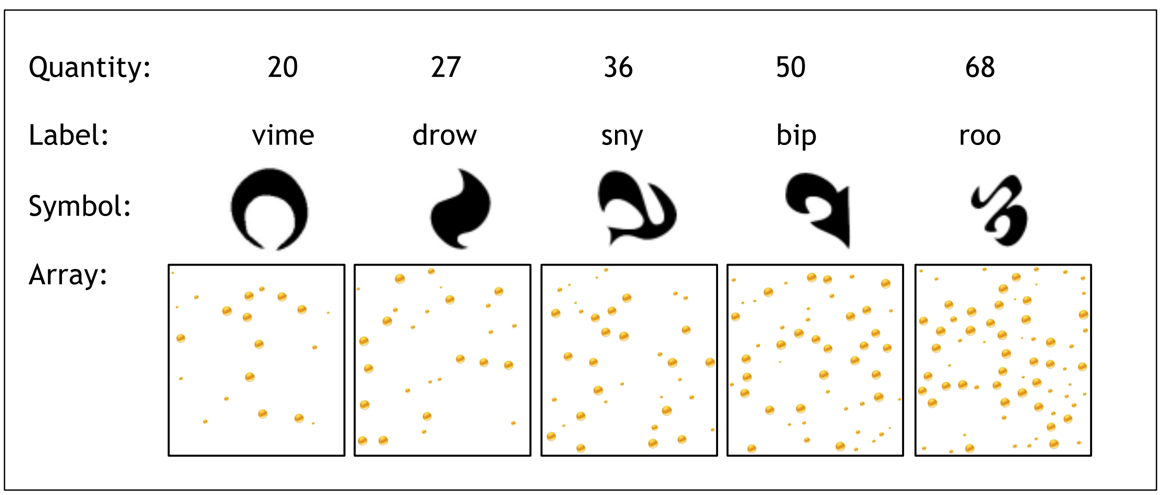

Figure 4.11. Stimuli in the learning phase.

Procedure. Participants were randomly assigned to the order condition or the non-symbolic magnitude learning condition. They first completed the training session, which lasted approximately 15 minutes. High-density EEG nets were then placed on participants and they completed a magnitude comparison task. Afterwards, they completed an ordinality test. The entire session took just over an hour.

Training. Participants in both groups were first exposed to each symbol paired with the associated label, displayed 4 times each, and instructed to try to remember them. The symbols were presented in the correct numerical order in the order condition, but in random order in the magnitude condition. Those in the order condition were also taught that the order was from smallest to largest. For participants in the order condition, the trials were presented in pseudo-random order, as they were always presented in the correct order sequence. Trials consisted of a fixation cross presented for 1000ms, followed by a symbol displayed for 1000ms. For participants in the magnitude condition, the randomly presented trials consisted of a fixation cross displayed for 1000ms, followed by a symbol displayed for 2000ms, and a corresponding non-symbolic array displayed for 1000ms, beginning 1000ms after the symbol was presented. There were 100 training trials and auditory labels were presented with the corresponding symbols throughout training in both conditions. Twenty check trials were displayed pseudo-randomly throughout the passive viewing trials in an effort to keep participants engaged. On these trials, participants in the order condition were shown 5 boxes in a horizontal line with one box containing a symbol and asked to indicate whether the symbol was in the correct ordinal position (see Figure 4.12). Participants in the magnitude condition were shown a symbol and a non-symbolic array and asked to indicate whether the symbol and array were paired correctly. In both conditions, check trials were displayed for a maximum of 2500ms, and participants were given feedback.

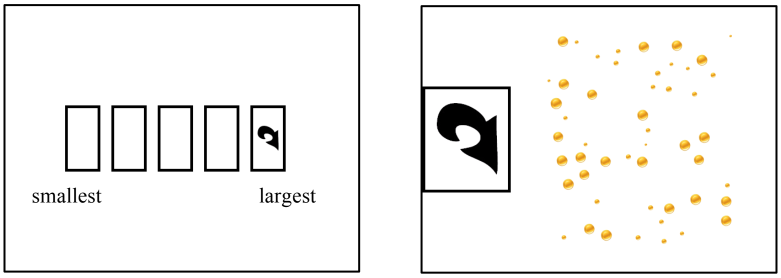

Figure 4.12. Check trials. Order condition on the left and magnitude condition on the right.

Magnitude comparison task. Participants were instructed to indicate whether a centrally presented learned symbol was bigger or smaller than an anchor symbol. The anchor was consistently the symbol that came third in the order of the symbols (and was associated with 36 coins in the magnitude condition). Response hands were counterbalanced between participants. The fixation was jittered between 500 and 600ms, and the trial until a response was made, or 2000ms at most. Each of the 4 other learned symbols was displayed 16 times for a total of 64 trials. The ratio between the symbol and the anchor was considered small if the ratio between the number of coins associated with each symbol in the magnitude condition was approximately .5 (e.g. the ratio between 36 and 68 is .53), and large if it was approximately .75 (e.g. the ratio between 36 and 50 is .72). Note that participants in the order condition did not have access to this magnitude information during training, and so small ratio trials are more accurately defined as having a distance of 2 from the anchor, and large ratio trials had a distance of 1 from the anchor. However, for simplicity, I refer to this distinction as ratio henceforth.

Ordinality test. This was the same as in Experiment 5.

EEG recording and processing. EEG signals were recorded with a 128-channel Hydrocel Geodesic Sensor Net connected to Net Amps 300 (ElectricalGeodesicsInc., Eugene, OR) and with NetStation 4.5 software. EEG signal was referenced online to the vertex and was sampled at 250Hz. Electrode impedances were kept below 50kΩ as recommended by the manufacturer. Eye movements and eye blinks were monitored with six eye channels placed on the outer canthi of both eyes and above and below the eyes. EEG data were filtered offline with band-pass filter 0.1-30Hz, rereferenced to the average reference, and segmented into specific time epochs that began 200ms before the onset of the symbol display and ended 800ms after. Artifacts were removed using NetStation's artifact detection routines and bad channels were replaced using spherical splines interpolation. ERP data were baseline-corrected using a 200ms pre-stimulus interval. Further ERP analyses were carried out on those trials with correct behavioural responses only.

ERP analyses. Statistical analyses focused on two components: N1 and P2p. Regions of interest and time windows for each component were selected a priori based on Temple and Posner (1998) and were defined as follows: N1 (120-170ms post-stimulus) and P2p (180-230ms post-stimulus); twelve electrode pairs were chosen in the inferior parietal and occipital regions: 67 and 77; 66 and 84; 60 and 85; 65 and 90; 71 and 76; 70 and 83 (see Figure 4.13).

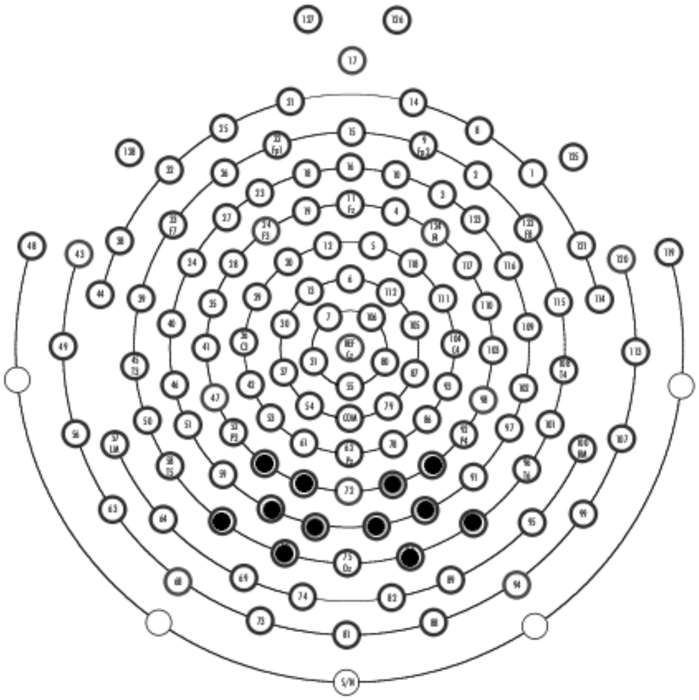

Figure 4.13. 128-channel geodesic sensor net with electrode montage for analysis shaded.

4.4.2. Results

Behavioural data. Median reaction times for correct responses on the magnitude comparison task were analysed with a mixed factorial ANOVA with Ratio (small or large) as a within subject factor, and Condition as a between subjects factor. Results revealed a significant main effect of Ratio, $F(1,27) = 9.43$, $p = .005$, $\eta^2_p = .26$, driven by faster response times for small ratio trials ($M = 593.98$, $SE = 15.34$) compared to large ratio trials ($M = 644.93$, $SE = 24.76$). Neither the main effect of Condition nor the interaction reached significance, $Fs < 1$ (see Figure 4.14).

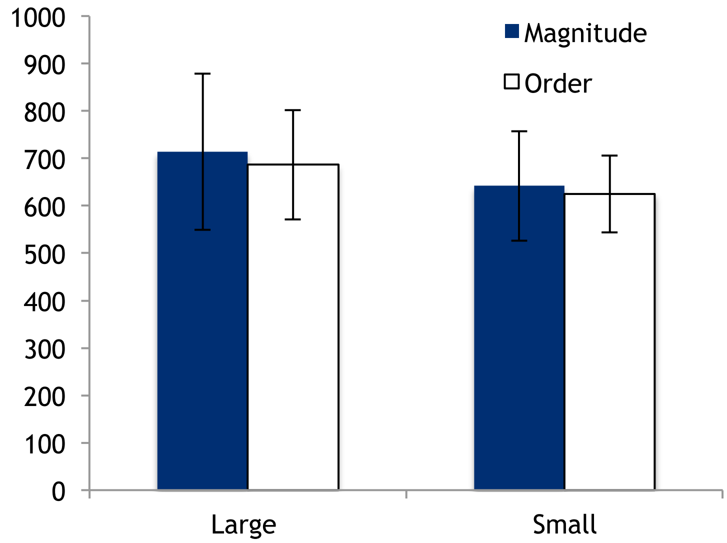

Figure 4.14. Reaction times separated by ratio and condition. Error bars show standard deviations.

Accuracy data were analysed with a mixed factorial ANOVA with Ratio (small or large) as a within subjects factor, and Condition as a between subjects factor. Kolmogorov-Smirnov tests showed that accuracy data were not normally distributed, suggesting caution in interpreting parametric statistics alone. We therefore accompanied these with the appropriate non-parametric equivalent. Results revealed a significant main effect of Ratio, $F(1,27) = 13.69$, $p = .001$, $\eta^2_p = .34$, and a significant interaction between Ratio and Condition, $F(1,27) = 4.38$, $p = .046$, $\eta^2_p = .14$. An analysis of simple main effects revealed a simple main effect of Ratio in the magnitude condition, $F(1,27) = 16.21$, $p < .001$, $\eta^2_p = .38$, but not in the order condition, $p = .26$. Participants in the magnitude condition were significantly more accurate at small ratio comparisons than large ratio comparisons, whereas participants on the order condition performed equally well on both comparison types (see Figure 4.15).

Furthermore, the simple main effect of condition on large ratio trials showed a trend towards significance, $F(1,27) = 3.57$, $p = .07$, $\eta^2_p = .12$. Nonparametric Wilcoxon Signed Ranks tests provided converging evidence for the interaction, showing a significant effect of ratio on accuracy in the magnitude condition, $Z = -3.08$, $p = .002$, but not the order condition, $Z = -.76$, $p = .45$. Additionally, Mann Whitney tests revealed there was no significant difference between groups on accuracy on small ratio comparisons, $Z = -.82$, $p = .42$, but the group difference showed a trend towards significance for large ratio comparisons, $Z = -1.96$, $p = .051$.

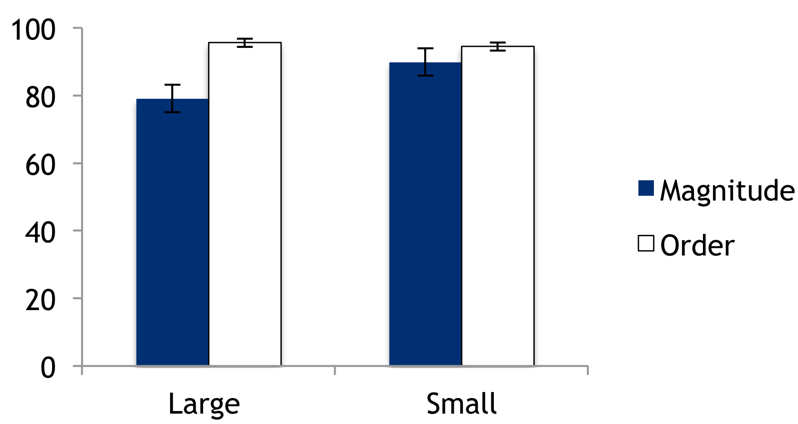

Figure 4.15. Mean accuracy by ratio and condition. Error bars represent standard error of the mean.

Results from the ordinality test showed that only six participants in the magnitude condition and three in the order condition were not completely accurate at the ordering task. Excluding these participants from the analysis of magnitude comparison accuracy did not change the effects.

ERPs.

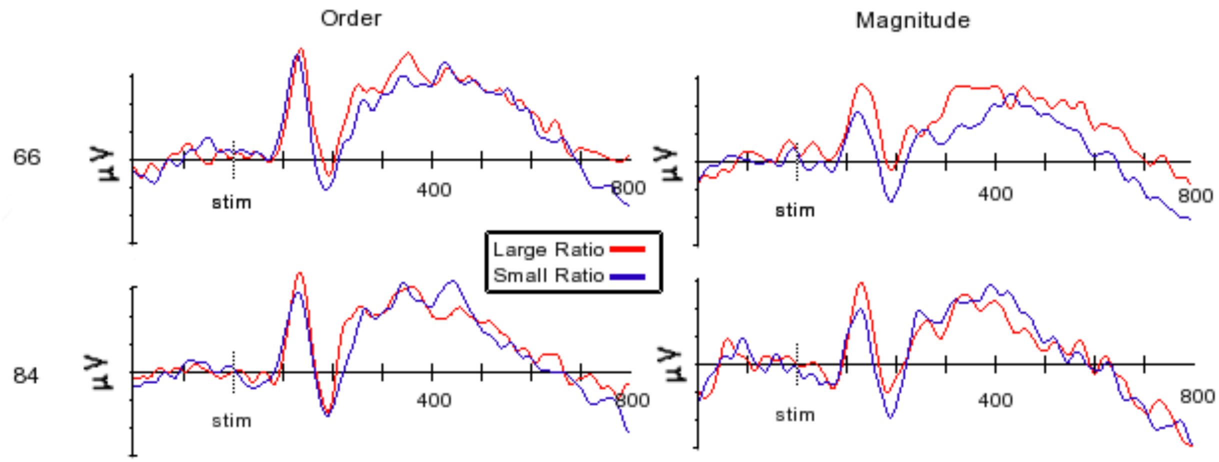

Figure 4.16. Grand average waveforms over representative bilateral electrodes (66 and 84) separated by condition and ratio.

Figure 4.16 represents waveforms for a representative bilateral electrode pair, for the two instruction conditions. A mixed factorial ANOVA was run on mean N1 amplitudes over the twelve electrodes, with Ratio, electrode Pair, and Side of the head as within subjects factors, and Condition as a between subjects factor. The Greenhouse-Geisser correction was used when the assumption of sphericity was violated. Results revealed a significant main effect of Ratio, $F(1,27) = 18.16$, $p < .001$, $\eta^2_p = .4$, driven by a higher mean amplitude for large ratio comparisons ($M = 1.99$, $SE = .3$) than for small ratio comparisons ($M = 1.01$, $SE = .27$). There was also a significant main effect of electrode Pair, $F(2.66, 71.72) = 6.6$ $p = .001$, $\eta^2_p = .2$, but the significant main effect of side did not reach significance, $F<1$, indicating that ratio effects were found bilaterally. Additionally, there was a significant interaction between Ratio and Pair, $F(2.79, 75.43) = 7.3$, $p < .001$, $\eta^2_p = .21$, driven by the fact that the difference across ratio trended toward significance at one pair (60 and 85), $p = .055$, but was significant at the other 5 pairs, $p < .01$. Crucially, the main effect of Condition failed to reach significance, $F < 1$, demonstrating that the ratio effects did not differ across learning conditions. None of the other interactions reached significance.

The same analysis was run on mean amplitudes for the P2p and revealed the same pattern of results. There was again a significant main effect of Ratio, $F(1,27) = 15.05$, $p = .001$, $\eta^2_p = .36$, driven by a higher mean amplitude for large ratio comparisons ($M = .186$, $SE = .44$) than for small ratio comparisons ($M = -.839$, $SE = .36$). There was also a significant main effect of electrode Pair, $F(2.87, 77.61) = 13.47$, $p < .001$, $\eta^2_p = .33$, as well as a significant interaction between Ratio and Pair, $F(2.3, 62.04) = 4.36$, $p = .013$, $\eta^2_p = .14$. Pairwise comparisons confirmed there were significant differences across ratio at each pair, with the interaction driven by larger effects of ratio for some electrode pairs compared to others. The main effect of Condition again failed to reach significance, $F < 1$. No other main effects or interactions reached significance.

4.4.3. Discussion

The aim of this experiment was to investigate whether the way in which a novel set of symbols was taught influenced subsequent magnitude comparison performance on the symbols, as well as associated ERPs. Magnitude comparison performance for these newly acquired symbols was associated with ratio effects on ERP components during the same time windows and over the same scalp locations as had previously been found with real numerical symbols (Dehaene, 1996; Libertus et al., 2007; Temple & Posner, 1998), suggesting that participants processed the learned symbols similarly to how adults process numerical symbols. Crucially, results showed that neither reaction time, nor electrophysiological indices of ratio effects, hypothesised a priori from the existing literature on numeral processing, differed across participants in different learning conditions. This finding does not fit with an account under which ratio effects measured with numerical symbols are *solely* reflective of underlying approximate representations of quantity. Particularly, participants in the order condition were not taught about approximate quantities, yet their later symbolic comparison performance and associated ERPs did not differ from that of participants who were. Furthermore, these results suggest that adults can form symbolic representations based on either numerical magnitude or order information alone.

Based in part on evidence that comparing numerical symbols and comparing non-symbolic magnitudes demonstrate highly similar effects of ratio on behaviour and electrophysiology, it has been proposed that processing numerical symbols involves automatically accessing their corresponding approximate quantity (e.g. Libertus et al., 2007). In line with this idea, one might hypothesize that, in the current study, behavioural and electrophysiological signatures of numerical processing would be influenced by knowledge of approximate quantity associated with the symbols. Our results failed to show this and instead lend further support to the symbolic estrangement hypothesis, which proposed that, in adults, numerical symbols are not tightly linked to non-symbolic representations (Lyons et al., 2012). Specifically, the learned symbols were estranged from quantity for participants in the order condition yet their performance on a symbolic comparison task did not differ in terms of ratio effects on reaction time or ERPs from that of participants who were taught to associate the symbols with approximate quantity. In other words, the notion that ratio effects observed with real numerical symbols reflect automatic activation of underlying representations of approximate quantity is difficult to reconcile with our finding that, in newly acquired symbols, similar ratio effects were observed regardless of whether approximate quantity information played a role in the acquisition of these symbols.

In order to test whether the newly acquired symbols showed ERP signals of numerical processing, electrode locations and time windows of interest for our analysis were chosen a priori based on the timing and location of significant ratio effects in previous studies of magnitude comparison of real numerical symbols (Dehaene, 1996; Libertus et al., 2007; Temple & Posner, 1998). Consequently, while our analysis did not reveal any differences across learning conditions, it does not rule out the possibility that there were differences in electrophysiology at other electrodes or time windows. Previous electrophysiological studies contrasting order and magnitude processing of numbers (Turconi et al., 2004), or contrasting order judgments of letters and numbers (Szűcs & Csépe, 2004) revealed convergent effects on some ERPs, but differing effects on others. However, performance was compared across different task demands (i.e., magnitude vs. order judgements, Turconi et al., 2004), or across stimuli with different perceptual characteristics (i.e., numerals vs. letters, Szűcs & Csépe, 2004), which complicates the interpretation of these differences. Similarly, a previous artificial learning study investigated novel symbol acquisition when given information about spatial order compared to when given information about non-symbolic magnitude (Zhao et al., 2012), but again the post-training comparison tasks differed in terms of instructions, confounding the interpretation of results. Importantly, in the current study, the perceptual characteristics of the stimuli in the final magnitude comparison task, as well as the task instructions, were identical across conditions, hence the only thing that differed was the way in which the novel symbols were taught. Therefore, if differences were found, they could have been attributed to the method of instruction and not to confounding factors. As there were no significant group differences, this strongly suggests that participants in both learning conditions formed similar symbolic representations.

The only significant difference observed between the learning conditions was that, for participants in the order condition, accuracy was not modulated by the ratio between the compared symbols. Mean accuracy on large comparison trials for participants in the order condition also showed a trend towards being higher than that of participants in the magnitude condition. In other words, when participants were taught the order of the symbols, in contrast with their magnitude, they were just as accurate at the more difficult large ratio comparisons than at easier small ratio comparisons. This supports the hypothesis that participants who received order information would form more precise representations than participants who received magnitude information. Furthermore, this largely fits with the proposal that symbolic numerical representations are more reliant on the ordinal relations between symbols than their corresponding approximate quantities (Lyons et al., 2012). Alternatively, as performance was overall highly accurate in the order group, it is possible that the absence of ratio effects in the order condition could be attributed to ceiling effects. Participants were taught a small set of five symbols, which is well within the capacity of the average adult short-term memory span of approximately seven items (Miller, 1956). Thus, if participants were asked to learn a set of symbols that exceeded short-term memory capacity, they may struggle to represent the order of symbols as precisely and accuracy may then be influenced by ratio. Nevertheless, reaction times are more sensitive indices of ratio effects than error rates are, as adults tend to be highly accurate at magnitude comparison of real numerical symbols (e.g. Turconi et al., 2004). Therefore, as the effect of ratio on reaction times did not differ across learning conditions, it does seem that participants in both conditions formed similar representations of the novel symbols.

While adults formed similar representations of novel symbols regardless of whether they were given information about order or magnitude, this does not necessarily suggest that young children would do the same. Interestingly, many participants in the current study, and also in Experiments 4 and 5, reported that they did not actually think they had learned to associate the symbols with numerical magnitudes, despite performing above chance on the post-training assessments. Furthermore, the adults in the current study already had an established symbolic system for numbers and therefore strategies for learning new symbols are likely different from those of children with limited experience with numerical symbols.

4.5. Experiment 7. Can children form novel symbolic representations?

In Experiment 7, the paradigm used in Experiment 6 was adapted to be appropriate for 6-year-old children in order to test whether they are able to form novel symbolic representations with abstract symbols. Young children who had only recently learned real numerical symbols would likely approach the artificial learning tasks differently than adults who had well-established symbolic numerical representations. We therefore hypothesized that congruency would influence children's novel symbolic representations even more strongly than adults' as children are more influenced by congruency on non-symbolic comparison tasks than adults are (Szűcs et al., 2013). Specifically, we expected that children taught to associate symbols with congruent non-symbolic arrays would perform better on the post-training comparison task than children taught to associate symbols with incongruent arrays. Furthermore, we hypothesized that children taught the explicit order of abstract symbols would perform better than children who were only exposed to the non-symbolic arrays.

4.5.1. Method

Participants. Thirty 5-6-year-old children (M = 5;8 SD = 3.5 months) in Year 1 were recruited from local schools and through Oxford Science Adventures, an event that ran over the Oxfordshire half term in which families came in to the department to participate in science-themed crafts and activities as well as developmental psychology experiments. Sixteen were female. Three children failed to complete the tasks.

Materials. Experimental tasks were presented on a laptop computer using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Responses were made with keyboard button presses.

Stimuli. As congruency was a between subjects variable in this experiment, two sets of images were created. For the incongruent arrays, area was held constant at 10,844 pixels so that the average individual coin size decreased as numerosity increased (see Figure 4.16). For the congruent arrays, the average individual coin size was held constant at 187 pixels so that total surface area increased as numerosity increased. Spatial locations of the coins varied randomly and each array was only displayed once in the experiment.

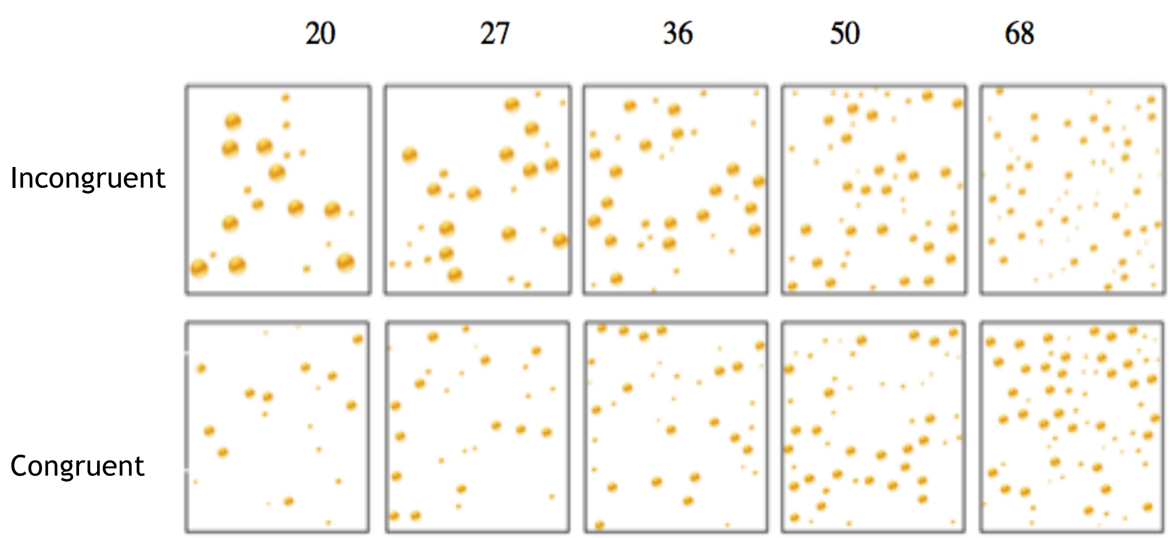

Figure 4.16. Examples of incongruent (top) and congruent (bottom) arrays.

Procedure. The paradigm from Experiment 6 was accompanied by a motivational narrative about helping Pete the Pirate find out how much treasure is in each chest. Participants were randomly assigned to the order condition, a non-symbolic-only condition with incongruent arrays, or a non-symbolic-only condition with congruent arrays. They completed the training session followed by a magnitude comparison task and an ordinality test.

Training. Children were told that Pete the Pirate was going to give them some clues to as to how much treasure was in each chest. They were first exposed to the symbols paired with the associated labels and instructed to try to remember them. In this phase, the symbols were presented in the correct numerical order in the order condition and children were told this was like the count sequence. All groups then saw 100 randomly presented trials, the timing of which is shown in Figure 4.17. In the order condition, the trials were presented in pseudo-random order, as they were always presented in the same order as the "count" sequence. Symbols were paired with incongruent arrays in both the order and incongruent conditions. In the congruent condition, symbols were paired with congruent arrays. Twenty check trials were displayed pseudo-randomly throughout the passive viewing trials in an effort to keep the children engaged. On these trials, participants in the order condition were shown 5 boxes in a horizontal line with a symbol in one box and asked to indicate whether the symbol was in the correct ordinal position. Participants in the other two conditions were shown a symbol and a non-symbolic array and asked to indicate whether the symbol and array were paired correctly. In all conditions response window was open until a response was received and feedback was given. The symbols and labels were paired throughout training in both conditions.

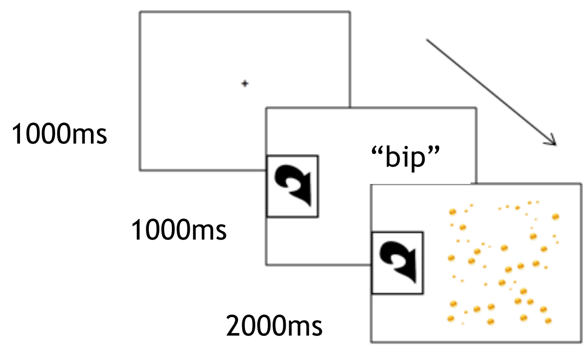

Figure 4.17. Timing of the training trials.

Magnitude comparison task. The task comprised of two blocks of 40 trials with a break in between and children were given unlimited time to respond. The ratio between symbols was considered small if it was approximately .5, and large if it was approximately .75.

Ordinality test. Participants were given five cards, and on each one was an image of a treasure chest and one of the symbols. They were instructed to put the treasure chests in order from smallest to largest, or from least to most treasure.

4.5.2. Results

Magnitude comparison. Two participants were excluded because their mean accuracy was more than 2 standard deviations below the group mean (see Figure 4.18). A mixed factorial ANOVA was run on accuracy data with Ratio as a within subjects factor and Condition as a between subjects factor. Results revealed a main effect of Ratio, $F(1,22) = 8.18$, $p = .009$, $\eta^2_p = .27$, which was driven by higher accuracy on small ratio ($M = 63.8\%$, $SE = 2.9\%$) compared to large ratio trials ($M = 56.8\%$, $SD = 3.2\%$) (see Figure 4.19). There was also a main effect of Condition $F(2,22)$

= 6.75, *p* = .005, η^2_p = .38. Pairwise comparisons showed that average performance in the order condition (*M* = 74.7%, *SE* = 4.7%) was significantly higher than in both the incongruent condition (*M* = 50.4%, *SE* = 5.3%), *p* = .003, and the congruent condition (*M* = 55.8%, *SE* = 4.7%), *p* = .01. There was no significant difference between the two non-symbolic-only conditions.

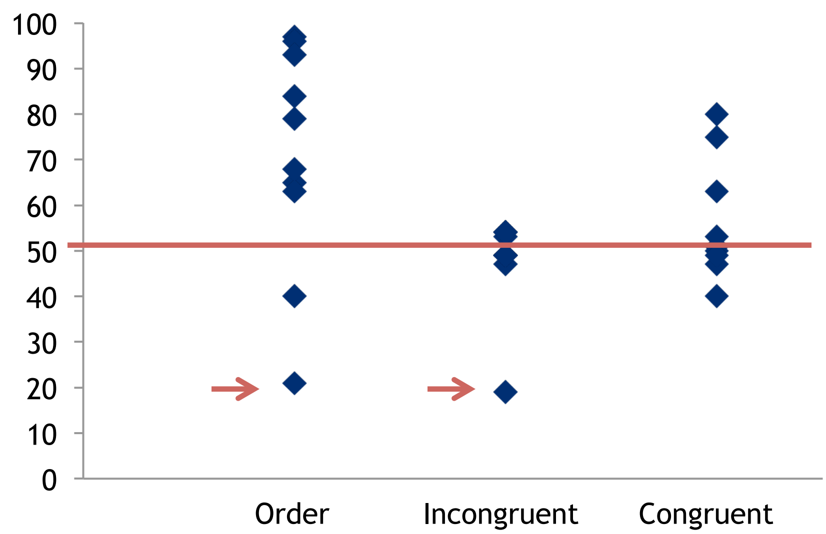

Figure 4.18. Individual mean accuracy separated by group. The red line indicates chance level performance and red arrows indicate excluded data points.

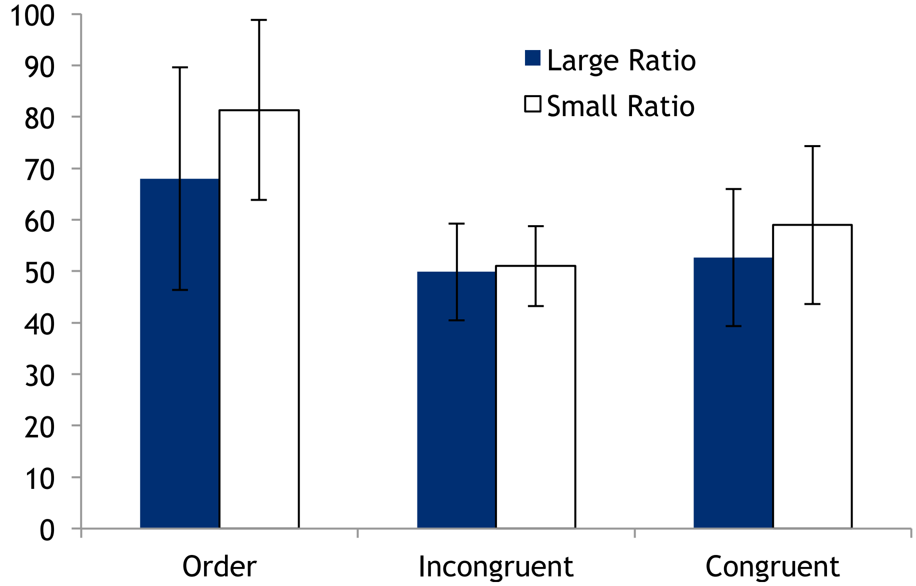

Figure 4.19. Magnitude comparison accuracy plotted by ratio and group. Error bars show standard deviations.

Ordinality. Accuracy on this task was rather low, with the median and mode being correctly placing one of the five symbols. A one-way ANOVA was run on mean accuracy with Condition as a between subjects factor and the main effect was significant $F(2,22) = 6.5$, $p = .006$, $\eta^2_p = .37$. Bonferroni corrected post-hoc tests showed that mean accuracy in the order condition (M = 55.6%, SD = 29.6%) was significantly higher than mean accuracy in the congruent condition (M = 15.6%, SD = 21.9%), $p = .006$, and tended to be higher than mean accuracy in the incongruent condition (M = 25.7%, SD = 19%), but did not reach significance, $p = .07$ (see Figure 4.20). There was no significant difference between the two non-symbolic-only conditions.

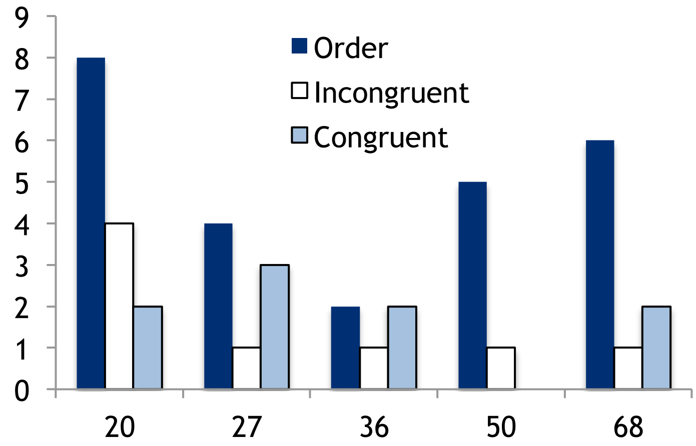

Figure 4.20. Number of participants who placed each symbol in the correct order, separated by condition.

Correlations. Accuracy on the training check trials was significantly positively correlated with both magnitude comparison accuracy, $r(25) = .76$, $p < .001$, and ordinality accuracy, $r(25) = .65$, $p < .001$.

Age effects. In addition to the anchor comparison task described in Experiment 6 above, adults who participated in that experiment completed a two alternative forced-choice version of the symbolic comparison task. This task therefore included trials of the same kind that the children in Experiment 7 completed. In order to investigate differences between children and adults, a mixed factorial ANOVA was run on symbolic comparison accuracy on this task by pooling data from Experiments 6 and 7, with Ratio (small vs large) was a within subjects factor and Age Group (child vs. adult) and Condition (order vs. no order) as between subjects factors. Box's M was significant ($p < .001$) and therefore parametric statistics should be interpreted with caution. Results revealed significant main effects of Ratio, $F(1,50) = 102.52$, $p < .001$,

η^2_p = .67, Age, $F(1,50) = 43.4$, $p < .001$, η^2_p = .47, and Condition, $F(1,50) = 12.29$, p = .007, η^2_p = .2. There was also a significant three-way interaction, $F(1,50) = 6.34$, p = .015, η^2_p = .11 (see Figure 4.21). An analysis of simple main effects revealed that adults' accuracy was significantly higher than children's on trials in both large, $F(1,50) = 22.29$, $p < .001$, η^2_p = .31, and small, $F(1,50) = 94.94$, $p < .001$, η^2_p = .66, ratio trials in the magnitude learning condition and on small ratio trials in the order learning condition, $F(1,50) = 5.69$, $p = .021$, η^2_p = .1, but not on large ratio trials in the order condition, $F(1,50) = 2.06$, $p = .157$.

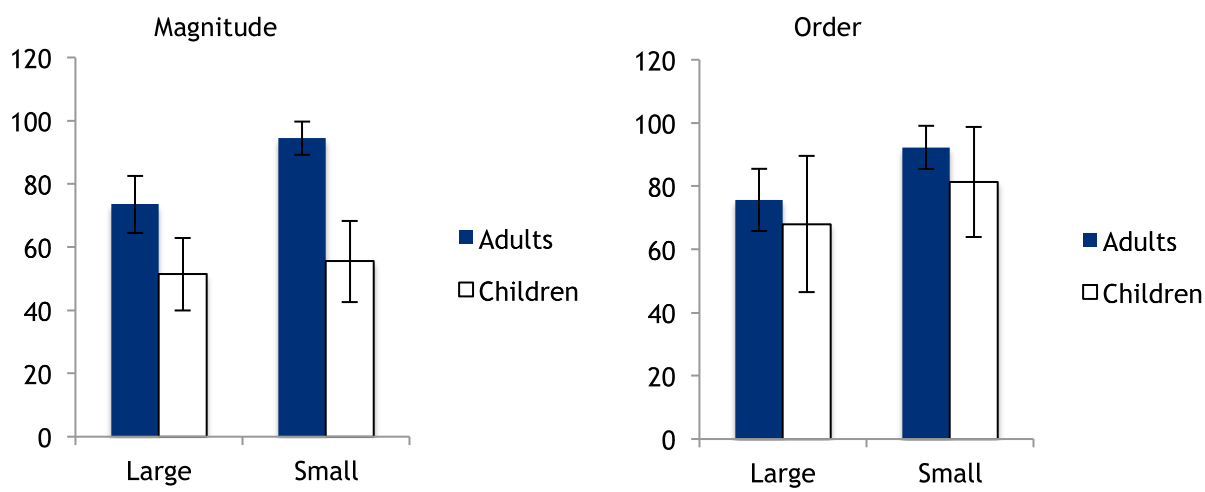

Figure 4.21. Mean symbolic comparison accuracy separated by Ratio, Age group, and Condition

4.5.3. Discussion

The results of the current experiment supported our hypothesis that children's formation of novel symbolic representations would be facilitated by ordinality cues. Crucially, children in the order condition successfully performed a symbolic comparison task with the learned symbols and showed the expected characteristic ratio effects after completing a brief (< 15 minute) training session. Furthermore, 6-year-old children in the order condition did not perform significantly differently from adults on small ratio trials on the comparison task. However, children in the two training conditions that did not provide the precise order of the symbols failed to perform above chance on the comparison task. This suggests that children, unlike adults, were not able to form numerical representations on the basis of non-symbolic magnitude information alone. Order, therefore, seems to play a particularly important role in learning in young children with more limited experience with numerical symbols. These findings align with Carey's (2009) proposal that the words in the count list act as placeholders onto which children attach numerical meaning. Furthermore, these results fail to support the hypothesis that ANS representations play a prominent role in the acquisition of numerical symbols (e.g. as proposed by Piazza, 2010).

Results did not support the hypothesis that the congruency between discrete and continuous quantity in non-symbolic arrays would influence children's formation of novel symbolic representations. This finding could be attributed to floor effects, as children failed to perform above chance in either the congruent or incongruent magnitude conditions. However, a group of 6-year olds was trained in a condition with order cues and congruent non-symbolic arrays in a subsequent experiment (Bishop, 2015, unpublished undergraduate dissertation), and there were no group differences between children in this group and those that completed training with order and incongruent arrays. This suggests that order is a more powerful cue than non-symbolic magnitude, and congruency, for younger children. Alternatively, in Experiments 5 and 6, congruency was a within subjects variable, whereas here it was a between subjects variable. In particular, this resulted in some comparisons in those previous experiments in which arrays with larger numerosities were associated with less total surface arrays than arrays with smaller numerosities. It is possible that this incongruence was more salient and thus had a stronger influence on the formation of novel representations than comparisons in the current experiment, in which total surface area was held constant across numerosities. We had further hypothesized that, if numerical symbols were mapped onto non-symbolic representations, inhibitory control would be required for this mapping as it is required for processing non-symbolic arrays (e.g. Clayton & Gilmore, 2015). However, as congruency did not influence 6-year-olds' formation of novel symbolic representations, this suggests that the training paradigm did not engage inhibitory control in this age group. Executive and attentional processes may be important for selecting information relevant to learning about numbers, but it remains unclear exactly which cues are most important for facilitating the formation of symbolic representations, and therefore to which information attention should be directed during the learning process.

4.6. General Discussion

4.6.1. Summary of findings

To summarize, results of the experiments using an artificial learning paradigm shed light onto the formation of symbolic representations. Specifically, numerical order emerged as the strongest facilitator of learning in these experiments as adults' performance was higher when they were given ordinal cues and 6-year-old children failed to perform above chance when not given the ordered sequence. In contrast, ratio effects on ERP components associated with adults' symbolic comparison did not differ between participants given magnitude or order information, suggesting that adults formed similar representations regardless of how they were taught about the symbols. This suggests that having more experience with real numerical symbols was associated with better performance on the artificial learning task. Alternatively, having more education more generally may have allowed for adults to use more efficient strategies for forming novel symbolic representations. In terms of the hypothesis that congruency between discrete and continuous quantity influences symbolic representations, we found that, when adults formed representations by mapping them onto non-symbolic arrays, congruency did have a significant effect, but that this effect was not seen in children, who failed to map the symbols onto non-symbolic arrays regardless of congruency.

4.6.2. Limitations

A caveat of the experiments discussed in the current chapter is that artificial symbol learning paradigms lack ecological validity. Therefore, while they are useful experimental tools for manipulating factors hypothesized to play a role in learning, the extent to which findings generalize to learning as it occurs in preschool classrooms remains unclear. Furthermore, the training was very brief and the new representations were not maintained long term. Future research could investigate whether more extensive training with abstract symbols could lead to better-established representations that would be recalled for longer following training. Additionally, the numerosities used in the artificial learning studies were all greater than twenty, and it is possible that this accounts for children's failure to perform above chance at associating symbols with non-symbolic arrays, as they may be less familiar with large quantities. However, we chose to include larger numerosities, because small sets are attended to differently from larger sets, as they can be subitized (Kaufman et al., 1949). A further limitation is that it is possible that learning was facilitated in the conditions where participants were given ordinality cues because the verbal sequence was a mnemonic device. Specifically, it is possible that memorizing a sequence of labels would facilitate the formation of novel symbolic representations even if the sequence were not in the correct numerical order. Therefore, while all participants in experiments 5 through 7 were exposed to the auditory labels during training in order to control for the effect of verbal labels, a further comparison group is needed in order to disentangle the effects of numerical order information from those of memorizing a verbal sequence.

4.6.3. Implications

It has been proposed that children learn numerical symbols by mapping them onto approximate representations of quantity (e.g. Piazza, 2010), but there is as yet surprisingly limited empirical supporting evidence for the functional role of a symbolic-to-non-symbolic mapping mechanism in the acquisition of numerals. Here we showed that adults were able to form novel symbolic representations by associating them with non-symbolic quantities, which is convergent with previous artificial learning studies (Lyons & Ansari, 2009; Lyons & Beilock, 2009; Zhao et al., 2012). However, we also found that adults formed similar representations when taught an ordinal sequence of symbols and non-word labels, analogous to the count sequence, instead of being given information about non-symbolic magnitude. This suggests that, at least in adults who already have an established symbolic system of number representation, different instructional methods lead to similar outcomes. Furthermore, adults who were taught about the abstract symbols without being exposed to non-symbolic quantity showed similar effects of ratio on behaviour and electrophysiology. This finding challenges the assumption that these effects, also seen on comparison tasks with real numerical symbols, are indicative of underlying ANS representations of quantity. Symbols are not tightly linked to corresponding quantity in adults (Lyons et al., 2012), and it remains unclear to what extent they are linked in young children, or whether the ANS plays a role in symbol acquisition.

The fact that young children were not able to learn novel symbols on the basis of non-symbolic magnitude alone suggests that order plays an important role in the acquisition of real numbers. When learning real numerical symbols, children learn the count sequence long before learning the cardinality of number words (Wynn, 1990), whereas within this artificial learning paradigm, participants attached numerical meaning to symbols following a few minutes of training. It is therefore likely that existing understanding of real numerical symbols and counting principles influenced their approach to this task. Children's artificial learning was qualitatively different from adults' in that they only succeeded at attaching numerical meaning to the abstract symbols when given order information. Thus, it appears that greater experience with real numbers is associated with more efficient formation of novel symbolic representations. In particular, it seems that having more experience with number exerts a top-down influence on the ability to extract numerosity from non-symbolic arrays. However, preschoolers approach learning numerical symbols without having any established symbolic representations and it remains unclear how they first learn to attend to the relevant numerical cues. Further longitudinal developmental research in younger children is necessary in order to glean insight into how children integrate order and magnitude information as well as multiple representational formats (words, numerals, non-symbolic quantities) to form coherent symbolic representations of real numbers.

CHAPTER 5: SYMBOLIC ESTRANGEMENT IN YOUNG CHILDREN'S NUMERICAL REPRESENTATIONS

5.1. Introduction

The symbolic estrangement hypothesis put forward by Lyons and colleagues (2012) posited that while symbols are not tightly linked to approximate quantity in adults, this does not rule out the possibility that mappings between symbols and ANS representations form much earlier in development, but symbols become estranged over time. However, results from Experiment 7 revealed that 6-year-old children were not able to form novel symbolic representations based on non-symbolic arrays alone, in contrast with the adults who participated in Experiments 4-6. This suggests that attaching symbols to non-symbolic representations may not be a viable learning strategy for young children, but as the symbols were abstract and non-symbolic arrays were for large quantities (>20) this cannot address whether children map numerical symbols to smaller quantities. Here, we aimed to test mappings between symbolic and non-symbolic representations of real numbers in young (4-6-year-old) children in order to see whether they are in fact tightly linked early in development.

5.1.1. The role of non-symbolic representations in the acquisition of numerical symbols

A proposed mechanism for the role of the ANS in the acquisition of the exact meaning of numerical symbols is that, in order for children to learn a given number symbol, their ANS should be precise enough to discriminate its corresponding quantity from numerically adjacent quantities (Piazza, 2010).

Supporting evidence for this proposal comes from correlational findings that 3-5-year-old children's ANS acuity, measured by a non-symbolic magnitude comparison task of quantities from 1-50, was significantly associated with their cardinality knowledge (Wagner & Johnson, 2011). Similarly, 3-6-year-old's ANS acuity was significantly related to their symbolic number knowledge over and above IQ and short-term memory (Mussolin, Nys, Leybaert, & Content, 2012). However, these studies could not prove a causal role of ANS precision in acquiring symbols, and, furthermore, it has recently been argued that these correlations were driven by young children's failure to understand the instructions of the comparison task (Negen & Sarnecka, 2014). Specifically, children who have not acquired the cardinality principle may be unlikely to interpret the word more as meaning more numerous and may instead make comparisons on the basis of associated continuous quantity, rather than discrete numerosity. Indeed, when a non-symbolic comparison task was modified to ensure that children responded based on the number of items, performance was not significantly correlated with cardinality knowledge (Negen & Sarnecka, 2014). Additionally, children with Williams syndrome, a rare genetic disorder, showed very impaired non-symbolic comparison abilities when compared to typically developing children, but were relatively strong at making verbal estimates of quantity on a numerical estimation task (Libertus, Feigenson, Halberda, & Landau, 2014). This clearly demonstrates that ANS precision is not a prerequisite for mapping between exact, symbolic representations of number and non-symbolic quantity. Further evidence refuting the notion that the ANS is involved in the acquisition of numerical symbols comes from the finding, elaborated in Chapter 1, that children were not able to map number words to approximate non-symbolic quantities before learning the cardinality principle (Le Corre & Carey, 2007).

A recent study replicated and extended the findings of Le Corre and Carey (2007) and showed that the development of mappings between numbers larger than four and the ANS is not bidirectional (Odic, Le Corre, & Halberda, 2015). While children who had not acquired the cardinality principle did not successfully map verbal estimates to sequences of taps, when they were instead told a number word and asked to tap that number of times without counting, their tap sequences did show scalar variability. This suggests that it is easier to map from exact to continuous representations and that words allow for the isolation of regions on a continuum. In a related vein, young children performed above chance on a cross-format comparison task only when comparisons were between numbers they knew the cardinality of (Batchelor, Keeble & Gilmore, 2015). Specifically, children who knew the cardinality of, for example, 3 were able to compare words and non-symbolic arrays for trials within their knower level, whereas children who were cardinality principle knowers performed above chance on the task as a whole. This supports the idea that knowing the meaning of number words allows for better processing of non-symbolic arrays. However, the authors' interpreted this result as suggesting that these mappings play a role in the acquisition process, as they formed before children learned the cardinality principle. A caveat of both of these interpretations is that subset knowers by definition map between symbols and quantities smaller than four. Therefore, as these quantities can be represented exactly through subitizing rather than approximately, it remains unclear how these mappings relate to the ANS.

While most previous research has focused on young children's associations between number words and non-symbolic quantity (e.g. Le Corre & Carey, 2007; Wagner & Johnson, 2011), fewer studies have investigated mappings between Arabic digits and non-symbolic quantities. The cross-sectional development of 3 to 5-year old children's mappings between non-symbolic arrays, number words, and digits was investigated using explicit mapping tasks (Benoit, Lehalle, Molina, Tijus, & Jouen, 2013). Participants either a) saw a target number (either digit or canonical dot array as typically presented on dice) and verbally produced the corresponding number word, b) heard a number word and then had to choose the matching digit or canonical dot array amongst six digits or arrays, or c) saw a target number (1-6) in one format and had to choose the corresponding response in the other visual format. Results showed that 3-year-olds only performed above chance when mapping between small (<4) number words and dot arrays, whereas five-year-olds reached ceiling on all tasks, suggesting that mappings are formed between these ages. Four-year-olds were less accurate at mapping between digits and words than between digits and arrays, and words and arrays, suggesting that they link digits to words only after mapping both to non-symbolic representations. Consistent with the findings of Le Corre & Carey (2007), these data showed that 3-5 year old children are more proficient at mapping between symbolic and non-symbolic representations of numbers that fall within the subitzing range than they are at mapping across formats for larger numbers. However, neither of these studies was able to address whether mapping between larger number symbols and quantities continues to improve once children have learnt the cardinality principle.

5.1.2. Children's imprecise mappings between symbolic and non-symbolic representations of numbers larger than four

A study investigating cross-format mapping ability in 6 to 8-year-old children, who were cardinality principle knowers, showed that they were able to map bi-directionally between Arabic numerals and non-symbolic arrays of large numbers (20-50) on two-alternative forced-choice tasks (Mundy & Gilmore, 2009). Children were shown a target, either a dot array or an Arabic numeral, and were asked to select the quantity that matched the target from two choices in the opposing format. The ratio between the target and distractor was manipulated so that on half of the trials it was small and on the other half it was large. Results revealed greater mapping accuracy rates were associated with higher scores on a standardized test of math achievement. However, when mapping digits to non-symbolic quantities, children did not perform above chance level on large ratio trials. This suggests that, even when school-aged children have acquired an understanding of large numerical symbols, they exhibit imprecise mapping between large quantities and their symbolic referent. Consistent with Mundy and Gilmore (2009), Brankaer, Ghesquière, and De Smedt (2014) found that 6-8-year-old children were also able to map across symbolic and non-symbolic representational formats of smaller numbers (1-9) and both studies showed that older children were more accurate overall at mapping compared to younger children. However, with small numbers (Brankaer et al., 2014), age-related improvement interacted with ratio such that older children were significantly more accurate on small ratio but not on large ratio trials. This again suggests that these children do not have very precise mappings between symbols and their corresponding quantities, as accuracy appears to reach a plateau for large ratio distractors. Brankaer and colleagues (2014) also assessed mathematics using a standardized test and found that mapping performance accounted for variance in math achievement over and above what was accounted for by symbolic and non-symbolic magnitude comparison performance. They therefore argued that the association between symbols and their corresponding non-symbolic quantities is important for the development of math achievement in young children, but that further research is necessary to determine the possible causal mechanism underlying this relationship.

Children’s demonstrated ability to map between different representational formats of number offers tentative support for Lyons and colleagues’ (2012) speculation that while symbolic representations are not tightly linked to ANS representations in adults, the two may be more strongly related in young children. Yet this remains controversial given the discrepancy in mapping performance between numbers within and outside of the subitizing range (Benoit et al., 2013; Le Corre & Carey, 2007) as numbers smaller than four do not seem to be linked to the ANS. Furthermore, 5-year-old children's performance on a non-symbolic comparison task was not related to performance on a symbolic comparison task either concurrently or longitudinally (Sasanguie et al., 2013). This suggests that there are separate systems for processing exact symbolic and approximate non-symbolic numbers and that the two do not overlap, even in very young children. While learning numerical symbols clearly entails learning the cardinality principle, and therefore that symbols refer to a given quantity (e.g. Le Corre & Carey, 2007), there is limited evidence supporting the idea that each symbol is ever tightly mapped onto its associated *approximate* non-symbolic representation. It therefore remains unclear how strong the connections are between symbolic and non-symbolic representations of number in children who know the cardinality principle, and how they may differ between numbers that fall within the subitizing range and larger numbers.

5.1.3. Differing attentional demands of enumerating small versus large non-symbolic quantities

The observed differences between mappings for numbers smaller than four than for numbers larger than four (e.g. Le Corre & Carey, 2007) point to a potential role for domain-general processing. Specifically, as four is the maximum number of items that can be attended to (Trick & Pylyshyn, 1994) or held in memory (e.g. Luck & Vogel, 2013) in parallel, it is likely that this constraint on attention and memory influences how children process non-symbolic arrays and in turn how they map to them. Evidence from neuroimaging studies have implicated attention in number processing in adults (e.g. Ansari et al., 2007; Demeyere et al., 2014; Sathian et al., 1999). For example, an fMRI study comparing small and large magnitude comparison showed that the right temporo-parietal junction (TPJ) had higher levels of activation in the non-symbolic small number processing condition compared to the non-symbolic large number processing condition (Ansari, et al., 2007). A contrast between small and large symbolic number processing did not show the same effect. The right TPJ is involved in stimulus-driven visual attention (Corbetta & Shulman, 2002), which suggests that stimulus-driven attention may be used during processing of non-symbolic arrays of small numbers but suppressed during processing of larger arrays.

By manipulating attentional demands of processing arrays of 1-3 objects, Hyde and Wood (2011) demonstrated that, when adults were not able to attend to the objects individually in parallel, their representations of small sets showed characteristics of the ANS. Specifically, under conditions of high attentional load or reduced viewing angle, changes in the numerosity of the set were associated with ratio effects on the P2p ERP component. However, when attentional capacity was not limited, changes in the numerosity of the set were associated with amplitude increases in the N1 component that varied as a function of the cardinality of the set. N1 is typically thought to reflect visuo-spatial attention, and therefore the authors interpreted this finding as reflective of the distribution of attention to individual objects. This highlights a causal role for visual attention in processing the numerosity of small sets of objects. It is worth noting here that while numerical cognition researchers have recognized the role of attention in enumerating small sets, there are some discrepancies in terminology usage in the literature. For example, some authors use the term parallel individuation to refer to the process of enumerating up to four objects (e.g. Hyde & Wood, 2011; Le Corre & Carey, 2007), whereas others use the term object tracking system (OTS) (e.g. Piazza, 2010). However, the OTS has been defined as a mechanism for representing numerosities up to four and this has been extended to sequential presentation of numerosities (Sella, Berteletti, Lucangeli, & Zorzi, 2015). It seems implausible that items presented sequentially would be attended to in parallel, which suggests that researchers who focus on domain-specific numeracy questions may neglect to fully account for domain-general processes. For the sake of clarity, throughout this chapter I use the term subitizing to refer to this ability of attending to up to four items in parallel.

5.2. Experiment 8. Young children’s mappings between words, digits, and non-symbolic arrays

Most existing research investigating mappings between number symbols and non-symbolic representations has focused on young children who have knowledge of some, but not all numerical symbols (e.g. Benoit et al., 2013; Le Corre & Carey, 2007; Wagner & Johnson, 2011) and few studies have investigated mappings in older children (Brankaer et al., 2014; Mundy & Gilmore, 2009). The goals of the current study were to investigate how tight the links are across symbolic and non-symbolic representations after cardinality knowledge has been acquired and to determine whether this differs for numbers that can subitized. Four to six-year-old children were asked to map across representational formats using explicit two-alternative forced-choice mapping tasks adapted from Mundy and Gilmore (2009). All three formats of numerical representations (verbal number words, Arabic digits and dot arrays) were included, and numbers were chosen from within and outside of the subitizing range to address both exact and approximate non-symbolic representations. We hypothesized that children would have strong cross-format mappings for numbers smaller than 5, consistent with Le Corre and Carey (2007). However, we did not expect to see tight links between larger symbols and approximate non-symbolic quantity. Additionally, we hypothesized that visuo-spatial memory capacity would be associated with mapping ability for mapping between non-symbolic representations of numbers in the subitizing range, but not for numbers outside of this range.

5.2.1. Method

Participants. One hundred and fifteen children ranging between 4 to 6 years of age were recruited to participate from elementary schools in the Thames Valley District School Board, Canada, as well as in Oxfordshire, England. Fourteen were excluded from further analyses because of failure to follow instructions or complete many tasks and one was excluded due to a diagnosis of a learning disorder. Therefore, the final sample consisted of 100 children (*M* = 68 months; *SD* = 8), of which 61 were female.

Materials.

Computerized mapping tasks. The four mapping tasks were created using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) and presented on a laptop computer. They were forced-choice tasks in which children were presented with a target numerosity in one representational format and two choices in a different format and asked to choose the one that matched the target. Four tasks were presented to all participants, in counterbalanced order between participants: non-symbolic-to-digit mapping (ND: in which participants were required to map a non-canonical dot array to one of two Arabic numerals), digit-to-non-symbolic (DN: mapping a single Arabic numeral to one of two dot arrays), verbal number word-to-non-symbolic (VN: in which a child heard a number word and was to be mapped to one of two dot arrays), and verbal number word-to- digit (VD: mapping a number word to one of two Arabic numerals) (see Figure 5.1). Number pairs (i.e., the target and a distractor) were chosen from within and outside of the subitizing range, with equal distance pairs (1, 2, 3) both within and outside of the subitizing range for all conditions. Pairs in the subitizing range included 2,3; 2,4; 1,4 and pairs outside the range included: 6,7; 7,9; & 5,8. Each task presented two blocks with an intervening break. Each block presented each number pair twice, counterbalancing for the side of the screen on which the target numerosity appeared, for a total of 24 trials. Trials were presented in randomized order. A fixation screen was presented for 500ms then the stimuli were displayed for 2500ms in order to prevent children from counting. If a response was not made during stimulus presentation, a blank screen was presented until children responded. A simple narrative about saving dinosaur eggs was included to motivate children. Children were instructed to choose the array or symbol that matched the target dot array, symbol, or word as quickly and accurately as possible. Words were presented in auditory format. To ensure that children did not reliably use total area or perimeter cues to select the correct numerosity in the context of non-symbolic stimuli, dot arrays were created using a Python script that controlled for total surface area in half of the trials and total perimeter in the other half of the trials (Price et al., 2012). More specifically, when total surface area was controlled, the array with more dots occupied a greater perimeter. In the perimeter-controlled dots, the total cumulative surface area was greater in the larger array of dots. This method of creating non-symbolic stimuli was used in order to be comparable to previous studies of mapping between symbols and non-symbolic arrays (e.g. Mundy & Gilmore, 2009; Brankaer et al., 2014).

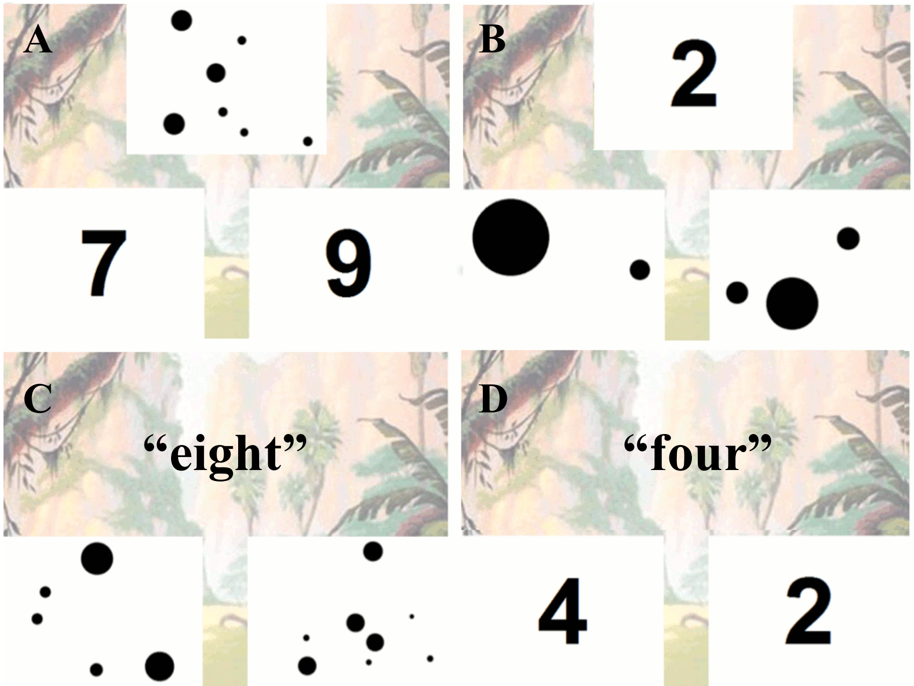

Figure 5.1. A) Non-symbolic-Digit task; B) Digit-Non-symbolic task; C) Verbal-Non-symbolic task; D) Verbal-Digit task

Cardinality knower task. The cardinality task was based on Wynn's Give-a-number task (1992) and children were asked to return x (where x was a number between 1 and 15) escaped animals back to the zoo. Numbers were presented in pseudo-random order. A distinction was made between knowing the Arabic numeral and knowing the number word. Children were first shown a card representing an Arabic numeral and asked to return that many animals to the zoo. If they responded correctly, they were deemed "digit knowers" for that numeral. If they did not respond correctly, they were asked to name the number on the card. If they named the number word incorrectly they were told which number it was and asked to return that many animals to the zoo. If they responded correctly, they were deemed "number word knowers". If an error was made when given the number word, the child was asked to return x-1 animals to the zoo and testing proceeded along the sequential order of numbers until knower level was established. The highest number for which a child answered at least two out of three trials correctly when shown a card was considered the digit knower level and the highest number for which a child answered at least two out of three trials correctly when told the number word was considered the number word knower level.

Counting. A task on which children were asked to complete count sequences of various numbers up to 102 was used to assess counting abilities. The experimenter said, "Now I am going to being counting, can you continue counting after I stop? Let's try a practice, Are you ready? 1, 2, 3… what numbers come next?" Children continued the sequence and the experiment stopped them when they reached the end point for each item. There were eleven counting items spanning all of the decades up to one hundred. A correct count sequence was given one point and the total score on the task was the number of items correct out of eleven. Testing was discontinued after three consecutive scores of 0.

Ordinality. Children were given a twelve-item ordinality task on which they were instructed to write the number that belongs in the middle. Each item had three boxes: the middle one was blank and on either side were digits from 1-20 in either ascending or descending order (i.e. 9 _ 11). If a child could not write numbers, the experimenter wrote for them. The dependent measure was the number of items correct.

Measures of executive and attentional processes. A subset of thirty-eight participants completed measures of EFs and sustained attention. An additional eleven children also participated as part of Oxford Science Adventures. Nine children were excluded for missing data from failure to complete tasks, and one was excluded because of a diagnosis of a learning disorder. The final sample consisted of thirty-nine children (M = 5 years, 9 months, SD = 7 months), twenty-one of which were female. The measures they completed were: The Attention Sustained subtest of the Leiter International Performance Scale – Revised (Leiter-R; Roid & Miller, 1997), Dot Matrix, Digit Recall, Backwards Recall, and Odd One Out subtests of the Automated Working Memory Assessment (AWMA; Alloway, 2007) (all described in Chapter 3), and Animal Size Stroop task described in chapter 2.

Procedure. Participants completed all of the tasks in a pseudo-randomized order over the course of one or two 25-40 minute sessions in a quiet room in their school. They were allowed to choose a sticker at the end of each session. A subset of participants (N = 49) did not complete the counting or ordinality tasks. Specifically, testing in the UK focused on domain-general processes associated with mapping, whereas testing in Canada focused on domain-specific skills.

5.2.3 Results

Cardinality. The number of participants who reached each cardinality knower level is reported in Table 5.1, split by age. More than half of participants reached ceiling on the task. Number word knower and symbols knower were strongly positively correlated, $r(98) = .907$, $p < .001$.

Table 5.1. Cardinality knower levels

	4-yr-olds (N =22)	5-yr-olds (N = 35)	6-yr-olds (N = 41)
Knower Levels	Word (Digit)	Word (Digit)	Word (Digit)
<4	2 (1)	0 (0)	0 (0)
5-8	4 (5)	0 (3)	0 (0)
9-14	9 (10)	9 (10)	4 (4)
15 (ceiling)	7 (5)	25 (21)	37 (37)

Mapping. Nine children who achieved digit knower levels lower than 9 were excluded from the mapping analysis, one child was excluded for failing to perform above chance on the mapping tasks and one was excluded for failing to complete all mapping tasks. To test hypotheses about the role of the ANS in mapping, accuracy was analyzed separately for numbers within the subitizing range (numbers 1-4) and outside the range (numbers 5-9). The justification for this cut off is also demonstrated by the substantial differences in performance between mapping numbers smaller than four and numbers larger than five shown in Figure 5.2. Additionally, preliminary analyses showed no main effect of Country (*F* = .036, *p* > .80), as well as no interaction between Country and Task (*F* = 1.75, *p* >.10); therefore, further analyses were conducted with participants from Canada and UK combined into one group.

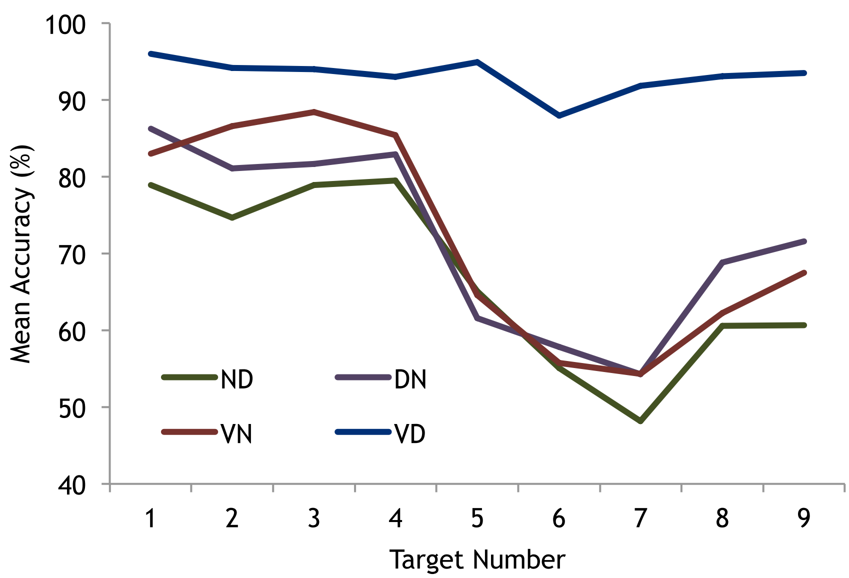

Figure 5.2. Mapping accuracy separated by condition and target number

A 4 (Condition) x 2 (Subitizing Range) x 3 (Age) Mixed Factorial ANOVA was conducted to investigate the development of mapping skills between different numerical representations and range on accuracy rates in 4-6-year-old children. The assumption of sphericity was violated for the Condition x Subitizing Range interaction, and so a Greenhouse-Geisser correction was applied. Results revealed a significant main effect of Task, $F(3, 258) = 58.95$, $p < .001$, $\eta^2_p = .41$. Additionally a main effect of Range, $F(1,86) = 218.95$, $p <$

.001, η^2_p = .72, and a main effect of Age, $F(2, 86) = 3.76$, $p = .027$, $\eta^2_p = .08$, were found. Overall ,six-year old children were significantly more accurate than both four-year-old, $t(55) = -2.14$, $p = .04$ and five-year-old participants $t(71) = -2.26$, $p = .03$. However, there were no significant differences between four and five-year old children $t(46) = -.47$, $p = .64$. The lack of interaction between age and subitizing range suggests that age-related improvements were associated with the entire range of numbers. Furthermore, the difference between accuracy on trials within and outside the subitizing range was not significantly correlated with age, $r(89) = -.059$, $p = .59$.

A significant two way interaction was found between condition and subitizing range $F(3, 225.16) = 25.82$, $p < .001$, $\eta^2_p = .23$. An analysis of simple main effects interrogating the interaction revealed a simple main effect of Range on the non-symbolic to digit mapping condition, $F(1,86) = 95.98$, $p < .001$, $\eta^2_p = .53$, driven by significantly higher accuracy on trials in the subitizing range (*M* = 82.3%, *SE* = 1.8%) compared to trials outside the subitizing range (*M* = 58.5%, *SE* = 2.1%), and the same effect was seen for the digit-to-non-symbolic and number word-to-non-symbolic conditions, but was not significant for the number word-to-digit mapping task.

There was also a simple main effect of Condition on trials within the subitizing range, $F(1,84) = 15.57$, $p < .001$, $\eta^2_p = .36$, driven by higher accuracy on the number word-to-non-symbolic and number word-to-digit conditions (*M* = 90.8%, *SE* = 1.5%; M = 93.7%, *SE* = 1.3% respectively) compared to on the non-symbolic-to-digit and digit-to-non-symbolic conditions (*M* = 82.3%, *SE* = 1.8%; *M*

= 82.8%, *SE* = 2.0% respectively). For trials outside the subitizing range, there was a simple main effect of Condition, $F(1,84) = 97.55$, $p < .001$, $\eta^2_p = .78$, driven by higher accuracy on the number word -to-digit condition than on all of the other conditions (see Figure 5.3). None of the other interactions reached statistical significance (task x age $F = 1.41$, $p = .21$; subitizing range x age, $F = 2.13$, $p = .12$; task x subitizing range x Age, $F = 1.31$, $p = .25$).

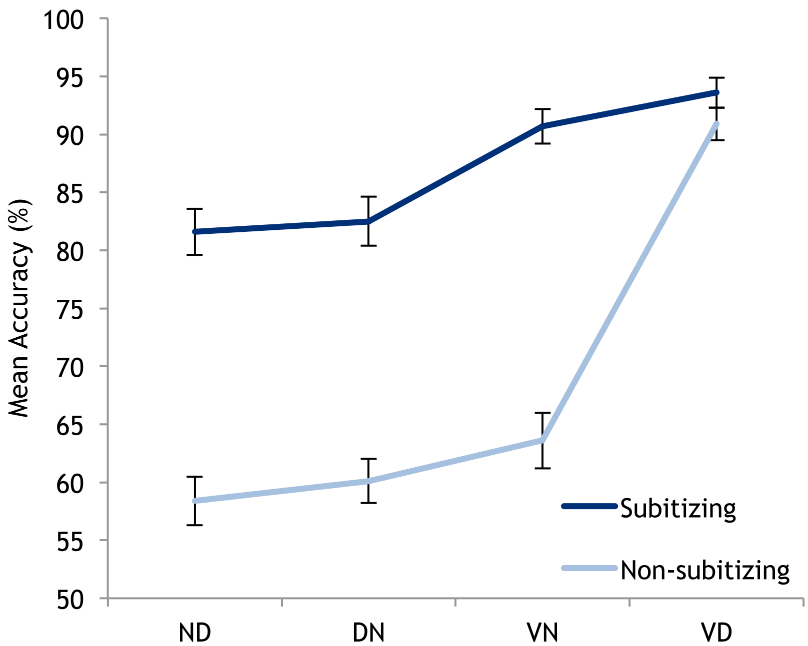

Figure 5.3. Mapping accuracy separated by subitizing range and condition. Error bars show standard errors.

Reaction time data is shown in Figure 5.4. No analyses were run on these data due to low accuracy in some conditions leading to limited numbers of correct trials for reaction time averaging.

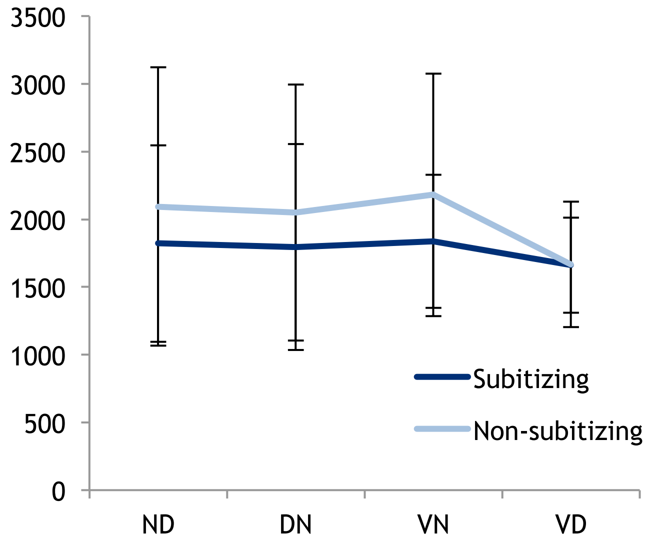

Figure 5.4. Reaction times on separated by subitizing and condition. Error bars show standard deviations

Counting and ordinality. Kolmogorov-Smirnov tests revealed that accuracy was not normally distributed on both the counting task and ordinality task. Kruskal-Wallis tests were run to test the effects of age and showed significant main effects of age on both the counting, $X^2(2) = 21.79$, $p < .001$ and the ordinality task, $X^2(2) = 11.62$, $p < .01$. Mann-Whitney tests were then run in order to locate the age differences. On the counting task, six-year-old were significantly more accurate than both five-year-old children, $Z = -3.01$ $p < .01$, and four-year-old children, $Z = -4.46$, $p < .001$ (see Figure 5.6). Five-year-olds were also significantly more accurate than four-year-olds, $Z = -2.4$, $p < .05$. On the ordinality task, both six-year-old children and five-year-old children were significantly more accurate than four-year-old children, $Z = -3.2$, $p < .01$ and Z

= -2.24, $p < .05$, but there was no significant difference in accuracy between five-year-olds and six-year-olds.

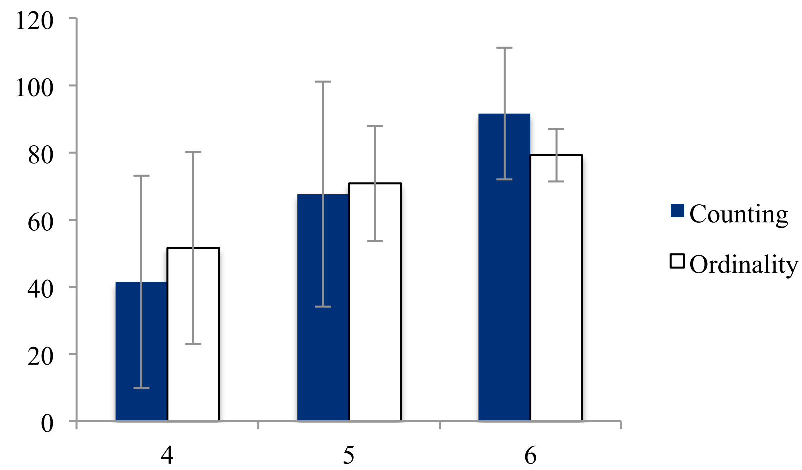

Figure 5.5. Accuracy on the counting and ordinality tasks. Error bars show standard deviations.

Correlations between mapping measures and domain-specific skills. Spearman correlations between the different measures are reported in Table 5.2. All measures were correlated with age and Pearson's partial correlations with age in months partialled out are also reported, although caution should be employed in interpreting them, given violations of parametric assumptions. Partial correlations are reported in text below, as Spearman and partial correlations typically provided converging results. Based on the results of the ANOVA on computerized mapping accuracy, accuracy on these tasks was separated into three conditions: accuracy on the verbal to digit task, accuracy on the other three tasks on trials within the subitizing range, and accuracy on the other three tasks on trials outside of the subitizing range. When age was partialled out, accuracy on trials within the subitizing range was significantly positively correlated with accuracy on trials outside of the subitizing range, $r(94) = .28$, $p = .005$, but neither of these was significantly correlated with accuracy on the verbal to digit task. In other words, accuracy on tasks including non-symbolic arrays was not significantly correlated with accuracy on the task that included only symbolic representational formats of number. Counting was significantly positively correlated with cardinality, $r(59) = .4$, $p = .001$, and ordinality, $r(58) = .42$, $p = .001$. However, somewhat surprisingly, mapping accuracy was not correlated with any other domain-specific numeracy abilities.

Table 5.2. Spearman's bivariate correlations (above diagonal) and Pearson's partial correlations controlling for age (months) (below diagonal)

Measure	2.	3.	4.	5.	6.	7.
1. Age	.47**	.52**	.42**	.1	.19	.22*
2. Cardinality		**.53****	.27*	-.01	.2	.12
3. Counting	**.4****		**.49****	.05	.18	.14
4. Ordinality	.09	**.42****		.1	-.04	-.01
5. Mapping Accuracy (Verbal-Digit)	-.17	-.08	.02		.2*	.09
6. Mapping Accuracy (Subitize)	.07	.03	-.22	.2		**.36****
7. Mapping Accuracy (Non-subitize)	.03	.07	-.14	.04	**.28****	

*Note: *p < .05; **p < .01*

Correlations between mapping and executive functions. Descriptive statistics are reported, split by age, in Table 5.3. Four and five-year-olds were combined due to the small number of four-year-olds. Kolmogorov-Smirnov tests revealed that mapping accuracy on the verbal-symbol condition, as well as within the subitizing range for the remaining conditions violated the assumption of normality, as did Stroop accuracy, and raw score on the Backwards Digit Recall subtest. Therefore, Spearman's correlations are reported in Table 5.4. The majority of the measures were not significantly correlated with age, therefore partial correlations were not run. Mapping accuracy on the verbal to digit condition was again not significantly correlated with any other mapping conditions, nor was it correlated with any domain-general processes. In contrast, mapping accuracy on the other three conditions for trials within the subitizing range was significantly positively correlated with visuo-spatial working memory, $r_s(34) = .34$, $p = .046$, verbal working memory, $r_s(32) = .42$, $p = .016$, and visuo-spatial short-term memory $r_s(34) = .4$, $p = .019$. Similarly, mapping accuracy on the other three tasks for trials outside of the subitizing range was significantly positively correlated with verbal working memory, $r_s(32) = .51$, $p = .002$ as well as visuo-spatial short-term memory $r_s(34) = .43$, $p = .011$.

Table 5.3. Descriptive statistics separated by age

		4-5-yos		*6-yos*	
Task	Measure	Mean (SD)	N	Mean (SD)	N
Animal	Accuracy	87.06% (15.86%)	16	95.91% (5.33%)	23
Stroop	Median RT	1416 (354.35)	16	1181.93 (383.8)	23
AWMA	Digit Recall	22.38 (4.66)	13	24.78 (3.8)	23
	Dot Matrix	14.3 (3.64)	13	17.32 (2.7)	22
	Odd One Out	11.23 (4)	13	14.18 (3.45)	22
	Backwards Recall	6.25 (2.67)	12	8.57 (2.71)	21
Visual Search	Adjusted Raw Score	45.18 (7.5)	16	51.47 (7.49)	23

Note: AWMA = Automated Working Memory Assessment

Table 5.4. Spearman correlations

Measure	2.	3.	4.	5.	6.	7.	8.	9.	10
1. Age	.42**	.22	.31	.32	.42*	.22	-.23	.01	.29
2. Stroop		.48**	.45**	.09	.38*	.23	.18	.16	.15
3. Digit Recall			.55**	.54**	.64**	.3	.22	.24	.26
4. Dot Matrix				.63**	.62**	.49**	.29	**.4***	**.43***
5. Odd One Out					.73**	.26	.18	**.34***	.28
6. Backwards						.15	.23	**.42***	**.51****
7. Visual Search							.1	.13	.17
8. Mapping (VD)								.29	.23
9. Mapping (Sub)									**.36***
10. Mapping (Non)									

Note: $^{*}p < .05$; $^{**}p < .01$; VD = Verbal – Digit, Sub = Subitize, Non = Non-subitize

5.3. Discussion

5.3.1. Summary of findings

The current study contributes to a growing body of literature demonstrating weak and inconsistent support for the commonly held assumption that numerical symbols are tightly linked to corresponding representations of approximate quantity. The present study was the first, to our knowledge, to investigate the strength of the associations between symbolic (i.e. Arabic numerals & verbal number words) and non-symbolic representations (i.e. arrays of dots) of number in young children who had recently acquired the semantics of symbols. Critically, the present data inform our understanding of children's ability to map between symbolic and non-symbolic representations. Specifically, we found that children were more accurate at mapping between dot arrays and both symbolic representations in the subitizing range compared to numbers that fell outside of the subitizing range (similar to Benoit et al., 2013; Le Corre & Carey, 2007). These findings suggest that symbols are directly linked to exact non-symbolic representations supported by subitizing, rather than approximate non-symbolic representations of number. Furthermore, mapping between symbolic and non-symbolic formats was not correlated with mapping between two symbolic formats, suggesting that there are different processing demands of mapping when non-symbolic arrays are involved. Indeed, mapping between symbolic and non-symbolic formats was associated with visuo-spatial short-term memory and verbal working memory, whereas mapping between two symbolic formats was not associated with any domain-general processes. Contrary to our hypothesis, mapping to non-symbolic representations within the subitizing range was not differentially associated with working and short-term memory than mapping outside the subitizing range. Additionally, mapping between representational formats of number was unexpectedly not significantly correlated with other domain-specific processes, counting, cardinality, and ordinality.

5.3.2. Limitations

A limitation of this study is that investigating correlations between mapping performance and domain-general processes does not shed light on causal mechanisms. Future work could experimentally tease apart domain-general contributions to exact and approximate representations of small sets in young children by manipulating attentional demands using methods similar to Hyde and Wood (2011). Furthermore, this study aimed to investigate mappings across different representational formats in children who knew the symbols, and therefore while results suggested there was estrangement in their representations, it remains unclear whether mapping between formats plays a role in the acquisition of symbols. Young children are slow to acquire the meaning of the first four number words (Le Corre & Carey, 2007; Wynn, 1992), but then seem to acquire the meaning of symbols beyond that much more quickly, which makes it difficult to investigate this learning process experimentally. Longitudinal studies investigating the underlying mechanisms involved in the learning process are required to shed some light on such claims. Future research should investigate learning trajectories by assessing children's number knowledge at multiple time points starting from the age of three until the end of the first year of formal education.

5.3.3. Implications

For now, I shall focus instead on the implications of the current findings for the estrangement hypothesis (Lyons et al., 2012), and the functional role proposed for ANS in the acquisition of symbols (Piazza, 2010). Highly relevant to this debate, children assessed here were significantly more accurate at mapping between different symbolic formats compared to mapping symbols to approximate non-symbolic representations of quantity. Based on these results, we propose that symbols are already estranged from approximate quantity in young children, supporting the notion that there are two distinct systems for processing exact symbolic and approximate non-symbolic quantity (e.g. Noël & Rousselle, 2011; Sasanguie et al., 2013).

Previous studies investigating mapping skills in older children (Mundy & Gilmore, 2009) and the association between symbolic and non-symbolic representational systems in adults (Lyons et al., 2012) have shown that the relationship between Arabic numerals and our intuitive approximate number sense may not be as strong as presently believed. The 'symbolic estrangement' hypothesis put forward by Lyons and colleagues (2012) proposed that symbols are initially mapped onto approximate numerosities when children acquire symbolic knowledge, but that these representations become estranged from each other with age and experience. Here we found that even children who had very recently acquired symbolic representations did not bind these larger representations to the approximate numerosities, which suggests that learning the semantic meanings of number symbols may not necessarily engage the ANS. Our data suggest instead that numerical symbols and the ANS may already be estranged once children have acquired a symbolic number system, in other words, it is plausible that they were not mapped onto each other in the first place. The finding that there seem to be distinct systems for processing symbolic and non-symbolic representations of number in 5-year-old children (Sasanguie et al., 2013) further supports this idea. Furthermore, mapping accuracy in the condition in which children mapped between two symbolic representations of number was not correlated with accuracy in the conditions that included a non-symbolic representation, which suggests that symbolic and non-symbolic representations of number are treated differently. Moreover, the clear differences in accuracy between the mappings within and outside the subitizing range observed in our study as well as in previous studies (Benoit et al., 2013; Le Corre & Carey, 2007) directly contradict arguments by Piazza (2010), who posited that the OTS, (subitizing) could not be the foundation for acquiring numerical symbols due to absence of evidence showing abrupt differences in processing numbers within and outside of the subitizing range.

Brankaer and colleagues (2014) used a mapping paradigm nearly identical to the one used in the current study with slightly older (6- to 8-year-old) children, but did not make a distinction between small and large numbers and therefore reached different conclusions based on similar data to ours. They set out to test whether mapping ability was related to math achievement rather than to investigate the role of core non-symbolic systems in mapping and took for granted that the mappings were based on ANS representations. I hypothesize that, if they had compared mapping for numbers within the subitizing range to mapping for numbers outside of it, they likely would have seen similar results to ours.

The results of the present study also demonstrated that children in all age groups performed significantly more accurately on the number word to digit mapping task across the whole range of numbers in comparison to tasks where children mapped between non-symbolic formats. This contradicts Benoit and colleagues' (2013) finding that children learn to map between words and digits only after learning to map between each symbolic format and non-symbolic arrays as even the youngest children in the current study were most accurate at this task. Participating children had all started formal education and received explicit instruction on the names of digits and so it is not surprising that they are were able to map accurately between words and digits. Moreover, children first learn the verbal count sequence and exact verbal representations of numbers within their count list starting around the age of three (Wynn, 1990). The discrepancy in findings could be explained by differences in task demands. The response formats were different between the verbal tasks and the other tasks in Benoit and colleagues' (2013) study and likely entailed differences in working memory demands, whereas response formats were equated across tasks in the present study. These differences in task demands may also account for the fact that 5-year-old children reached ceiling on their tasks, but 6-year-olds did not reach ceiling in the current study. The non-symbolic arrays used by Benoit and colleagues were canonical dot patterns and could be construed as symbolic representations as they are easily recognizable and strongly associated with their corresponding quantities. Furthermore, participants were allowed up to thirty seconds to respond and so could easily have counted the dot arrays, rather than rely on approximate numerical representations.

While previous studies have found that the ability to map across representational formats of number was associated with arithmetic abilities (Brankaer et al., 2014; Mundy & Gilmore, 2009), here we found that mapping ability was not significantly associated with other domain-specific processes, counting, cardinality, and ordinality. This suggests that knowing the cardinality of a number may not entail automatically associating a symbol with a corresponding approximate representation of quantity. However, it is particularly surprising that mapping between words and digits was not associated with domain-specific skills, as a similar task was previously found to be highly correlated with a numerical ordering task (Lyons et al., 2014). As accuracy was rather high in the word to digit mapping condition, it is possible that ceiling effects masked individual differences. Mapping between two symbolic formats was also not associated with any domain-general processes. However, mapping performance in conditions that included a non-symbolic representation was associated with visuo-spatial short-term memory as well as verbal working memory. This supports the hypothesis that visuo-spatial short-term memory contributes to the processing of non-symbolic arrays, but surprisingly extended to arrays outside of the subitizing range rather than being exclusive to arrays that could be attended to and held in memory in parallel.

In sum, the results of the current study highlight differences in the way children are able to map between symbolic and non-symbolic representations of number within and outside the subitizing range. Children who had already learned the meanings of numerical symbols seemed to have strong links between all three representational formats for small numbers, but only between the two symbolic formats for larger numbers. These results suggest that once children have semantic representations of symbolic numerosities, their symbolic representations of numbers larger than four are not strongly related to corresponding ANS representations. This is consistent with the finding that non-symbolic and symbolic number processing are not related in 5-year-old children who have recently acquired numerical symbol knowledge (Sasanguie et al., 2013) and suggests that the ANS and symbolic representations of number are already separate by the age of 5. Therefore, an even stronger conclusion is that they may not ever be tightly linked. Furthermore, domain-general processes were associated with mappings involving non-symbolic representations, suggesting that additional processing costs are required when mapping across formats.

CHAPTER 6: GENERAL DISCUSSION

"Almost anything you pay close, direct attention to becomes interesting."
– David Foster Wallace, *The Pale King*

In the above quotation, the term *attention* conveys conscious alertness and awareness, as a theme of Wallace's writing is that we have the freedom to choose what we attend to in our environment and therefore what we attribute meaning to. This suggestion mirrors to an extent the notion from a cognitive science perspective that attention serves as a gatekeeper for filtering information relevant for learning (Posner & Rothbart, 2007), which has shaped my thinking on the role domain-general processes play in learning about mathematics. This quotation coincidentally also serves as motivation for persevering at reading (or writing) a rather long thesis on an ostensibly narrow topic.

The research presented in this thesis collectively highlights the need to consider factors beyond number sense in the development of early numeracy skills. In particular, my findings contribute to a growing body of evidence challenging the prevalent ideas that the ANS is a system for processing discrete number exclusively (as also suggested by, e.g., Cantrell & Smith, 2013), and that numerical symbols automatically activate corresponding ANS representations (as also suggested by, e.g., Lyons et al., 2012). Results from my research revealed that another domain-specific factor, numerical order, influenced the formation of novel symbolic representations in adults (Experiment 5), and that young children did not succeed at attaching numerical meaning to abstract symbols in the absence of order information (Experiment 7). Furthermore, inhibitory control, a domain-general factor, was required for extracting numerosity from non-symbolic arrays (Experiment 2) and visual attention was found to constrain children's mapping between symbolic and non-symbolic representations of number (Experiment 8). Going forward, longitudinal research with multiple time points between the ages of 2 and 4 years is necessary for closely examining dynamic interactions between attention and numeracy and determining the role that domain-general factors play in learning about number.

6.1. Summary of Key Findings

Here I will briefly summarize the main findings from this thesis. The experiments presented in Chapter 2 showed that a measure of inhibitory control was associated with non-symbolic magnitude comparison, as well as a standardized measure of early numeracy skills. Specifically, performance on incongruent trials on the inhibitory control task was correlated with performance on incongruent trials on the magnitude comparison task, lending further support to the argument that these tasks measure inhibitory control in addition to ANS acuity (Gilmore et al., 2013), especially in early childhood. Furthermore, age-related improvements on both the inhibitory control and magnitude comparison tasks were seen. In Chapter 3, domain-general cognitive training was pitted against domain-specific training, but results revealed that neither training group showed significantly greater improvements than those observed in a passive control group. This contributes to the existing body of mixed evidence for transfer effects of training executive functions suggesting that manipulating cognitive processes in typically developing preschoolers requires a more rigorous intervention. The experiments presented in Chapter 4 all used variations of an artificial learning paradigm to test the influences of different factors on the formation of novel symbolic representations. Adults' representations were influenced by the congruency between discrete and continuous quantity of non-symbolic arrays, and also by whether or not they were given the numerical order of the abstract symbols. Furthermore, their symbolic comparison performance was associated with ratio effects on ERPs at the same electrodes and in the same time windows as had been previously observed when adults compared real numerical symbols. Intriguingly, these effects did not differ between participants who had been taught to associate the symbols with non-symbolic arrays and participants who had been taught the numerical order of symbols but given no information about magnitude. Children, on the other hand, performed below chance when not given information about numerical order and hence no effect of congruency of non-symbolic arrays was observed. Finally, in Chapter 5, mappings between different representational formats of number were investigated in 4-6-year-old children who had recently learned numerical symbols. Children were significantly more accurate at mapping between symbolic and non-symbolic representations of numbers one to four, which can be represented precisely, than of numbers larger than four, which are represented approximately. This finding fails to support the notion that processing numerical symbols entails automatically activating an approximate representation of magnitude (Dehaene, 2011) and instead provides evidence for symbolic estrangement (Lyons et al., 2012) in young children's numerical representations.

In sum, the findings from the experiments presented in this thesis poke holes in prevalent numerical cognition theories and highlight the need to incorporate domain-general factors in revisions of these theories. First, non-symbolic magnitude processing is influenced not only by the discrete number of items in a set, but also by the congruency between continuous and discrete quantity, as well as the size of the set. This points to the requirement of executive control for selecting discrete quantity as the relevant stimulus dimension when performing non-symbolic magnitude comparison. Second, symbols do not seem to be as strongly associated with the ANS as was previously believed. Evidence supporting this claim comes from the finding that symbols for numbers larger than four do not seem to be tightly linked to corresponding ANS representations, even in young children who have recently learned the symbols. Furthermore, ratio effects on ERPs seen when adults performed a comparison task with novel symbols were the same whether or not they had been exposed to non-symbolic magnitude, which challenges the assumption that these ratio effects are indicative of underlying representations of approximate quantity. Taken together, these findings highlight the need to look beyond number sense in order to understand how children learn the meaning of numerical symbols.

6.2. Theoretical Implications

Number sense is related to formal mathematics achievement (e.g. Halberda et al., 2008), yet tasks designed to assess number sense also require inhibitory control. It therefore could be that the observed relationship between non-symbolic comparison and mathematics achievement is accounted for by individual differences in inhibitory control rather than precision of ANS representations. Indeed, Fuhs & McNeil (2013) showed that the relationship between non-symbolic comparison and a measure of early numeracy skills in preschoolers from low SES backgrounds was mediated by inhibitory control. Furthermore, as discussed in Chapter 1, there is a large body of evidence linking executive functions and attentional processes to maths correlationally (e.g. Bull & Scerif, 2001; Steele et al., 2012). Additionally, executive functions and attention were found to mediate the relationship between kindergarten number sense and year 1 mathematics outcomes (Hassinger-Das et al., 2014). In a related vein, knowledge of numerical symbols has also recently emerged as a stronger predictor of mathematics achievement in 6-year-old children than non-symbolic comparison (e.g. Göbel et al., 2014), and in younger children the relationship between number sense and mathematics achievement was mediated by cardinality knowledge (Chu et al., 2015). Taken together, it is clear from these studies that the relationship between ANS acuity and symbolic mathematics can be accounted for by other domain-specific factors, such as cardinality and numerical symbol knowledge, as well as by domain-general factors, such as executive functions and attention.

It still remains unclear how approximate non-symbolic representations of number relate to the acquisition of numerical symbols. Specifically, as discussed in Chapter 1, if numerical symbols are mapped onto corresponding non-symbolic representations of number, then visual features associated with continuous quantity likely influence these mappings. Indeed, results of Experiment 4 showed that adults' novel symbolic representations were influenced by the congruency between discrete and continuous quantity. However, 6-year-old children failed to show the same effect as they did not successfully attribute numerical meaning to abstract symbols when not also given information about numerical order, analogous to the count sequence. This suggests that adults, who have more experience with number, are better at extracting numerical information from non-symbolic arrays than children, suggesting that experience with symbolic number refines ANS precision. This idea is corroborated by recent research showing that young children who do not know the cardinality principle fail to perform above chance on magnitude comparison (Negen & Sarnecka, 2014), and further that children who are subset knowers perform above chance only on comparisons that include numbers they know the cardinality of (Batchelor et al., 2015). Therefore, rather than children's symbolic knowledge being built on their approximate number representations, it could instead be that knowledge of cardinality facilitates children's ability to select discrete number as the relevant dimension in non-symbolic arrays. This idea is further supported by Slusser & Sarnecka's (2011) finding that young children were able to sort arrays of objects based on colour or shape, but only cardinality principle knowers sorted based on the number of objects in an array. In contrast, individual differences in 3-6-year-olds' spontaneous focusing on numerosity (SFON) were related, concurrently and longitudinally, to early numeracy abilities (Hannula & Lehtinen, 2005). Hannula & Lehtinen (2005) argued that SFON differs from enumeration skills in that children are not instructed to attend to number in tasks used to measure this construct. They further suggested that SFON scaffolds the development of numeracy skills, which implies that subsequently acquired numeracy skills build on the ability to attend to number in the environment, rather than that increased numerical knowledge facilitates attention to number. It is clear that attention to number and knowledge of cardinality and number symbols interact, which could account for the mixed theories and evidence for the relationship between non-symbolic representations of number and numerical symbol acquisition. Further research is required to disentangle the mechanisms underlying this relationship.

6.3. Potential Applications for Education

While the findings presented in this thesis are unlikely to lead to direct applications to education, I believe they provide a starting point for linking research from developmental psychology to preschool curriculum. In particular, the findings from this thesis demonstrate the importance of multiple factors contributing to early numeracy success and suggest educators should focus on all of them. From discussions with other developmental psychologists as well as policy advisors working in the UK Department for Education, it is clear that the politicians setting the standards for early years education know very little about what cognitive skills should be expected of a typically developing 4-year-old (e.g. Norbury et al., 2015). In a related vein, given the increasing nursery provisions in the UK, the Office for Standards in Education, Children's Services and Skills (Oftsted) has proposed that nursery providers should be held accountable for preschool education outcomes (Ofsted, 2014). This is therefore an optimal time to positively influence early childhood education policy.

Studying learning in classroom environments could also contribute to our understanding of cognitive mechanisms involved in learning, as education is in and of itself an environmental manipulation. An example of the dramatic effect formal education has on children's understanding of numerical symbols comes from one child who participated in Experiment 8. Prior to starting reception, he had very little knowledge of Arabic digits and admitted after completing the mapping tasks that he had been pushing the buttons at random. When he returned a few weeks after starting school, his accuracy on the mapping tasks increased substantially. It is clear that his first few weeks of reception strongly influenced his understanding of numerical symbols. Therefore, studying how teachers foster this knowledge in their students could shed light on mechanisms underpinning the acquisition of numerical symbols. The majority of existing attempts to bridge cognitive science and education have involved scaling research findings up to the more complex classroom environment (Nathan & Alibali, 2010). For example, the finding that ANS acuity is correlated with mathematics ability (Halberda et al., 2008) has led to ANS training interventions designed to improve arithmetic (e.g. Wilson et al., 2006). In contrast, scaling down from the classroom and rigorously investigating existing teaching practices could lead to better understanding of what works in the classroom as well as how it works, and therefore impact educational practice.

6.4. Limitations

The experiments presented in this thesis have methodological limitations that could be improved upon in future work. In particular, given the consensus that continuous visual properties of number influence judgments of non-symbolic magnitude (e.g. Gebuis & Reynvoet, 2012a), the non-symbolic stimuli used in Experiments 1 and 8 should ideally not have equated variables associated with continuous quantity across arrays. Additionally, although my studies focused on a relatively narrow age range (3-6-years), executive functions show rather steep trajectories of improvement across those ages (e.g. Garon et al, 2008) and therefore some of the tasks included in Experiments 1, 2, and 3 were not sensitive to individual differences across the entire range. For example, the Backwards Recall subtest of the AWMA, used in Experiment 3, resulted in floor effects in the youngest children, and the Animal Stroop task showed ceiling effects in the oldest children in Experiment 2. A further limitation is that, due to the task impurity problem, it has been recommended that multiple measures of each construct should be used in order to obtain valid and reliable measures of executive functions (Miyake et al., 2000). Ideally, I should have included more than one measure of each executive construct of interest in my experiments, but opted to keep my testing sessions shorter in order to impinge less on schools and nurseries.

My experiments accomplished my goal of investigating the role of domain-general processes in *learning* about mathematics, to varying degrees. Specifically, the null training results seen in Experiment 3 rendered addressing mechanistic hypotheses impossible. As there is now increased evidence that executive function training, or at least training working memory in particular, does not consistently lead to far transfer effects (e.g. Melby-Lervåg & Hulme, 2012), it is clear that effectively manipulating executive functions with the goal of transferring to educational outcomes will require more than computerized training. For example, training working memory with the goal of transferring to early numeracy was more effective when domain-specific aspects were highlighted in the working memory training (e.g. Kroesbergen et al., 2014). Embedding training in existing curricula could prove to be even more effective as the context in which domain-general executive skills and attentional processes are deployed matters (Ristic & Enns, 2015). In contrast, studying the formation of novel symbolic representations using artificial learning in adults proved much easier than implementing cognitive training in preschoolers. However, the extent to which those findings are applicable to early learning is unclear, as adults' previous experience with real numerical symbols likely had a strong influence on their approach to the novel learning challenge, whereas children approach learning numbers with much more limited experience. Future work investigating learning in early childhood should implement more extensive environmental manipulations, as well as observe change over shorter periods in preschool, as change happens rapidly during this period.

6.5. Future Directions

The research conducted for this thesis has led to the refinement of my research questions, as well as the generation of additional hypotheses I would like to pursue in future work. In particular, mechanisms underlying relationships between learning to attend to number and learning numerical symbols require further study. It has been proposed that the degree to which one attends to discrete number changes over the course of development: Young children first discriminate magnitudes based on continuous properties but come to rely more on discrete numerosity with age and experience (Leibovich & Henik, 2013). However, as discussed in Chapter 1 and section 6.2, the direction of the relationship between learning the cardinality of numerical symbols and attending to discrete number in non-symbolic arrays remains unclear. In my future work, I would like to more closely examine this relationship with longitudinal research in children between the ages of 2 and 4-years of age in order to better understand how children acquire the meaning of numerical symbols and the role played by domain-general processes. As discussed in section 6.4 above, I wish to study learning as it happens in the classroom in order to scale findings down and glean insight into cognitive mechanisms underlying successful learning. This will require close collaboration and engagement with educators and early childhood carers. In a complementary approach, I also plan to develop neuroimaging analysis skills, particularly functional magnetic resonance imaging (fMRI), in my future training in order to investigate the development of brain structure and function associated with learning in early childhood.

Language also plays an important role in the process of learning about numbers. Cardinality is just one property of number and number words are also used to describe ordinal and nominal properties of objects. For this reason, language may be the basis of numerical thinking as number words represent all three properties whereas non-symbolic arrays represent just the one (Wiese, 2003). Bilingual children have certain cognitive advantages compared to their monolingual peers in executive control in particular (e.g. Barac & Bialystok, 2012). Therefore, one might hypothesize that bilingual children should be showing an advantage in mathematics achievement, since the two are related. However, at the same time, there seem to be costs associated with performing mathematical operations in a language other than the language of instruction (Saalbach, Eckstein, Andri, Hobi, & Grabner, 2013). People tend to perform mental calculations in their native language, even when fluent in a second language (Dehaene, 2011), which suggests that number words may be less flexible when it comes to switching between languages. Relationships between bilingualism, cognitive control, and mathematics warrant further investigation and could have important implications for English as an additional language (EAL) students, who make up an increasingly large number of students in classrooms in the UK, as well as Canada and the US.

6.6. Concluding Remarks

To conclude, the research presented in this thesis has contributed to our understanding of the development of numeracy in early childhood and how it relates to domain-general cognitive processes. Cognitive processes develop dynamically and while it has been proposed that attention constrains learning, there is little evidence to support a unidirectional relationship. Therefore, future research should focus on interactions across domains and not only on how attention constrains learning but also how previous experience influences attention. The research presented here demonstrates that theories of early learning need to account for factors beyond domain-specific knowledge. Further experimental investigations of relationships between domain-general and domain-specific factors are required in order to uncover the mechanisms underpinning these relationships that have been established through correlational evidence.

References

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–9. doi:10.1016/j.jecp.2009.11.003

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

Ansari, D., Lyons, I. M., van Eimeren, L., & Xu, F. (2007). Linking visual attention and number processing in the brain: the role of the temporo-parietal junction in small and large symbolic and nonsymbolic number comparison. *Journal of Cognitive Neuroscience*, *19*(11), 1845–53. doi:10.1162/jocn.2007.19.11.1845

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

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: role of language, cultural background, and education. *Child Development*, *83*(2), 413–22. doi:10.1111/j.1467-8624.2011.01707.x

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. doi:10.1016/S1364-6613(00)01538-2

Batchelor, S., Keeble, S., & Gilmore, C. (2015). Magnitude Representations and Counting Skills in Preschool Children. *Mathematical Thinking and Learning*,*17*(2-3), 116-135.

Benoit, L., Lehalle, H., Molina, M., Tijus, C., & Jouen, F. (2013). Young children's mapping between arrays, number words, and digits. *Cognition*, *129*(1), 95–101. doi:10.1016/j.cognition.2013.06.005

Bergman Nutley, S., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: a controlled, randomized study. *Developmental Science*, *14*(3), 591–601. doi:10.1111/j.1467-7687.2010.01022.x

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between Executive Function and Academic Achievement from Ages 5 to 17 in a Large, Representative National Sample. *Learning and Individual Differences*, *21*(4), 327–336. doi:10.1016/j.lindif.2011.01.007

Blair, C., Knipe, H., & Gamson, D. (2008). Is There a Role for Executive Functions in the Development of Mathematics Ability? *Mind, Brain, and Education*, *2*(2), 80–89. doi:10.1111/j.1751-228X.2008.00036.x

Blair, C., & Raver, C. C. (2014). Closing the Achievement Gap through Modification of Neurocognitive and Neuroendocrine Function: Results from a Cluster Randomized

Controlled Trial of an Innovative Approach to the Education of Children in Kindergarten. *PloS One*, *9*(11), e112393. doi:10.1371/journal.pone.0112393

Borst, G., Poirel, N., Pineau, A., Cassotti, M., Houdé, O. (2012). Inhibitory control in number-conservation and class-inclusion tasks: A neo-Piagetian inter-task priming study. *Cognitive Development, 27*, 283-298. doi:10.1016/j.cogdev.2012.02.004

Brankaer, C., Ghesquière, P., & De Smedt, B. (2014). Children's mapping between non-symbolic and symbolic numerical magnitudes and its association with timed and untimed tests of mathematics achievement. *PloS one*, *9*(4), e93565.

Bryce, D., Szũcs, D., Soltész, F., & Whitebread, D. (2011). The development of inhibitory control: an averaged and single-trial Lateralized Readiness Potential study. *NeuroImage*, *57*(3), 671–85. doi:10.1016/j.neuroimage.2010.12.006

Bull, R., Espy, K. A., Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical acheivement at age 7 years. *Developmental Neuropsychology*, *33*(3), 205-228. doi:10.1080/87565640801982312.

Bull, R., Espy, K. A., Wiebe, S. A., Sheffield, T. D., & Nelson, J. M. (2011). Using confirmatory factor analysis to understand executive control in preschool children: sources of variation in emergent mathematic achievement. *Developmental Science*, *14*(4), 679–92. doi:10.1111/j.1467-7687.2010.01012.x

Bull, R., & Lee, K. (2014). Executive Functioning and Mathematics Achievement. *Child Development Perspectives*, *8*(1), 36–41. doi:10.1111/cdep.12059

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*(3), 273–93. doi:10.1207/S15326942DN1903_3

Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage*, *87*, 311–22. doi:10.1016/j.neuroimage.2013.10.049

Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2015). Visual Number Beats Abstract Numerical Magnitude. *Journal of Cognitive Neuroscience*, 1–10. doi:10.1162/jocn

Burgess, P. W. (1997). Theory and methodology in executive function research. *Methodology of frontal and executive function*, 81-116.

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, *46*(1), 3–18. doi:10.1111/j.1469-7610.2005.00374.x

Carey, S. (2009). *The Origin of Concepts.* New York: Oxford University Press.

Cantrell, L., & Smith, L. B. (2013). Open questions and a proposal: a critical review of the evidence on infant numerical abilities. *Cognition*, *128*(3), 331–52. doi:10.1016/j.cognition.2013.04.008

Cantrell, L., Kuwabara, M., & Smith, L. B. (2015). Set size and culture influence children's attention to number. *Journal of Experimental Child Psychology*, *131*, 19-37.

Carroll D. J., Cragg L. (2012). Precursors to cognitive flexibility: rule-switching in two-year-olds, Paper presented at the Development of Executive Functions Workshop (Utrecht, Netherlands)

Chacko, A., Bedard, A. C., Marks, D. J., Feirsen, N., Uderman, J. Z., Chimiklis, A., ... & Ramon, M. (2014). A randomized clinical trial of Cogmed Working Memory Training in school-age children with ADHD: A replication in a diverse sample using a control condition. *Journal of Child Psychology and Psychiatry*,*55*(3), 247-255.

Chooi, W. T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*(6), 531–542. doi:10.1016/j.intell.2012.07.004

Chu, F. W., vanMarle, K., & Geary, D. C. (2015). Early numerical foundations of young children's mathematical development. *Journal of Experimental Child Psychology*, 1–8. doi:10.1016/j.jecp.2015.01.006

Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology*, *46*(5), 1176–91. doi:10.1037/a0019672

Clark, C. A. C., Sheffield, T. D., Wiebe, S. A., Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Development, 84*(2), 662-677. doi: 10.1111/j.1467-8624.2012.01854.x

Clayton, S., & Gilmore, C. (2015). Inhibition in dot comparison tasks. *ZDM*. 1-12. doi:10.1007/s11858-014-0655-2

Cohen Kadosh, R., & Walsh, V. (2009). Numerical representation in the parietal lobes: abstract or not abstract? *The Behavioral and Brain Sciences*, *32*(3-4), 313–28; discussion 328–73. doi:10.1017/S0140525X09990938

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201-215.

Cordes, S., & Brannon, E. M. (2010). The relative salience of discrete and continuous quantity in young infants, *12*(3), 453–463. doi:10.1111/j.1467-7687.2008.00781.x

Cragg, L., & Chevalier, N. (2012). The processes underlying flexibility in childhood. *The Quarterly Journal of Experimental Psychology*, *65*(2), 209-232.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics : The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education.* doi: 10.1016/j.tine.2013.12.001

Dehaene, S. (2008). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In Haggard, P. & Rossetti, Y. (Eds.) *Sensorimotor Foundations of Higher Cognition: Attention and Performance 22* (pp. 527-74). New York: Oxford University Press.

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition.* New York: Oxford University Press.

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

De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, 1–8. doi:10.1016/j.tine.2013.06.001

Defever, E., Reynvoet, B., & Gebuis, T. (2013). Task- and age-dependent effects of visual stimulus properties on children's explicit numerosity judgments. *Journal of Experimental Child Psychology, 116,* 213-233. doi: 10.1016/j.jecp/2013/04.006

Demeyere, N., Rotshtein, P., & Humphreys, G. W. (2014). Common and dissociated mechanisms for estimating large and small dot arrays: value-specific fMRI adaptation. *Human Brain Mapping*, *35*(8), 3988–4001. doi:10.1002/hbm.22453

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science (New York, NY)*, *318*(5855), 1387.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science (New York, N.Y.)*, *333*(6045), 959–64. doi:10.1126/science.1204529

Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Developmental Science*, *11*(5), 650–4. doi:10.1111/j.1467-7687.2008.00713.x

Dunn, L. M., & Dunn, D. M. (2009). *The British picture vocabulary scale*. GL Assessment Limited.

Elliot, C. D., Smith, P., & McCulloch, K. (1996). British ability scales II. *Windsor: NFER-Nelson*.

Espy, K. A. (1997). Developmental Cognitive Neuroscience Laboratory - Faculty and Staff Publications The Shape School : Assessing executive function in preschool children. *Developmental Neuropsychology*, *13*(4), 495–499.

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*, 53–72. doi:10.1016/j.jecp.2014.01.013

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–14. doi:10.1016/j.tics.2004.05.002

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology. General*, *133*(1), 101–35. doi:10.1037/0096-3445.133.1.101

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: contributions of inhibitory control. *Developmental Science*, *16*(1), 136–48. doi:10.1111/desc.12013

Gabay, S., Leibovich, T., Henik, A., & Gronau, N. (2013). Size before numbers: conceptual size primes numerical value. *Cognition*, *129*(1), 18–23. doi:10.1016/j.cognition.2013.06.001

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60. doi:10.1037/0033-2909.134.1.31

Gathercole, S. E., Dunning, D. L., & Holmes, J. (2012). Cogmed training: Let's be realistic about intervention research. *Journal of Applied Research in Memory and Cognition*, *1*(3), 201–203. doi:10.1016/j.jarmac.2012.07.007

Gebuis, T., & Reynvoet, B. (2012a). The role of visual information in numerosity estimation. *PloS One*, *7*(5), e37426. doi:10.1371/journal.pone.0037426

Gebuis, T., & Reynvoet, B. (2012b). The interplay between nonsymbolc number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*(4), 642-648. doi:10.1371/journal.pone.0037426

Gelman, R., & Gallistel, C. R. (1978). The child's understanding of number. Cambridge, MA: Harvard University Press.

Gerstadt, C., Hong, Y.,& Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2 –7 years old on a Stroop-like day-night test. Cognition, 53, 129–153.

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., Attridge, N., De Smedt, B., & Inglis, M. (2014). Measuring the approximate number system in children: Exploring the relationships among different tasks. *Learning and Individual Differences*, *29*, 50-58.

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PloS one*, *8*(6), e67374. doi:10.1371/journal.pone.0067374

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *ZDM,* 1-12. doi:10.1007/s11858-014-0659-y

Ginsburg, H. P., & Baroody, A. J. (2003). The Test of Early Mathematics Ability (3rd ed.). Austin, TX: Pro Ed.

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), 789–98. doi:10.1177/0956797613516471

Goldberg, E. (2002). *The executive brain: Frontal lobes and the civilized mind*. Oxford University Press.

Goldin, A. P., Hermida, M. J., Shalom, D. E., Elias Costa, M., Lopez-Rosenfeld, M., Segretin, M. S., … Sigman, M. (2014). Far transfer to language and math of a short software-based gaming intervention. *Proceedings of the National Academy of Sciences of the United States of America*, 1–6. doi:10.1073/pnas.1320217111

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457–65. doi:10.1037/a0012682

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–8. doi:10.1038/nature07246

Hannula, M. M., & Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learning and Instruction*, *15*(3), 237-256.

Hannula, M. M., Lepola, J., & Lehtinen, E. (2010). Spontaneous focusing on numerosity as a domain-specific predictor of arithmetical skills. *Journal of Experimental Child Psychology*, *107*(4), 394–406. doi:10.1016/j.jecp.2010.06.004

Hassinger-Das, B., Jordan, N. C., Glutting, J., Irwin, C., & Dyson, N. (2014). Domain-general mediators of the relation between kindergarten number sense and first-grade mathematics achievement. *Journal of Experimental Child Psychology*, *118*, 78–92. doi:10.1016/j.jecp.2013.09.008

Heckman, J. J. (2006). Investing in Disadvantaged Children, *312*(June), 2005–2007.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–15. doi:10.1111/j.1467-7687.2009.00848.x

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

Howard-Jones, P. (2012). The good, the bad and the ugly in neuroscience and education: an. *Neuroscience in Education: The good, the bad, and the ugly*, 337.

Huang, Y. T., Spelke, E., & Snedeker, J. (2010). When is four far more than three? Children's generalization of newly acquired number words. *Psychological Science*, *21*(4), 600–6. doi:10.1177/0956797610363552

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: a latent variable approach. *Developmental Neuropsychology*, *35*(1), 20–36. doi:10.1080/87565640903325691

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), 2017-2036.

Hurewitz, F., Gelman, R., and Schnitzer, B. (2006). Sometimes area counts more than number. Proceedings of the National Academy of Science, *103*, 19599-19604. doi: 10.1073/pnas.0609485103

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*(1), 92–107. doi:10.1016/j.cognition.2013.12.007

Hyde, D. C., & Wood, J. N. (2011). Spatial attention determines the nature of nonverbal number representation. *Journal of Cognitive Neuroscience*, *23*(9), 2336-2351.

Jacques, S., & Zelazo, P. D. (2012). Developmental Neuropsychology The Flexible Item Selection Task (FIST): A Measure of Executive Function in Preschoolers. *Developmental Neuropsychology*, (February).

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–33. doi:10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, *42*(3), 464–80. doi:10.3758/s13421-013-0364-z

Jolles, D. D., & Crone, E. A. (2012). Training the developing brain: a neurocognitive perspective. *Frontiers in human neuroscience*, *6*.

Jordan, K. E., & Baker, J. (2010). Multisensory information boosts numerical matching abilities in young children. *Developmental Science*, *14*(2), 205–213. doi:10.1111/j.1467-7687.2010.00966.x

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*(6), 978-990.

Karmiloff-Smith, A., D'Souza, D., Dekker, T. M., Van Herwegen, J., Xu, F., Rodic, M., & Ansari, D. (2012). Genetic and environmental vulnerabilities in children with neurodevelopmental disorders. *Proceedings of the National Academy of Sciences*, *109*(Supplement 2), 17261-17265.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, 498-525.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in cognitive sciences*, *14*(7), 317-324.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*(6), 781–91. doi:10.1076/jcen.24.6.781.8395

Kloo, D., & Perner, J. (2003). Training transfer between card sorting and false belief understanding: Helping children apply conflicting descriptions. *Child development*, *74*(6), 1823-1839.

Konkle, T., & Oliva, A. (2012). A familiar-size Stroop effect: real-world size is an automatic property of object representation. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(3), 561–9. doi:10.1037/a0028294

Kroesbergen, E. H., van 't Noordende, J. E., & Kolkman, M. E. (2014). Training working memory in kindergarten children: Effects on working memory and early numeracy. *Child Neuropsychology*, *20*(1), 23-37.

Kroesbergen, E. H., Van Luit, J. E. H., Van Lieshout, E. C. D. M., Van Loosbroek, E., & Van de Rijt, B. A. M. (2009). Individual Differences in Early Numeracy: The Role of Executive Functions and Subitizing. *Journal of Psychoeducational Assessment*, *27*(3), 226–236. doi:10.1177/0734282908330586

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size. *PloS One*, *9*(9), e105825. doi:10.1371/journal.pone.0105825

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438. doi:10.1016/j.cognition.2006.10.005

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.

LeFevre, J.-A., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S.-L., & Smith-Chant, B. L. (2013). The role of executive attention in the acquisition of mathematical skills for children in Grades 2 through 4. *Journal of Experimental Child Psychology*, *114*(2), 243–61. doi:10.1016/j.jecp.2012.10.005

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*(1), 59-80.

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology*, *4*(June), 375. doi:10.3389/fpsyg.2013.00375

Libertus, M. E., Woldorff, M. G., & Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behavioral and Brain Functions : BBF*, *3*, 1. doi:10.1186/1744-9081-3-1

Libertus, M. E., Feigenson, L., Halberda, J., & Landau, B. (2014). Understanding the mapping between numerical approximation and number words: evidence from Williams syndrome and typical development. *Developmental Science*, *17*(6), 905–19. doi:10.1111/desc.12154

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in cognitive sciences*, *17*(8), 391-400.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking redundant labels facilitate learning of novel categories, *18*(12), 1077–1083.

Lyons, I. M., & Ansari, D. (2009). The cerebral basis of mapping nonsymbolic numerical quantities onto abstract symbols: an fMRI training study. *Journal of Cognitive Neuroscience*, *21*(9), 1720–35. doi:10.1162/jocn.2009.21124

Lyons, I. M., & Beilock, S. L. (2009). Beyond quantity: individual differences in working memory and the ordinal understanding of numerical symbols. *Cognition*, *113*(2), 189–204. doi:10.1016/j.cognition.2009.08.003

Lyons, I. M., Ansari, D., & Beilock, S. L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology. General*. doi:10.1037/a0027248

Lyons, I. M., Ansari, D., & Beilock, S. L. (2014). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping*, *00*, n/a–n/a. doi:10.1002/hbm.22641

Lyons, I. M., & Beilock, S. L. (2013). Ordinality and the nature of symbolic numbers. *The Journal of Neuroscience* , *33*(43), 17052–61. doi:10.1523/JNEUROSCI.1775-13.2013

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental science*, *17*(5), 714-726.

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.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' Precision of the Approximate Number System Predicts Later School Mathematics Performance. *PLoS ONE*, *6*(9), e23749. doi:10.1371/journal.pone.0023749

Melby-Lervåg, M., & Hulme, C. (2012). Is Working Memory Training Effective? A Meta-Analytic Review. *Developmental Psychology*. doi:10.1037/a0028228

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Multiple cues for quantification in infancy: Is number one of them? *Psychological Bulletin*, *128*(2), 278–294. doi:10.1037//0033-2909.128.2.278

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. doi:10.1006/cogp.1999.0734

Moyer, R. S.,and Landauer,T. K. (1967). Time required for judgments of numerical inequality. *Nature 215*, 1519–1520.

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*(4), 490–502. doi:10.1016/j.jecp.2009.02.003

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends in Neuroscience and Education*, *1*(1), 21–31. doi:10.1016/j.tine.2012.09.003

Naglieri, J. A., & Das, J. P. (1997). Cognitive Assessment System.

Nathan, M. J., & Wagner Alibali, M. (2010). Learning sciences. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(3), 329–345. doi:10.1002/wcs.54

Negen, J., & Sarnecka, B. W. (2014). Is there really a link between exact-number knowledge and approximate number system acuity in young children? *The British Journal of Developmental Psychology*. doi:10.1111/bjdp.12071

Nieder, A.,and Dehaene, S. (2009). Representation of number in the brain. *Annual Reviews of Neuroscience, 32*, 185–208.

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. doi:10.1037//0033-2909.126.2.220

Noël, M. P., & Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: an explanation based on a double exact-and-approximate number representation model. *Frontiers in Human Neuroscience*, *5*.

Norbury, C. F., Gooch, D., Baird, G., Charman, T., Simonoff, E. and Pickles, A. (2015), Younger children experience lower levels of language competence and academic progress in the first year of school: evidence from a population study. Journal of Child Psychology and Psychiatry. doi: 10.1111/jcpp.12431

Nys, J., & Content, A. (2012). Judgement of discrete and continuous quantity in adults: Number counts!. *The quarterly journal of experimental psychology*,*65*(4), 675-690.

Odic, D., Le Corre, M., & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition*, *138C*, 102–121. doi:10.1016/j.cognition.2015.01.008

Ofsted (2014, April). Early Years Annual Report 2012/2013. Retrieved April 29, 2014, from: https://www.gov.uk/government/publications/ofsted-early-years-annual-report-201213

Omniglot. (2013). *Tékumel writing systems.* Retrieved April 22, 2012, from http://www.omniglot.com/writing/tekumel.php

Park, J., & Brannon, E. M. (2013). Training the Approximate Number System Improves Math Proficiency. *Psychological Science*, (August). doi:10.1177/0956797613482944

Passolunghi, M. C., & Costa, H. M. (2014). Working memory and early numeracy training in preschool children. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, (November), 1–18. doi:10.1080/09297049.2014.971726

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*(12), 542–51. doi:10.1016/j.tics.2010.09.008

Poirel, N., Borst, G., Simon, G., Rossi, S., Cassotti, M., Pineau, A., & Houdé, O. (2012). Number conservation is related to children's prefrontal inhibitory control: an fMRI study of a piagetian task. *PloS One*, *7*(7), e40802. doi:10.1371/journal.pone.0040802

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42. doi:10.1146/annurev.ne.13.030190.000325

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annu. Rev. Psychol.*, *58*, 1-23.

Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica*, *140*(1), 50–7. doi:10.1016/j.actpsy.2012.02.008

Rabipour, S., & Raz, A. (2012). Training the brain: Fact and fad in cognitive and behavioral remediation. *Brain and Cognition*, *79*(2), 159-179.

Ristic, J., & Enns, J. T. (2015). The Changing Face of Attentional Development. *Current Directions in Psychological Science*, *24*(1), 24–31. doi:10.1177/0963721414551165

Robinson, K. M., & Dubé, A. K. (2013). Children's additive concepts: Promoting understanding and the role of inhibition. *Learning and Individual Differences*, *23*, 101–107. doi:10.1016/j.lindif.2012.07.016

Roid, G., & Miller, L. (1997). *Leiter International Performance Scale–Revised*.

Rousselle, L., & Noël, M.-P. (2008). The development of automatic numerosity processing in preschoolers: evidence for numerosity-perceptual interference. *Developmental Psychology*, *44*(2), 544–60. doi:10.1037/0012-1649.44.2.544

Rousselle, L., Palmers, E., & Noël, M.-P. (2004). Magnitude comparison in preschoolers: what counts? Influence of perceptual variables. *Journal of Experimental Child Psychology*, *87*(1), 57–84. doi:10.1016/j.jecp.2003.10.005

Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(41), 14931–6. doi:10.1073/pnas.0506897102

Saalbach, H., Eckstein, D., Andri, N., Hobi, R., & Grabner, R. H. (2013). When language of instruction and language of application differ: Cognitive costs of bilingual mathematics learning. *Learning and Instruction*, *26*, 36–44. doi:10.1016/j.learninstruc.2013.01.002

Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2013). The approximate number system is not predictive for symbolic number processing in kindergartners. *The Quarterly Journal of Experimental Psychology*, (May), 1–26. doi:10.1080/17470218.2013.803581

Sathian, K., Simon, T. J., Peterson, S., Patel, G. A., Hoffman, J. M., & Grafton, S. T. (1999). Neural evidence linking visual object enumeration and attention. *Journal of Cognitive Neuroscience*, *11*(1), 36-51.

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1177-1196.

Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2015). Spontaneous non-verbal counting in toddlers. *Developmental Science*, n/a–n/a. doi:10.1111/desc.12299

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). Cogmed Working Memory Training: Does the Evidence Support the Claims? *Journal of Applied Research in Memory and Cognition*, 1–9. doi:10.1016/j.jarmac.2012.06.003

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–54. doi:10.1037/a0027473

Slusser, E. B., & Sarnecka, B. W. (2011). Find the picture of eight turtles: A link between children's counting and their knowledge of number word semantics.*Journal of experimental child psychology*, *110*(1), 38-51.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology (2006)*, *59*(4), 745–59. doi:10.1080/17470210500162854

St Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, *30*(2), 203-219.

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, 1–5. doi:10.1073/pnas.1302751110

Steele, A., Karmiloff-Smith, A., Cornish, K., & Scerif, G. (2012). The multiple subfunctions of attention: differential developmental gateways to literacy and numeracy. *Child Development*, *83*(6), 2028–41. doi:10.1111/j.1467-8624.2012.01809.x

Szűcs, D., & Csépe, V. (2004). Similarities and differences in the coding of numerical and alphabetical order using acoustic stimulation as revealed by event-related potentials in humans. *Neuroscience Letters*, *360*(1-2), 65–8. doi:10.1016/j.neulet.2004.02.038

Szűcs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in Psychology*, *4*(July), 444. doi:10.3389/fpsyg.2013.00444

Szűcs, D., Soltész, F., Bryce, D., & Whitebread, D. (2009). Real-time tracking of motor response activation and response competition in a Stroop task in young children: a lateralized readiness potential study. *Journal of Cognitive Neuroscience*, *21*(11), 2195–206. doi:10.1162/jocn.2009.21220

Temple, E., & Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(13), 7836–41. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=22775&tool=pmcentrez&rendertype=abstract

Thompson, T. W., Waskom, M. L., Garel, K.-L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., … Gabrieli, J. D. E. (2013). Failure of Working Memory Training to Enhance Cognition or Intelligence. *PLoS ONE*, *8*(5), e63614. doi:10.1371/journal.pone.0063614

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, *12*(1), 106–13. doi:10.1111/j.1467-7687.2008.00745.x

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80.

Turconi, E., Campbell, J. I. D., & Seron, X. (2006). Numerical order and quantity processing in number comparison. *Cognition*, *98*(3), 273–85. doi:10.1016/j.cognition.2004.12.002

Turconi, E., Jemel, B., Rossion, B., & Seron, X. (2004). Electrophysiological evidence for differential processing of numerical quantity and order in humans. *Brain Research. Cognitive Brain Research*, *21*(1), 22–38. doi:10.1016/j.cogbrainres.2004.05.003

Tzelgov, J., Yehene, V., Kotler, L., & Alon, A. (2000). Automatic comparisons of artificial digits never compared: learning linear ordering relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 103.

Van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2012). The development of executive functions and early mathematics: A dynamic relationship. *The British Journal of Educational Psychology*, *82*(Pt 1), 100–19. doi:10.1111/j.2044-8279.2011.02035.x

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2014). Does numerical processing uniquely predict first graders' future development of single-digit arithmetic? *Learning and Individual Differences*. doi:10.1016/j.lindif.2014.12.004

Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. Cognition, 119,10–22. doi:10.1016/j.cognition. 2010.11.014

Wallace, D. F. (2011). *The Pale King*. Little, Brown.

Wass, S. V., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory – Is younger, better? *Developmental Review*, *32*(4), 360–387. doi:10.1016/j.dr.2012.07.001

Wiebe, S. A. Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*(3), 436–52. doi:10.1016/j.jecp.2010.08.008

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, *44*(2), 575–87. doi: 10.1037/0012-1649.44.2.575.

Wiese, H. (2003). Iconic and non-iconic stages in number development: the role of language. *Trends in cognitive sciences*, *7*(9), 385-390.

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., & Cohen, D. (2006). Principles underlying the design of “The Number Race”, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions: BBF*, *2*, 19. doi:10.1186/1744-9081-2-19

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of “The Number Race”, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions: BBF*, *2*, 20. doi:10.1186/1744-9081-2-20

Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an Adaptive Game Intervention on Accessing Number Sense in Kindergarten Children. *Mind, Brain, and Education*, *3*(4), 224–234.

Witt, M. (2011). School based working memory training : Preliminary finding of improvement in children ’ s mathematical performance. *Advances in Cognitive Psychology, 7*(0), 7–15.

Wright,I., Waterman, M., Prescott, H., Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function development: typical developmental trends. *Journal of Child Psychology and Psychiatry, 44*(4), 561-75.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155-193.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220-251.

Xenidou-Dervou, I., De Smedt, B., van der Schoot, M., & van Lieshout, E. C. D. M. (2013). Individual differences in kindergarten math achievement: The integrative roles of approximation skills and working memory. *Learning and Individual Differences*, *28*, 119–129. doi:10.1016/j.lindif.2013.09.012

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences*, *23*, 1–9. doi:10.1016/j.lindif.2012.10.004

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nature Protocols*, *1*(1), 297–301. doi:10.1038/nprot.2006.

Zhao, H., Chen, C., Zhang, H., Zhou, X., Mei, L., Chen, C., Chen, L., Cao, Z., Dong, Q. (2012). Is order the defining feature of magnitude representation? An ERP study on learning numerical magnitude and spatial order of artificial symbols. *PloS One*, *7*(11), e49565. doi:10.1371/journal.pone.0049565

Appendix A: Six month follow up to Experiment 3

In order to investigate whether the training effects were maintained, a subset of the children (N = 17) was assessed again six months following the initial post-assessment. Given the null training effects, this was instead seen as an opportunity to investigate longitudinal predictors of early numeracy. It was not possible to see all of the children at this time as many had left Oxford and some parents did not have time to participate and therefore the sample consisted of 17 children between the ages of 43 and 72 months (*M* = 60, *SD* = 8). Eight of the participants were female, and participants were equally split across the two training groups. Many of the children seen in this follow up had transitioned from nursery to reception in the time since training. Correlations with age in months partialled out are reported in Table 1. Verbal and nonverbal IQ, early numeracy, and selective attention at Time 1 were all significantly positively correlated with both TEMA and FIST at time 3. Visual-spatial short-term memory at Time 1 was significantly positively correlated with TEMA at Time 3, and inhibition was significantly negatively correlated with FIST at Time 3. FIST and TEMA were related to each other concurrently at Time 3 despite not having been significantly correlated with each other at Time 1 or Time 2.

Table 1. Partial correlations between baseline assessments and assessments at Time 3.

	TEMA at Time 3	FIST at Time 3
BPVS	.71**	.78**
BAS	.86**	.59*
TEMA	.9**	.71**
Visual Search	.81**	.61*
FIST Switch Accuracy	.23	.43
Digit Recall	.03	-.18
Dot Matrix	.53*	.42
Odd-One-Out	-.14	-.27
Spatial Conflict Accuracy Conflict	-.14	-.52*
Spatial Conflict RT Conflict	-.23	-.15
TEMA (Time 2)	.92**	.77**
FIST (Time 2)	.53	.77**
TEMA (Time 3)	-	.69**

*Note: *p < .05; **p < .001; Assessments at Time 1 unless otherwise specified*