THE ROLE OF INTUITIVE ARITHMETIC IN DEVELOPING MATHEMATICAL SKILL

Emily Maja Szkudlarek

A DISSERTATION

in

Psychology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

____

Elizabeth M. Brannon

Professor of Psychology

Graduate Group Chairperson

____

John Trueswell, Professor of Psychology

Dissertation Committee

Daniel Swingley, Professor of Psychology

Allyson Mackey, Assistant Professor of Psychology

ACKNOWLEDGMENT

First, I would like to thank my advisor, Dr. Elizabeth Brannon, for her guidance and support throughout my time in graduate school. I am forever grateful for the opportunity to work in the Brannon lab. I would also like to thank the members of my committee, Dr. Daniel Swingley and Dr. Allyson Mackey for their advice and feedback regarding this research. I am further thankful to Dr. Whitney Tabor and Dr. Etan Markus for their endless support of my research interests as an undergraduate at the University of Connecticut. Without their mentorship, this dissertation would not be possible.

Many past and present members of the Brannon lab have offered help and advice during my time in graduate school, and to all of them I am thankful. Two members of the lab in particular, Nick DeWind and Stephanie Bugden, have contributed to the current work in innumerable ways. I cannot imagine graduate school without either of them. Our near daily chats about work and life (whether in the office or not) have impacted this research and beyond. I am immensely grateful for our friendship.

I would also like to thank Eleni Kursten and Emily Kibbler for their consistent belief that I am moving in the right direction.

I am also thankful to my family for their continual interest in what I've been working on, especially Kim Szkudlarek, Celeste Truskolawski, Małgorzata Podłuska, Alicja Szkudlarek, Marek Szyndel, and my grandmother, Valerie Truskolawski. Though my grandmother is not here to see the end of my time in graduate school, I know she was already excited for the next things to come.

To my parents, Justine and Lech Szkudlarek, thank you for everything. There are truly too many things to list, but a few include helping me to move (three) times, timely reminders that my car is due for an oil change, and sparking my interest in science.

Finally, thanks to my parakeets Pickle, Parsnip, and Ollie for their happy chirps each morning that became the soundtrack of my graduate career.

ABSTRACT

THE ROLE OF INTUITIVE ARITHMETIC IN DEVELOPING MATHEMATICAL SKILL

Emily Maja Szkudlarek

Elizabeth M. Brannon

Symbolic mathematics allows humans to represent and describe the logic of the world around us. Although we typically think about math symbolically, humans across the lifespan and a wide variety of animal species spontaneously exhibit numerical competence without reference to formal mathematics. This intuitive ability to approximately compare, estimate, and manipulate large non-symbolic numerical quantities without language or symbols is called the Approximate Number System. The four chapters of this dissertation explore whether non-symbolic, approximate calculation can function as a bridge between our Approximate Number System and symbolic mathematics for children at the beginning of formal math education and university undergraduates. Chapter 1 explores how non-symbolic and symbolic ratio reasoning relates to general math skill and Approximate Number System acuity in elementary school children. Chapter 2 examines whether children and adults can perform a non-symbolic, approximate division computation, and how this ability relates to non-symbolic and symbolic mathematical skill. Chapter 3 tests the robustness and mechanism of a non-symbolic, approximate addition and subtraction training paradigm designed to improve arithmetic fluency in university undergraduates. Chapter 4 investigates whether the negative relation between math anxiety and symbolic math performance extends to approximate, non-symbolic calculation. Together, Chapters 1 and 2 provide evidence that non-symbolic calculation ability functions as a mechanism of the relation between Approximate Number System acuity and symbolic math. Chapters 3 and 4 identify populations of students for whom practice with non-symbolic calculation may or may not be beneficial. In sum, this dissertation describes how non-symbolic, approximate calculation allows students harness their intuitive sense of number in a mathematical context.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

GENERAL INTRODUCTION

Introduction to approximate number representations

Imagine that you are snorkeling and you encounter many schools of brightly colored tropical fish. When you swim back to the beach, your friend asks whether you saw more black angelfish or yellow butterfly fish. In this scenario you could probably provide an answer to this question without too much thought. Moreover, if asked, you could likely provide an estimate of how many yellow butterfly fish you saw, flexibly switch between comparing the relative number of orange and red fish to the blue and yellow fish, and then compare these numbers to the quantities you remember from your last snorkeling trip. This intuitive ability to approximately compare, estimate, and manipulate large non-symbolic numerical quantities without language or symbols is called the Approximate Number System (ANS; Feigenson, Dehaene, & Spelke, 2004). Humans across the lifespan and a wide variety of animal species spontaneously exhibit this numerical competence. For example, the ANS supports the ability of both wild North Island robins and captive elephants to pick the greater quantity of food, newborn infants to match numerical information across visual arrays of objects and a series of sounds, and child and adult members of an indigene group in the Amazon to perform intuitive arithmetic without formal education in mathematics (Garland, Low, & Burns, 2012; Izard, Sann, Spelke, & Streri, 2009; Perdue, Talbot, Stone, & Beran, 2012; Pica, Lemer, Izard, & Dehaene, 2004). After practice comparing quantities of dots in a laboratory setting, adults can successfully discriminate between 23 dots from 26 dots when they are flashed on a screen for 500ms, even when the size, spatial layout, and total surface area of the dots change (Cochrane, Cui, Hubbard, & Green, 2018; DeWind, Adams, Platt, & Brannon, 2015). Thus, the ANS allows humans and non-humans to intuitively use numerical information present in the world around us.

The approximate number system is not the only way in which humans represent number. We can precisely enumerate collections of less than four items (Feigenson, Carey, & Hauser, 2002), and with the help of language we can exactly count how many objects there are, or we can use numerals to symbolically represent a cardinal value. In contrast to these other systems of numerical representation, ANS representations are approximate and noisy. This quality results in the behavioral hallmark of the ANS of ratio dependence, or Weber's law. When we use the ANS the difficulty in discriminating two numerical quantities is dependent on the ratio between the two numerosities and not their absolute value. For example, it is hard to tell which school of fish is larger when comparing 15 black fish and 16 yellow fish but it is much easier when comparing 25 black fish and 5 yellow fish. Moreover comparing 125 black fish and 25 yellow fish is just as easy as comparing 25 and 5. Theoretically, this behavioral signature suggests that the ANS represents numbers as a series of overlapping Gaussian distributions with logarithmic compression as numerosity increases along an ordered mental number line (Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). In this model, there is more overlap in the distributions representing 15 and 16 than there are in in the distributions representing 25 and 5, and therefore 25 and 5 are easier to tell apart. Individual differences in the acuity of the ANS are usually captured by modelling the width of these distributions with a dot comparison task using the Weber fraction (w) (Halberda, Ly, Wilmer, Naiman, & Germine, 2012). In this task subjects see two arrays of dots appear on a screen for a short period of time. The goal is to pick which array of dots has the greater number. Researchers manipulate the difficulty of this task by varying the ratio between the two dot arrays. The closer the ratio is to one, the harder it is to discriminate between the two arrays. The weber fraction corresponds to the ratio closest to one that a person can reliably discriminate, and theoretically to the width of the Gaussian neural tuning curves that instantiate the ANS (Dehaene, Piazza, Pinel, & Cohen, 2003; Piazza et al., 2004). ANS acuity varies meaningfully across the population (Halberda et al., 2012; Halberda, Mazzocco, & Feigenson, 2008). More precise ANS acuity corresponds to narrower Gaussian distributions with less overlap between numerosities, while less precise ANS representations correspond to wider distributions for each numerical value.

The abilities afforded by the ANS – non-symbolic and approximate representation of number - is not a skill normally thought of as mathematics. Math is highly dependent on exact, symbolic representations of number. For example, to perform symbolic, exact calculations a student needs to learn that the difference between 8 and 7 is the exact same difference as between 1,000 and 1,001. Due to the approximate nature of ANS representations, this concept cannot be represented using the ANS. Rather, a student must rely on symbolic methods of representing number to capture this mathematical concept. Despite this fundamental difference between math and our approximate number sense, there is evidence of a meaningful relationship between mathematics and the ANS. Three meta-analyses on the relation between symbolic math and ANS acuity have found evidence of a significant correlation between these skills from preschool into adulthood (Chen & Li, 2014; Fazio, Bailey, Thompson, & Siegler, 2014; Schneider et al., 2016). Work with children who have a math specific learning disability has suggested that an ANS deficit plays a role in at least a subset of children with math learning disorders (Piazza et al., 2010; Skagerlund & Träff, 2016). There is also some evidence for a neural representation of number independent of symbolic or non-symbolic format in the intraparietal sulcus (Dehaene et al., 2003; L. He, Zuo, Chen, & Humphreys, 2014; Lussier & Cantlon, 2016; Piazza, Pinel, Le Bihan, & Dehaene, 2007). Taken together, this work suggests a meaningful link between the approximate and symbolic systems of number representation and lends support to the prominent theory that numerical symbols are grounded in approximate number representations (Dehaene, 1997; Gallistel & Gelman, 1992). This idea has important implications for math education. Understanding the cognitive building blocks that underlie mathematical thought may allow educational researchers and teachers to develop pedagogy that incorporates our understanding of how the brain represents and processes numerical information. However, a substantial number of studies have failed to find a relation between ANS acuity and mathematics (Bonny & Lourenco, 2013; Gilmore et al., 2013; Holloway & Ansari, 2009; Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013), an ANS acuity deficit in children with math learning disability (Iuculano, Tang, Hall, & Butterworth, 2008), or a format independent neural code for number (Lyons, Ansari, & Beilock, 2015). These mixed results in the literature call for a deeper understanding of potential mechanisms of a link between ANS acuity and symbolic mathematics to delineate the specificity and usefulness of this relation for improving math education.

Potential mechanisms of the link between the ANS and mathematics

The mixed results on the relation between ANS acuity and math in the literature reflects the fact that a prominent mechanistic theory for the relation between ANS acuity and symbolic math skill has not yet emerged. Does ANS acuity correlate with all types of mathematics? Does sharper ANS acuity only facilitate the acquisition of symbolic math as a child, and then lose relevance into adulthood? Are there particular populations of students who tend to use their ANS representations during math learning? Understanding why a relation between ANS acuity and mathematics could exist is crucial for research that seeks to utilize the ANS to improve math education.

One preliminary theory suggests the ANS could function as an online error detection system during calculation (Feigenson, Libertus, & Halberda, 2013; Lourenco, Bonny, Fernandez, & Rao, 2012). Students that have a better representation of the numerical magnitudes involved in calculation can imagine a better estimate of the correct solution to a math problem. Therefore, these students would be able to more adeptly reject incorrect solutions resulting in higher accuracy when solving calculation problems. This is an open possibility because there is currently little empirical evidence to judge this claim.

Another theory suggests that children map number words onto ANS representations during their initial acquisition. To understand the meaning of number words, children must acquire the cardinal principle. This is the concept that the number of items in a set corresponds to the last number produced when counting the objects in the set. Learning the cardinal principle is an important milestone in children's mathematical learning. The timing of cardinal principle acquisition predicts later math achievement (Geary et al., 2017). If children base their understanding of number words on their ANS representations, children with sharper ANS acuity will have an easier time mapping to their more precise representations of number. In support of this idea, the relation between ANS acuity and math achievement in preschool is mediated by a children's understanding of the cardinal principle (Chu, vanMarle, & Geary, 2015; van Marle, Chu, Li, & Geary, 2014). Relatedly, a longitudinal study by Shusterman and colleagues found that improved ANS acuity coincided with the understanding of the cardinal principle (Shusterman, Slusser, Halberda, & Odic, 2016). Thus, increased ANS acuity may facilitate acquisition of the cardinal principle. However, if facilitating number word acquisition were the only function ANS acuity played in symbolic math ability then we might expect the relationship to completely dissipate by late childhood once number words are automatized. While the relationship may get weaker over time, a positive correlation between the ANS and symbolic math skills has been demonstrated in multiple adult samples (e.g, DeWind & Brannon, 2012; Halberda et al., 2012).

Another theory is based in the current evidence that practice with non-symbolic approximate arithmetic tasks can improve the symbolic math skills of children and adults (Dillon, Kannan, Dean, Spelke, & Duflo, 2017; Hyde, Khanum, & Spelke, 2014; Obersteiner, Reiss, & Ufer, 2013; Park, Bermudez, Roberts, & Brannon, 2016; Park & Brannon, 2014; Szkudlarek & Brannon, 2018; Wilson, Revkin, Cohen, Cohen, & Dehaene, 2006). Sharper ANS acuity may allow for better non-symbolic, approximate calculation skill because more precise numerical magnitude representation yields a more accurate sense of the quantities involved in calculation. In turn, better non-symbolic calculation skill allows children and adults to create conceptual models of mathematical operations which may facilitate symbolic calculation understanding. The current dissertation explores this hypothesis.

Non-symbolic, approximate calculation

Adults, children, and even infants can successfully calculate non-symbolically and approximately. They can add and subtract arrays of objects (Barth et al., 2006; Barth, La Mont, Lipton, & Spelke, 2005; Gunderson, Ramirez, Beilock, & Levine, 2012; Knops, Viarouge, & Dehaene, 2009; McCrink & Wynn, 2004; McNeil, Fuhs, Keultjes, & Gibson, 2011; Pica et al., 2004; Pinheiro-Chagas et al., 2014; Xenidou-Dervou, van Lieshout, & van der Schoot, 2014). Rhesus macaques, honeybees, and robins also demonstrate basic addition and subtraction ability, indicating this ability is not dependent on language or culture (Cantlon, Merritt, & Brannon, 2015; Garland & Low, 2014; Howard, Avarguès-Weber, Garcia, Greentree, & Dyer, 2019). Young children are also capable of performing multistep increasing and decreasing scaling operations on large quantities of objects (Barth, Baron, Spelke, & Carey, 2009; McCrink, Shafto, & Barth, 2016; McCrink, Spelke, Dehaene, & Pica, 2013; McCrink & Spelke, 2010, 2016), solving addend unknown algebra problems (Kibbe & Feigenson, 2015, 2017), placing a non-symbolic quantity on a number line (Honoré & Noël, 2016), and comparing ratios (Falk, Yudilevich-Assouline, & Elstein, 2012; Ruggeri, Vagharchakian, & Xu, 2018). The ability to represent non-symbolic ratios is even shared by many non-human primates (Drucker, Rossa, & Brannon, 2015; Rakoczy et al., 2014; Tecwyn, Denison, Messer, & Buchsbaum, 2016).

Non-symbolic and approximate calculation tasks are non-symbolic because the numerical values involved in the computation are a concrete depiction of numerical magnitude. They are approximate because a subject does not have to provide an exact answer, but instead compares their answer to another quantity. For example, in one version of a non-symbolic, approximate addition task an array of dots falls into an opaque bucket and then a second array of dots falls into the same bucket. The subject imagines the sum of the two addends, which corresponds to the total amount in the bucket. Then, the subject is asked to judge whether their imagined sum or a new array of dots is greater in quantity. In contrast to this non-symbolic, approximate addition task, you can create exact, symbolic calculation tasks (3 + 4 = 12) or approximate symbolic calculation tasks (estimate the answer of 3 + 4 = ?), or exact non-symbolic tasks (tell me how many objects are in the bucket). While all these tasks may have a place in math education, only non-symbolic, approximate tasks require subjects to calculate with ANS representations.

The majority research on the ANS has focused on individual differences in ANS acuity, rather than its affordance for non-symbolic calculation. Thus, the relation between ANS acuity and the ability to non-symbolically and approximately calculate is not firmly established. While non-symbolic and approximate calculation tasks use ANS representations by definition, it unclear whether sharper ANS acuity results in better non-symbolic calculation skill. From a theoretical perspective, sharper ANS acuity should allow for better representation of the quantities involved in the calculation, and in turn should lead to better approximate computation. Indeed a few studies have found a correlation between sharper ANS acuity and better non-symbolic calculation performance (Matthews, Lewis, & Hubbard, 2016; Pinheiro-Chagas et al., 2014; Starr, Roberts, & Brannon, 2016). Interestingly, this research has also found that non-symbolic calculation is a better predictor of symbolic math than ANS acuity. This suggests that non-symbolic calculation may be an important link between ANS acuity and symbolic mathematics. Sharper ANS acuity creates better non-symbolic calculation skill. In turn, better non-symbolic calculation creates a better conceptual scaffold for symbolic arithmetic. Interventions that attempt to improve ANS acuity alone have largely failed to improve symbolic math skills (Cochrane et al., 2018; Maertens, De Smedt, Sasanguie, Elen, & Reynvoet, 2016; Park & Brannon, 2014; Szűcs & Myers, 2017). Creating tasks where students use ANS representations in a mathematical context rather than attempting to change the acuity of their underlying numerical representations may be a better strategy to increase transfer to symbolic math skills.

The relation between non-symbolic calculation and symbolic math

A small number of experiments have already investigated the link between non-symbolic, approximate arithmetic and symbolic calculation. Children's performance on a non-symbolic, approximate addition task predicts their symbolic math achievement two months later at the end of their kindergarten year (Gilmore, McCarthy, & Spelke, 2010). Children perform with higher accuracy on non-symbolic, approximate addition trials that are arranged with the addends on the left and the comparison array on the right than when the animation is presented on the reverse side (i.e. addition on the right and the target array on the left) (McNeil et al., 2011). The canonical presentation of symbolic addition problems is with the addition portion on the left and the answer on the right (ex. 1 + 1 = 2). An overgeneralization of this convention suggests that children connect their symbolic addition knowledge to the non-symbolic version of addition. Functional MRI indicates common neural activation during non-symbolic and symbolic addition in the bilateral intraparietal sulcus (Venkatraman, Ansari, & Chee, 2005). The activation patterns in the intraparietal sulcus are more similar to each other during non-symbolic and symbolic addition than when performing a color comparison with the same visual stimuli. This finding indicates there is a representation of addition regardless of non-symbolic or symbolic format in the brain

(Bugden, Woldorff, & Brannon, 2019). Finally, a small number of experiments have directly tested whether there is a causal relation between non-symbolic, approximate arithmetic and symbolic math skills using a pretest-training-posttest experimental design. In support of a causal relation between non-symbolic, approximate arithmetic and math, practice adding and subtracting arrays of dots has improved the arithmetic skills of both children and adults (Hyde, Khanum, & Spelke, 2014; Obersteiner, Reiss, & Ufer, 2013; Park, Bermudez, Roberts, & Brannon, 2016; Park & Brannon, 2014; Szkudlarek & Brannon, 2018). Together, this is promising evidence that non-symbolic and symbolic addition and subtraction share a meaningful link.

In further support of this relation, several existing math interventions have included an ANS calculation task within a larger training regime. While a mixed format intervention cannot provide conclusive evidence that the non-symbolic component is necessary for the improvement of symbolic math skills, it does suggest that non-symbolic, approximate arithmetic tasks can be successfully included into a broader math curriculum. Many of the current combined non-symbolic and symbolic math interventions include some variation of a number line task. In a non-symbolic number line task, children see dot arrays or Arabic numerals anchoring both ends of a number line with the goal of placing another dot array on the correct place of the line. A non-symbolic number line task as part of a larger intervention has improved the arithmetic skills and symbolic number line skills of elementary school children (Käser et al., 2013; Kucian et al., 2011; Kuhn & Holling, 2014; Maertens et al., 2016). Non-symbolic addition and subtraction problems as part of a larger intervention have also proved successful at improving arithmetic and basic numeration skills (Dillon et al., 2017; Obersteiner et al., 2013; Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009; Sella, Tressoldi, Lucangeli, & Zorzi, 2016; Wilson, Dehaene, Dubois, & Fayol, 2009; Wilson et al., 2006). This is positive evidence that non-symbolic, approximate calculation can be successfully incorporated into current math instruction.

The current literature on the relation between non-symbolic calculation and symbolic math is still at an early stage. A handful of studies have begun to investigate approximate non-symbolic arithmetic, number line, and ordinal comparison tasks; however, the range of operations that can be represented non-symbolically and approximately is unknown. It is an open question as to whether more advanced math topics are usefully presented in a non-symbolic, approximate format. There is no consensus on the math skills most malleable to non-symbolic calculation intervention, or who is the ideal target for this training. It is unclear how ANS acuity relates to a child's ability to perform non-symbolic calculation. The current dissertation offers evidence for novel non-symbolic, approximate calculation abilities, a meaningful relation between ANS acuity, non-symbolic calculation skill, and symbolic mathematics, and explores which target populations may benefit from practice with ANS calculation tasks.

Dissertation approach

The goal of this dissertation is to explore the relation between intuitive non-symbolic calculation and symbolic mathematics. To examine this link this dissertation will take a behavioral and psychophysical approach to understand how elementary school aged children and adults solve non-symbolic calculation problems, and how this skill relates to the acuity of their non-symbolic representations of quantity, and their symbolic mathematical skill. The first two chapters examine the extent to which children and adults can calculate non-symbolically, and whether this non-symbolic calculation skill functions as a link between ANS acuity and symbolic math. The third chapter examines the limits of training with non-symbolic addition and subtraction to improve the symbolic addition and subtraction ability of adults. The fourth chapter explores whether math anxiety affects the accuracy of both symbolic and non-symbolic calculation. Taken together, these studies provide insight into the relation between our intuitive and symbolic systems of mathematics and identify novel targets and populations of students for whom non-symbolic arithmetic interventions may be an effective way to improve math learning.

The first empirical chapter explores the ability of children to compare and calculate non-symbolic ratios. Ratios are notoriously challenging for children to learn, but despite this, human infants and non-human primates solve ratio comparisons non-symbolically using the ANS (Denison & Xu, 2014; McCrink & Wynn, 2007; Siegler, Fazio, Bailey, & Zhou, 2013). This disconnect between non-symbolic ratio intuition and symbolic ratio calculation offers the possibility that intuitive ratio reasoning could be used as a scaffold for symbolic ratio learning. Practice with non-symbolic, approximate ratio reasoning may strengthen a child's conceptual ratio reasoning skill, and in turn promote correct symbolic ratio reasoning strategies. To examine this hypothesis, 6-8 year old children completed a binary choice ratio comparison task in two formats: symbolic (numerals) and non-symbolic (dots). They also completed a measure of their ANS acuity and a standardized symbolic math assessment of general math skills and probability concepts. This chapter explores whether children can accurately compare non-symbolic and symbolic ratios before formal mathematical instruction in ratio calculation, even when controlling for the use of incorrect unidimensional heuristics. I also examine whether completing the non-symbolic ratio comparison task first improves symbolic ratio comparison accuracy by reducing the use of incorrect heuristics during the symbolic ratio comparison task. Finally, this chapter investigates whether non-symbolic ratio comparison accuracy mediates the relation between ANS acuity and math in the domain of probabilities and general numeration skills. A complete or partial mediation is evidence that stronger non-symbolic calculation skill can function as a mechanism of the relation between ANS acuity and symbolic math ability.

The second empirical chapter explores whether children and adults can use their approximate number system to compute a division operation. Previous work has shown that children can perform a scaling operation that increases or decreases dot arrays by a factor of 2 or 4 (McCrink et al., 2016; McCrink & Spelke, 2010, 2016). Scaling by a factor, however, is distinct from true division where both the initial set and the divisor can vary. To examine whether children have a true sense of intuitive division, this chapter describes performance on a novel approximate division task in both 6 to 9 year-old children and college undergraduates. In this division task subjects were presented with non-symbolic (dot arrays) or symbolic (Arabic numerals) dividends ranging from 32 to 185, and non-symbolic divisors ranging from 2 to 8. Subjects compared their imagined quotient to a comparison quantity. This chapter explores the degree to which children and adults can successfully divide in both non-symbolic and symbolic format, and whether subjects can generalize to divide with novel divisors. I also examine whether the ability to perform intuitive division is dependent on formal knowledge of the division operation. Finally, this chapter also tests whether non-symbolic division ability mediates the relation between ANS acuity and math in both children and adults. In the same way as Chapter 1, a significant mediation effect is evidence that non-symbolic calculation may be a mechanism of the relation between ANS acuity and symbolic math.

The third chapter explores whether non-symbolic, approximate addition and subtraction training can benefit the symbolic arithmetic skills of college undergraduates across four experiments. Previous research demonstrated that non-symbolic, approximate addition and subtraction training could improve symbolic arithmetic fluency in adults (Au, Jaeggi, & Buschkuehl, 2018; Park & Brannon, 2013; Park & Brannon, 2014). In this paradigm, participants who trained for 6 or 10 days with approximate arithmetic answered more symbolic addition and subtraction problems correctly at post-test compared to pretest than participants who trained with numeral ordering, knowledge training, dot comparison, or visual spatial short-term memory tasks. To further investigate the extent of training necessary, the mechanism of this transfer, and the robustness of the effect Chapter 3 reports four attempted extensions and replications of the original effect using the exact same training and testing paradigms. Overall, this chapter explores whether non-symbolic, approximate arithmetic training can benefit the basic arithmetic skills of adult subjects already expert in arithmetic calculation.

The fourth and final chapter explores how non-symbolic and symbolic calculation relate to math anxiety. Math anxiety is a negative reaction specific to situations involving numbers or mathematics (Beilock & Maloney, 2015; Richardson & Suinn, 1972). Students high in math anxiety tend to perform poorly on tests of mathematics (Ashcraft & Krause, 2007; Hembree, 1990; Lee, 2009; Ramirez, Gunderson, Levine, & Beilock, 2013). The cause of math anxiety and its relation with math performance is complex and bidirectional. Extrinsic factors such as social learning and aversive conditioning may lead to math-focused anxiety which functions to deplete domain general cognitive skills necessary for mathematics (Ashcraft & Kirk, 2001; Ramirez, Chang, Maloney, Levine, & Beilock, 2016; Ramirez et al., 2013). In other cases a fundamentally poor representation of numerical magnitude may lead to math difficulties and math focused anxiety (Dietrich, Huber, Moeller, & Klein, 2015; Maloney, Ansari, & Fugelsang, 2011; Núñez-Peña & Suárez-Pellicioni, 2014). Chapter 4 examines whether the negative relation between math anxiety and math performance extends to math calculation done without numerals. Non-symbolic calculation requires participants to use their sense of numerical magnitude to calculate. If the relation between math anxiety and math performance is due to weaker numerical magnitude representation among math anxious students, there will be a correlation between math anxiety and non-symbolic calculation accuracy. Alternatively, if math symbols (numerals) provoke math anxiety and the corresponding depletion of domain general skills, there should be no correlation between math anxiety and non-symbolic calculation. The same participants from Chapters 1 and 2 and additional college undergraduates completed measures of math and reading anxiety, symbolic math, non-symbolic math, and Approximate Number System (ANS) acuity. I then test whether math anxiety level is correlated with calculation accuracy in both non-symbolic and symbolic format. If non-symbolic calculation accuracy is relatively unrelated to math anxiety level, non-symbolic calculation may be an effective way to introduce mathematical concepts to math anxious students.

CHAPTER 1: FIRST AND SECOND GRADERS SUCCESSFULLY REASON ABOUT RATIOS WITH BOTH DOT ARRAYS AND ARABIC NUMERALS

Introduction

Ratio reasoning is prevalent in everyday life. We reason about ratios when we estimate how long it will take to walk to the grocery store, decide how many hands we need to carry all the grocery bags into the house, or when we slice a birthday cake to ensure that everyone at the party will get a piece. Mathematics provides a symbolic system to calculate exact solutions to these ratio reasoning problems, however, in everyday life we typically avoid such precise calculations and instead arrive at an approximate, workable solution. These approximate ratio reasoning skills are present from infancy and are shared with other animal species. Comparative and developmental research demonstrates that animals and human babies possess an intuitive capacity for ratio reasoning that is not dependent on language or knowledge of symbolic mathematics. For example, when an ape is given a choice between two human hands that each randomly drew one food item from one of two buckets, the ape will tend to pick the hand that drew from the bucket that contained the more favorable ratio of banana pellets to carrots (Eckert, Call, Hermes, Herrmann, & Rakoczy, 2018; Rakoczy et al., 2014). Similarly, rhesus macaques can be trained to pick the array with the more favorable ratio of a rewarded shape, or to match different colored line lengths based on ratio, for a juice reward (Drucker et al., 2015; Vallentin & Nieder, 2008). When 6-month-old human infants are habituated to a particular ratio of yellow to blue objects within an array, they look longer at a novel compared to a familiar ratio (McCrink & Wynn, 2007). Ten to 12-month-old infants will reliably crawl toward a cup that contains a lollipop drawn from a bucket with a more favorable ratio of their preferred color lollipop (Denison & Xu, 2014). Relatedly, it violates infants' expectations when a sample is drawn from a population of items that does not reflect the population distribution (Kayhan, Gredebäck, & Lindskog, 2017; Xu & Garcia, 2008). For example, when 8-month-old infants see a box filled with ping-pong balls where the majority is red and only a few are white, they look longer when 4 out of the 5 balls randomly drawn from this box are white (Xu & Garcia, 2008). Thus, infants and non-human primates are equipped with intuitive ratio reasoning skills that are flexible to probabilistic context and the depiction of the ratio. An important question is whether this powerful ratio reasoning mechanism can be harnessed to improve symbolic ratio reasoning ability in school-age children.

Children reason approximately about ratios before they are capable of calculating exact ratios, and this intuitive ratio reasoning skill continues to grow during early education. Six year old children can solve proportional reasoning problems displayed with continuous magnitudes (e.g. line lengths), but continue to have difficulty with the same ratio problems presented with discrete magnitudes (e.g. squares) until around age 10 (Boyer & Levine, 2015; Boyer, Levine, & Huttenlocher, 2008). Seven and ten year old children can compare ratios of items when presented as a frequency (Y out of X tokens are gold), but only 10 year old children succeed when the ratio between the two proportions is small (Ruggeri et al., 2018). Elementary and preschool aged children (6 to 11 years old) can successfully pick which of two jars has the more favorable ratio of their preferred color object, but the strategies children use to solve this task become more sophisticated with development (Falk et al., 2012; Yost, Siegel, & Andrews, 1962). For example, four year-old children tend to use a unidimensional strategy based on the relative number of the preferred (or non-preferred) color in each set and by around age 7 children tend to use a true ratio strategy that compares the relation between relations of each set (Falk et al., 2012). In sum, children's intuitive grasp of ratio improves around the same period of time children begin intensive formal math education. This timing suggests a link between symbolic and non-symbolic ratio representations. But in contrast to children's prodigious intuitive ratio reasoning skill, exact, symbolic ratio reasoning is very difficult for children to acquire.

According to the US Common Core standards, children begin formal education about ratio in 3rd grade, but by 4th grade only 32% of students are able to identify whether simple fractions are greater or less than ½, and only 27% of 8th graders can successfully place a rational number in the correct spot on a number line (National Assessment of Educational Progress

2017). Ratios, fractions, and proportions are critically important for advanced math education, thus deficits in symbolic ratio reasoning are a major challenge for math educators (Butterworth, Varma, & Laurillard, 2011; Duncan et al., 2007; Parsons & Bynner, 2005; Siegler et al., 2012). One possible way to improve children's symbolic ratio reasoning is to use a child's intuitive understanding of ratio to scaffold their symbolic learning (Ahl, Moore, & Dixon, 1992; Falk & Wilkening, 1998; Fujimura, 2001). Prior research has established that children are able to approximately calculate over non-symbolic dot arrays by adding, subtracting, multiplying, dividing, solving for unknown values in algebra problems, and placing dot arrays on a number line (Barth et al., 2006; Honoré & Noël, 2016; Kibbe & Feigenson, 2017; McCrink et al., 2016; Park et al., 2016). Moreover, training with some non-symbolic and approximate math operations yields improvement in symbolic math ability. For example, practice adding and subtracting arrays of objects improves the general math skill of preschoolers after 10 days of training (Park et al., 2016; Szkudlarek & Brannon, 2018). Similarly, practice with a symbolic and non-symbolic number line task improves number line and arithmetic performance in 8-10 year old children (Kucian et al., 2011). It is therefore possible that practice with non-symbolic ratio reasoning could benefit symbolic ratio reasoning in a similar way.

The first goal of the current study is to examine the non-symbolic ratio comparison skills of children in early elementary school, and to test whether children can extend their intuitive ratio reasoning ability to a ratio comparison task in a symbolic format. To this end, we created both a non-symbolic and symbolic version of the same approximate ratio reasoning task. On each trial children were presented with an illustration of two gumball machines both filled with blue and white colored gumballs (dots) or blue and white numerals (Figure 1). The task was to pick the machine with the best chance of producing a blue (or white, counterbalanced) gumball. Crucially, the task did not require exact calculation of the ratio of ratios, but instead only required children to identify which ratio of blue to white was more favorable. Due to the approximate nature of the task, we predicted that children would perform above chance despite their lack of formal knowledge of ratio or probabilistic reasoning with both formats of the task (e.g., see Gilmore, McCarthy, & Spelke, 2007 for a similar finding with approximate addition and subtraction). We further hypothesized that if children were able to connect their intuitive, non-symbolic ratio reasoning skills to the symbolic version of the calculation, engaging in non-symbolic ratio reasoning should provide a foothold for children as they attempt to make the isomorphic judgment with symbols. To test this, each child completed both the symbolic and non-symbolic ratio comparison tasks with order counterbalanced across children. If children's symbolic ratio reasoning can be scaffolded by engaging in the non-symbolic ratio reasoning task, then performance on the symbolic task should be modulated by the order of the two tasks.

The second goal of the current study was to examine the degree to which children made true ratio comparisons or instead relied on simpler unidimensional heuristics. Prior work found that when children 7 years of age or younger were presented with two ratios they tend to erroneously compare only the number of preferred items in an integer-based strategy (Clarke & Roche, 2009; Falk et al., 2012; Jeong, Levine, & Huttenlocher, 2007; Obersteiner, Bernhard, & Reiss, 2015; Shaklee & Paszek, 1985; Siegler, Strauss, & Levin, 1981). Slightly older children tend to choose the stimulus with fewer non-preferred items, and/or the array with the overall greater number of items. As children progress in their mathematical knowledge, they are more likely to perform a true ratio comparison and consider both the preferred and non-preferred items in each array. However, previous research on ratio comparison strategies has relied on self-report or differences in accuracy between trial types to test for the use of a particular heuristic. These approaches have serious drawbacks. First, self-report may be inaccurate, especially among young children who have not yet learned formal ratio reasoning in school. Second, a comparison of accuracy by trial type does not account for the fact that many binary choice ratio comparisons can be solved using more than one strategy. For example, imagine two gumball machines where the left machine has 25 blue gumballs and 15 white, and the right machine has 12 blue gumballs and 18 white. A child needs to pick the machine with the best chance of getting a blue gumball on their first try. The correct answer based on a comparison of ratios is the left machine. But a child could also pick the left machine because it has a larger total number of gumballs, or because it has less white gumballs. Greater accuracy on trials of this type does not disambiguate between the use of these three strategies. A more flexible analysis of strategy is needed to describe the potential use of multiple strategies within one subject. Here, we constructed the stimuli such that the correct answer was orthogonal to the total number of items in a gumball machine, and the correlation between the correct choice and the one with more of the preferred or non-preferred colored items was minimized. We then modeled the left/right choices a child would make under each of three heuristics children are known to use on a binary choice ratio comparison task. This stimulus structure allowed us to detect the degree to which children used unidimensional heuristics and/or a ratio comparison to solve the ratio tasks by looking at a child's pattern of behavior across all trials instead of only a subset split by trial type. Our strategy analysis had three goals. The first was to determine whether children perform a true ratio comparison independent of incorrect heuristics in both symbolic and non-symbolic format. The second was to determine whether there is evidence of incorrect heuristic use, again in both task formats. The third was to assess the degree to which task order (non-symbolic or symbolic first) affects not only overall accuracy, but the strategies children employ to complete the task. We hypothesized that engaging in non-symbolic ratio comparison would shift the strategies children use to solve the symbolic ratio task and lead to improved symbolic ratio reasoning. More specifically, we expected that children might rely less on heuristics in the symbolic version of the task if they first engaged in the non-symbolic task.

The third goal of the current study was to determine whether the Approximate Number System (ANS) is a cognitive foundation underlying intuitive ratio reasoning. The ANS supports the ability to represent numerical magnitude without symbols, and is present in infants, adults, and many animal species (Feigenson et al., 2004). Previous work suggests that the ANS is involved in binary choice ratio comparison tasks, because performance on these tasks is dependent on the ratio of ratios being compared, as predicted by *Weber's* law (Drucker et al., 2015; Eckert et al., 2018; McCrink & Wynn, 2007). Ratio dependent numerical discrimination following Weber's law is a hallmark of the ANS (Feigenson et al., 2004). Moving beyond the description of a ratio effect, in the current study we independently measured each child's ANS acuity to directly test whether children with sharper ANS acuity are better at intuitive ratio reasoning.

We also examined the relationship between ratio reasoning skill, ANS acuity, and formal math ability. A multitude of studies have linked individual differences in the acuity of ANS representations to a variety of symbolic math skills, but the mechanism for this relation is unknown (Halberda, Mazzocco, & Feigenson, 2008; see Schneider et al., 2016 for meta-analysis). Recent work suggests that non-symbolic calculation may be a better predictor of symbolic math ability than ANS acuity (Matthews et al., 2016; Pinheiro-Chagas et al., 2014; Starr et al., 2016). Indeed, non-symbolic ratio comparison accuracy was a stronger predictor than ANS acuity of a variety of symbolic math tasks in adults, leading to the proposal of a Ratio Processing System separate from the ANS (Matthews et al., 2016). Here we test the possibility that non-symbolic calculation is a mechanism of the relation between ANS acuity and symbolic math. Specifically, we hypothesize that sharper ANS acuity allows for better non-symbolic ratio calculation. In turn, better non-symbolic ratio calculation allows children to conceptually ground symbolic ratio representations in their intuitive non-symbolic sense of ratio, leading to better symbolic calculation skill. To test this, we asked whether non-symbolic ratio skill mediates the relation between ANS acuity and symbolic math skill as measured by our symbolic ratio comparison test and two subtests of the Key-Math-3 standardized test. The Key-Math-3 test is divided into subtests that represent different symbolic math skills (numeration, algebra, geometry, measurement, data analysis and probability). Based on our hypothesis that non-symbolic ratio calculation leads to a conceptual grounding of ratio computation, we administered two subtests with the goal of targeting both ratio based and non-ratio based symbolic math concepts. The Data Analysis and Probability section includes questions about probability and graphical representations of data. These math concepts overlap with the concepts involved in a non-symbolic probabilistic ratio comparison judgement. We also administered the Numeration section, which is a test of general counting and calculation skill. These general math skills do not involve the concept of ratio. Thus, we hypothesize that non-symbolic ratio skill will mediate the relation between ANS acuity and questions about Data Analysis and Probability but will not mediate the relation between ANS acuity and general math skill.

To summarize, the present experiment was designed with three major goals. First, to determine whether children's intuitive ratio reasoning skill functions to scaffold symbolic ratio reasoning. Second, to examine the degree to which children utilize true ratio reasoning or instead rely on incorrect unidimensional heuristics to solve a binary choice ratio comparison task. Third, to investigate the relationship between ANS acuity, intuitive ratio reasoning, and symbolic math scores. Taken together, these questions have important implications for the value of including non-symbolic ratio calculation in early math education.

Materials and Methods

Subjects

Eighty-five 6-8-year old children were tested (mean age = 7.6 years old, standard deviation = .83 years; 35 female, 47 male, 3 chose not to report). Written parental consent and children's verbal assent were obtained in accordance with a protocol accepted by the University of Pennsylvania's Institutional Review Board. Ten additional children were consented but were excluded from the final sample because they did not complete both the non-symbolic and symbolic ratio comparison tasks. The parents of 43 children in our sample elected to fill out a detailed demographics questionnaire. Of this subset of the sample, 37% identified as Hispanic or Latino, 49% identified as African American, 21% as Caucasian, 5% as Asian, 12% as more than one race, and 13% chose not to report. Our sample included a large proportion of children from families with household incomes of 50,000 or less (12% $150,000+, 7% $100,000 - $150,000, 2% $75,000 - $100,000, 5% $50,000 - $75,000, 42% $25,000 - $50,000, 23% $0 - $25,0000, and 9% chose not to report). All subjects were recruited from six after school programs in the Philadelphia, PA area. A subset of the children who completed both the non-symbolic and symbolic ratio comparison tasks completed additional assessments (Dot comparison, n = 75; Key-Math Numeration assessment, n = 78; the Key-Math Data Analysis and Probability assessment, n = 73; the Woodcock-Johnson Basic Reading Skills cluster, n = 72; and a measure of numeral identification, n = 82). All participants received a small toy as a thank you gift after completion of the experiment.

Procedure

Children completed all tasks individually with an experimenter in a quiet room at their after school program. Children completed the non-symbolic and symbolic ratio comparison tasks first and the order of the tasks was counterbalanced across children. All but 9 participants received the two ratio comparison tasks on the same day. The order in which all other tasks were administered was random across participants and was dependent on the child's availability. Each participant was tested for a total of 45-60 minutes across 2-3 days. Children also completed a math anxiety questionnaire, but these results are not included in this paper. Children received stickers throughout the session to maintain motivation.

Experimental Tasks

Both the symbolic and non-symbolic tasks were run in MATLAB and programmed using the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007). The programs were run on a 15-inch touch screen laptop computer.

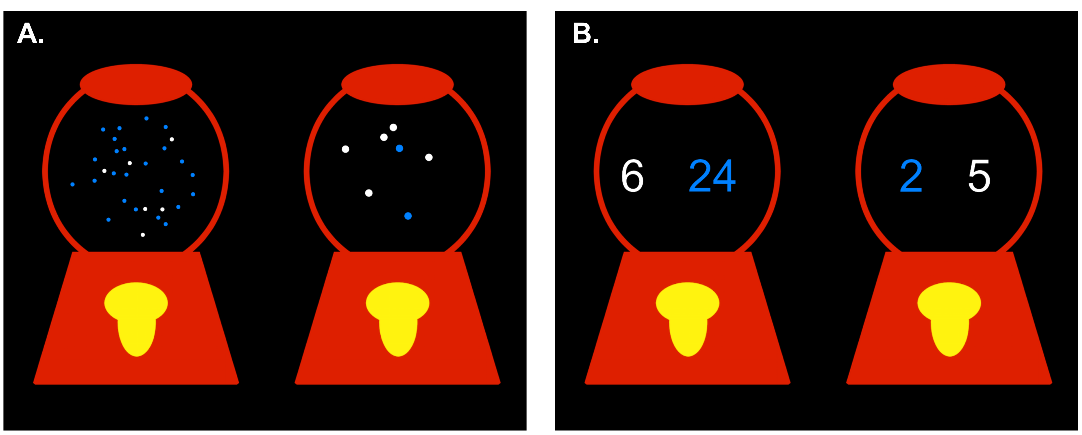

Figure 1. Illustration of one trial of the A) non-symbolic ratio comparison task and B) symbolic ratio comparison task.

Introduction to the Ratio Comparison Tasks: All children regardless of whether they were given the symbolic or non-symbolic version of the ratio task first were introduced to an alien character next to a single red and yellow gumball machine that held 11 orange colored gumballs (dots). Children were shown that only one gumball comes out of the machine at a time and that the alien doesn't care if he gets a big or small gumball[1]. The child was also told that all of the gumballs regardless of their placement in the machine have an equal chance of coming out.

Non-symbolic Ratio Comparison Task

Children were given 5 non-symbolic practice trials. On the first trial children were presented with two gumball machines where one machine contained 10 green gumballs and the other contained 10 orange gumballs. The child was told that the alien's favorite color gumball is orange. As the experimenter pointed to each machine the child was told "There are this many green gumballs in this machine and this many orange gumballs in this machine." The child was then asked "Which machine gives Mr. Alien the best chance of getting an orange gumball on his first try?" The child was asked to touch the gumball machine of their choice. Children were given verbal feedback by the experimenter. The remaining four practice trials all had mixed green and orange gumballs in each machine. In one trial the correct choice had more orange gumballs, in one trial the correct choice had more orange gumballs and a greater number of total gumballs, in one trial the correct answer had fewer total items, and in the last trial the correct answer had fewer orange gumballs. Regardless of accuracy on this last trial, the child was told "See, the machine with the *most* orange gumballs is not always the one with the *best chance* of an orange gumball." On all practice trials the experimenter did not proceed until the child picked the correct machine. After the practice trials children were told that the alien wants a blue or white gumball, counterbalanced across participants. Children completed 60 trials. After a correct response, children saw three smiling aliens, a spaceship blasting off, and the word "Great!". After an incorrect response, children saw a spaceship crashing into a planet, and the words "Let's try again!". The only

[1] Results suggest children understood this instruction. There was no significant difference in accuracy dependent on whether or not the correct ratio had the larger or smaller gumballs (t_{81} = .94, p = .35).

instructions a child was given throughout testing were “Which machine has the best chance of giving a (blue/white) color gumball?”. Children were instructed not to count the dots if they appeared to be doing so, and to respond as quickly as possible. The dots moved altogether in a circular motion on the screen for 300ms at the beginning of each trial to encourage interest. There were 30 unique ratio comparisons, and each ratio comparison occurred twice. The number of blue or white gumballs within one machine ranged from 1 – 30. The ratios of ratios (ratio of preferred to non-preferred in machine 1 / ratio of preferred to non-preferred in machine 2) ranged from 1.5 to 10. The correct answer was orthogonal to the total number of items in a gumball machine (no correlation), and there was a small correlation between the correct choice and the one with the greater number of the preferred color (r =.2) and between the correct choice and the one with less of the non-preferred items (r =.2). The stimuli are provided in Appendix A2.

Symbolic Ratio Comparison Task

The procedure and numerical values were identical to that described for the non-symbolic task. In the symbolic version of the task, the number of gumballs in each gumball machine was represented by blue and white Arabic numerals 1 to 30, instead of dots (see Figure 1). The introduction to Mr. Alien and how the gumball machine works were shown with gumballs (dots) in the machine to exactly match the non-symbolic version of the task. The five practice trials used green and orange numerals, and as in the non-symbolic version, the experimenter did not use number words when pointing to the contents of the gumball machines.

Dot Comparison Task

Two dots arrays appeared on a black screen for 750ms. The dot arrays were subsequently occluded and the child was required to touch the array with the greater quantity. Children completed 200 trials of this task, with feedback on every trial. The number of dots ranged from 8 to 32. The stimuli were created to evenly sample a stimulus space that varied by the ratio between the number, size, and the spacing of the dots. To encourage greater reliability of the measurement, trial level difficulty was titrated (Lindskog, Winman, Juslin, & Poom, 2013). The titration procedure calculated the percentage correct over the last 5 trials. The ratio between the two dot arrays moved one log level farther apart if the accuracy was less than 70% and moved one log level closer together if the accuracy was greater than 80%. A quantitative index of each child's ANS acuity was calculated with the Weber fraction (*w*) as specified in (DeWind et al., 2015). This model accounts for the effects of non-numerical features of dot arrays on numerical discrimination (DeWind et al., 2015).

Numeral Identification Task

Each child's numeral recognition ability was assessed by presenting the numerals 1-30 individually on index cards. The numerals were displayed in random order, and the child was asked "What number is this?" The accuracy of each child's response was recorded.

Key Math-3 Diagnostic Assessment

The Numeration and Data Analysis and Probability sections of the Key Math-3 Diagnostic Assessment Form B (Connolly, 2007) were administered. The Numeration section is a test of general basic math skills such as place value, counting, the relative magnitude of numbers, and an understanding of fractions, decimals, and percentages. For example, children are presented with 4 numerals and told "Read the numbers in order, from least to greatest". The Data Analysis and Probability section targets math content related to concepts of probability, statistics, and graphical representations of data. For example, this section includes questions such as "Which spinner gives you an equal chance of landing on green or white?". We used the age standardized scale score for both sections of the test.

Woodcock-Johnson IV Test of Cognitive Abilities

Participants' reading abilities were assessed using the "Basic Reading Skills" cluster of the Woodcock-Johnson. This cluster is comprised of the "Letter-Word Identification" and "Word Attack" subtests. In the "Letter-Word Identification" subtest, participants named letters and read words aloud. In "Word Attack," participants read nonsense words and identified letter sounds. We used the age standardized Basic Reading Skills score.

Analysis Plan

Strategy Analysis. We examined three incorrect heuristics: picking the machine with a greater absolute number of the preferred color gumballs, picking the machine with a greater number of total items, and picking the machine with less of the non-preferred color gumball. We will refer to these strategies as the “More Good”, “More Items”, and “Less Bad” strategies respectively. Three subjects were excluded from these analyses because they did not complete the full 60 trials of one of the two ratio tasks. The first goal of this analysis was to test whether children were performing a ratio comparison, or simply relying on an alternative heuristic. To test this, we created an ideal model for each heuristic by coding the choices a child would make on each trial if they were exclusively using a given heuristic (1 indicating the choice of the right machine, and 0 indicating a choice of the left machine). Then, we subtracted each ideal heuristic model from the correct response on each trial and took the absolute value of this subtraction. This yielded three error models: a model of the errors that would be made if children used a More Good strategy, a model of the errors that would be made if children used a More Items strategy, and a model of the errors that would be made if children used a Less Bad strategy. These error models indicate the trials where each incorrect heuristic would differ from a correct ratio comparison response (0 indicating the heuristic and a ratio response give the same answer, 1 indicating they give different answers). Trials indicated by a 1 in each error model cannot be correctly solved by using each particular incorrect heuristic. We then subtracted each ideal heuristic model from children’s actual responses on the task to remove any trials where children could have used an incorrect heuristic to get the correct answer and took the absolute value. This created a vector where a 0 indicates a response corresponding to an incorrect heuristic, and a 1 indicates a response opposite of the incorrect heuristic. Then, we calculated a Pearson correlation coefficient between this vector and each error model. A positive correlation indicates choices based on a true ratio comparison when controlling for use of the incorrect heuristic. The outcome of this analysis is six correlation coefficients for each subject, one for each heuristic on both the non-symbolic and symbolic ratio comparison tasks. We then tested at a group level whether children’s correlation coefficients differed from zero with a one sample t-test for each heuristic on both the non-symbolic and symbolic ratio tasks.

The second goal of this analysis was to characterize which, if any, incorrect heuristics children used. To do this, we took the error models we created for the first part of the analysis and correlated each error model with a child's actual errors on the task. A significant positive correlation indicates use of the heuristic. This analysis tests if the errors children made on the ratio comparison tasks were random or could be described by a bias towards using a specific heuristic. This analysis also yields six correlation coefficients for each subject, one for each heuristic for both task formats. Again, we use a one sample t-test to test whether the correlation differs from zero at a group level.

Mediation Analysis. We first removed any outliers on each task that were greater or less than three times the interquartile range. This process removed two ANS acuity scores. ANS acuity and symbolic ratio comparison scores were log transformed to approach a normal distribution (ANS acuity Shapiro-Wilk W = .96; symbolic ratio comparison W = .98). To ensure that correlations between measures were not simply due to age or demographics of each school location, we partialed out age and school location from our measures of ANS acuity and symbolic and non-symbolic ratio comparison, and school location from the age standardized scores of the Woodcock-Johnson and the Key Math. The age and school location standardized Pearson correlations are reported in Table 1.

Children completed three symbolic math measures: symbolic ratio comparison, Key-Math-3 Numeration, and Key-Math-3 Data Analysis and Probability. ANS acuity was no longer correlated with accuracy on the symbolic ratio comparison task when controlling for age and school location (comparison of Appendix A.1 and Table 1). Consequently, we ran mediation models with only the Key-Math 3 Numeration and Data Analysis and Probability sections as symbolic math outcome measures[2]. As seen in Table 1, the Key-Math-3 Numeration section score was significantly correlated with children's score on the Woodcock Johnson Basic Reading Skills. To ensure that this outcome measure is a test of general math skill, and not general academic performance, we partialed out children's Woodcock Johnson Basic Reading score from the Key-Math Numeration score. The residuals from the models with age, school location, and the Basic Reading Skills test as predictors were used in the mediation analysis.

Finally, we ran two mediation models: one for each symbolic math measure. Mediation analyses test for a significant indirect effect (the product of the standardized coefficients a and b) that accounts for some portion of the original direct effect (c). The remaining direct effect is represented as c'. First, we ran a model testing whether non-symbolic ratio comparison accuracy was a mediator of the relation between ANS acuity and children's scores on the Key-Math-3 Numeration test. Second, we ran the same model, but with the Key-Math-3 Data Analysis & Probability section as the symbolic math outcome measure. The goal of this analysis is not to determine causality, but rather to examine whether non-symbolic ratio calculation and ANS acuity account for the same or different variance in each symbolic math outcome measure. A full or partial mediation, indicated by a significant indirect effect, would be consistent with our hypothesis that non-symbolic ratio calculation is a mechanism of the relation between ANS acuity and symbolic math.

Results

Symbolic and non-symbolic ratio comparison performance. Children performed both the non-symbolic (Figure 2A; $t_{84} = 14.15$ $p < .001$, Cohen's $d = 1.54$) and symbolic (Figure 2A; $t_{84} = 8.22$ $p < .001$, Cohen's $d = .89$) ratio tasks with above chance accuracy. Children were more accurate on non-symbolic ratio comparison than symbolic ratio comparison ($t_{84} = 7.20$ $p <.001$, Cohen's $d = .78$). This format effect remained with the subset of children (N=64) who recognized all numerals 1-30 on the numeral identification test, indicating that numeral recognition ability was

[2] We ran all mediation analyses using the "mediation" package in R (Tingley et al. 2014).

not driving this difference (t_{63} = 6.99, p <.001, Cohen's d = .87). As shown in Figure 2b, accuracy on the non-symbolic and symbolic ratio comparison tasks were highly correlated (Appendix A.1; r = .62, t_{83} = 7.22 p <.001). This relation held when controlling for the age and school location of the participants (Table 1; r = .52, t_{83} = 5.61, p < .001).

To test whether the ratio of ratios of the two arrays impacted accuracy, we ran a generalized mixed effects linear model (GLMM) following a binomial distribution predicting accuracy as a function of task format (symbolic or non-symbolic), the ratio of ratios as fixed effects, and a random effect of subject. This model indicated significant main effects of ratio of ratio and task (ratio of ratio β = .11, z = 9.46, p < .001; task β = -.16, z = -2.21, p = .03), and an interaction between ratio of ratio and task format (β = -.06, z = -3.90, p < .001). The interaction indicates that accuracy was less dependent on the ratio of ratios of the two arrays for the symbolic than for the non-symbolic task, however, the main effect indicates that the ratio of ratios of the stimuli impacted accuracy for both tasks.

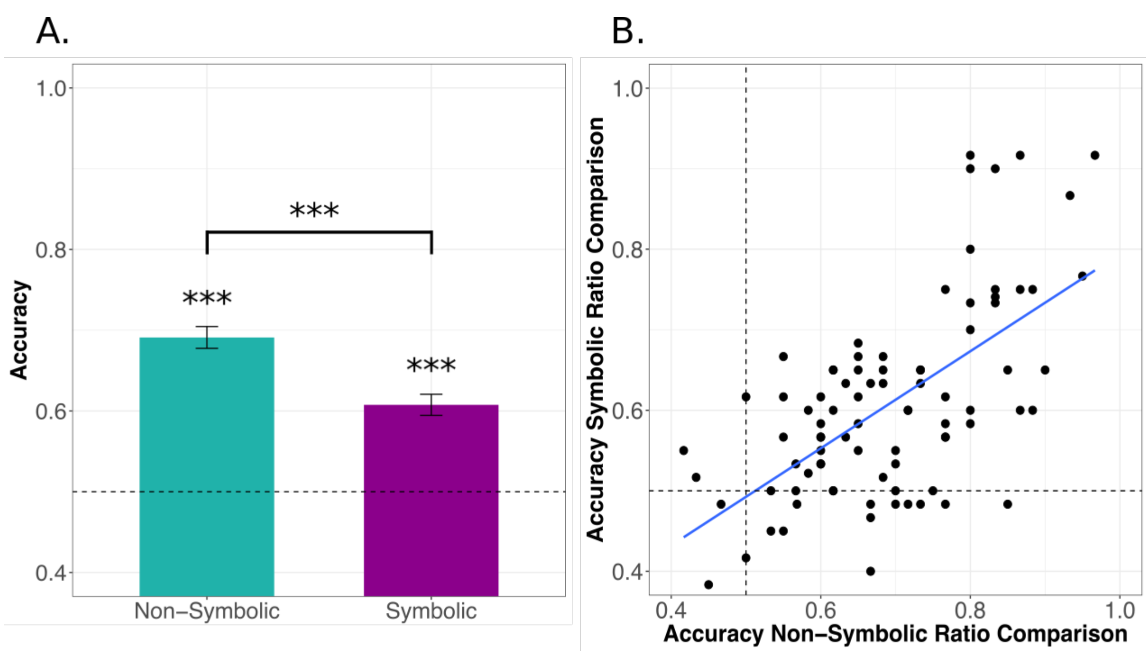

Figure 2. A) Accuracy was above chance on both the non-symbolic and symbolic ratio comparison tasks. Children were more accurate during non-symbolic than symbolic ratio comparison. Error bars indicate standard error of the mean. B) Performance on the symbolic and non-symbolic ratio comparison tasks were highly correlated. Dotted lines indicate chance performance.

Non-symbolic ratio comparison facilitates symbolic ratio comparison. To test the hypothesis that experience with non-symbolic ratio comparison scaffolds understanding of the symbolic version of the task, we ran a mixed effects ANOVA with a main effect of task order (first or second), task format (symbolic or non-symbolic), a task order by task format interaction, and a random effect of subject. As planned, we removed any children who could not identify all numerals 1-30 from these analyses. This left 64 children spanning the age range of the entire sample (11 six year olds, 21 seven year olds, 32 eight year olds). In line with our prediction, there was a significant task order by task format interaction (Figure 3; $F_{1,62} = 4.56$, $p = .04$). This effect was driven by significantly better performance on the symbolic ratio comparison test after completion of the non-symbolic version of the task ($t_{62} = -2.21$, $p = .03$, Cohen's $d = .56$) and a non-significant effect of task order on non-symbolic ratio comparison accuracy ($t_{62} = -1.63$, $p = .11$). The ANOVA also revealed a significant main effect of task ($F_{1,62} = 48.6$, $p < .001$), consistent with our analysis of the larger data set, and a non-significant effect of task order ($F_{1,62} = .543$, $p = .46$). These findings are consistent with our hypothesis that non-symbolic ratio reasoning serves as a scaffold for symbolic ratio reasoning.

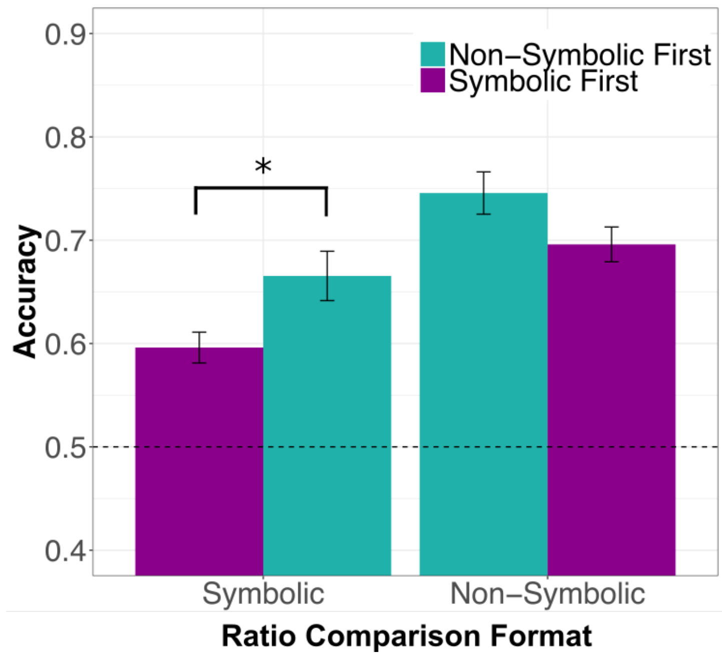

Figure 3. There is a significant interaction between task order and task format. Children who completed the non-symbolic ratio comparison task first were significantly more accurate on the symbolic ratio comparison task. Error bars indicate standard error of the mean

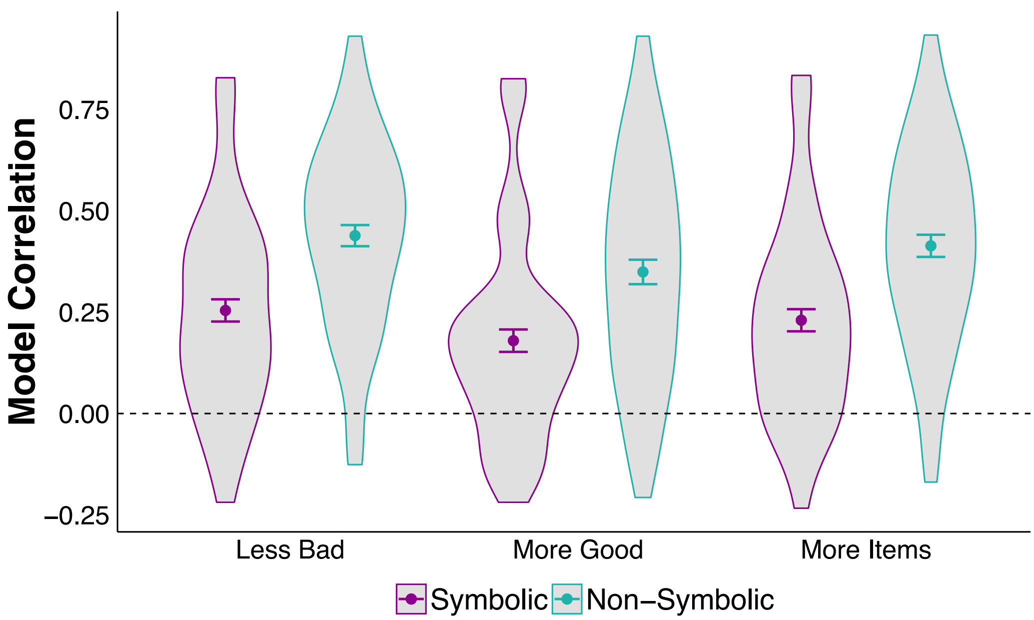

Figure 4. Correlation between performing a true ratio comparison and children's actual choices on trials where a ratio comparison and each incorrect heuristic provided opposing answers. A correlation significantly above zero indicates children made choices consistent with ratio comparison, even when controlling for the use of each incorrect heuristic. Children's responses

were significantly positively correlated with the responses indicated by a ratio comparison strategy for all three heuristics on both formats of the ratio comparison task $p < .001$. Points indicate the mean correlation of all subjects, and the grey area of the violin plot indicates the distribution of individual correlations. Error bars indicate standard error of the mean.

Alternative strategy analysis. We examined whether children performed a ratio comparison when controlling for heuristic use on both the symbolic and non-symbolic ratio tasks. Figure 4 displays the correlation between each error model and a model of where each child's actual response differed from the incorrect heuristic. A one-sample t-test revealed a significant positive correlation between each error model and the model of children's choices controlling for the corresponding heuristic (Figure 4; Non-symbolic More Good $r = .35$, $t_{81} = 11.6$, $p < .001$; More Items $r = .41$, $t_{81} = 15.1$, $p <.001$; Less Bad $r = .44$, $t_{81} = 16.8$, $p <.001$; Symbolic More Good $r = .18$, $t_{81} = 6.49$, $p < .001$; More Items $r =.23$, $t_{81} = 8.42$, $p < .001$; Less Bad $r = .25$, $t_{81} = 9.28$, $p <.001$). This analysis revealed that children's choices were significantly driven by a comparison of ratios in both the symbolic and non-symbolic ratio comparison tasks even when controlling for the potential use of each incorrect heuristic.

Although the previous analysis demonstrates that children were not solely relying on any particular unidimensional heuristic to solve the ratio comparison tasks, it remains possible that children used one or more heuristic on some trials. Figure 5 displays the correlation between the error model for each strategy and a child's actual errors on the non-symbolic and symbolic ratio comparison tasks. For the non-symbolic ratio comparison task, a one sample t-test on the mean correlation indicated that the errors children made were not random, but instead were in accordance with the More Good or More Items strategies (Figure 5; More Good $r = .36$, $t_{81} = 10.7$, $p < .001$; More Items $r = .29$, $t_{81} = 9.12$, $p <.001$). The negative correlation between errors and the Less Bad strategy indicated that children were significantly doing the opposite of the Less Bad strategy ($r = -.24$, $t_{81} = -8.11$, $p <.001$). The same pattern of results was found for the symbolic ratio comparison task. Specifically, children's errors were in accord with the More Good ($r = .25$, $t_{81} = 7.83$, $p <.001$) or the More Items ($r = .20$, $t_{81} = 6.61$, $p <.001$) strategy. Children also avoided the Less Bad strategy on the symbolic version of the task ($r = -.17$, $t_{81} = -6.17$, $p <.001$). Thus, the errors children made on the ratio task were qualitatively similar regardless of presentation format.

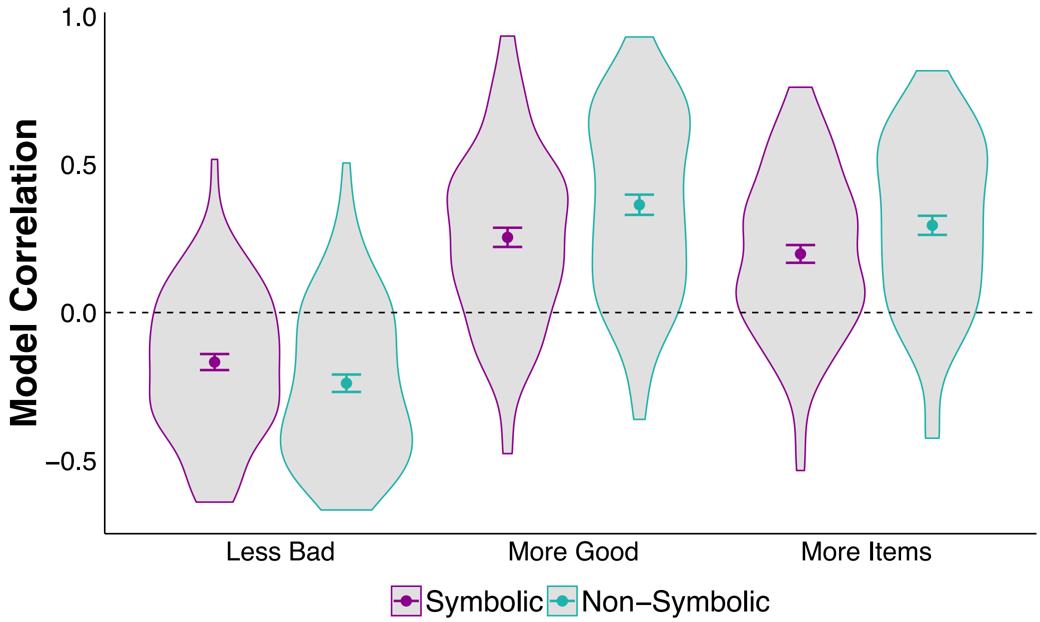

Figure 5. Correlation between the errors a child would make if they were exclusively using a given strategy and the errors a child actually made on the non-symbolic and symbolic ratio comparison tasks. A correlation significantly above zero indicates use of the strategy. The errors children made on both the symbolic and non-symbolic ratio tasks were consistent with use of the More Good or More Items strategy $p < .001$. The errors children made were significantly opposite of using the Less Bad strategy $p < .001$. Points indicate the mean correlation of all subjects, and the grey area of the violin plot indicates the distribution of individual correlations. Error bars indicate standard error of the mean.

No evidence of a strategy shift after non-symbolic ratio practice. Finally, we tested our hypothesis that engaging in non-symbolic ratio comparison might facilitate children's subsequent symbolic ratio comparison by allowing them to shift away from using incorrect heuristics. We used the same strategy analysis described above to investigate heuristic use, but this time split participants by the order in which they completed the ratio tasks (non-symbolic first $n = 27$, symbolic first $n = 36$) and analyzed their performance on only the symbolic version of the task. Contrary to our hypothesis, there was no significant difference in the degree to which children's errors reflected use of the More Good and More Items strategies by the order in which they completed the symbolic task (two-sample t test More Good $t_{61} = -.66$, $p = .51$; More Items $t_{61} = -.82$, $p = .41$). There was also no effect of task order on the degree to which children avoided the Less Bad heuristic ($t_{61} = 1.09$, $p = .28$).

Table 1. Age and school location standardized Pearson correlations between each measure.

		1	2	3	4	5
1	**ANS acuity**					
2	**Non-symbolic Ratio Comparison**	-0.24*				
3	**Symbolic Ratio Comparison**	-0.14	0.52***			
4	**Key-Math 3 Numeration**	-0.41***	0.30**	0.32**		
5	**Key-Math 3 Data Analysis & Probability**	-0.28*	0.37**	0.30**	0.64***	
6	**Woodcock-Johnson Reading Custer**	-0.18	0.18	0.30*	0.28*	0.23

*Note: *** p<.001 ** p < .01 * p < .05*

The relation between ANS acuity, non-symbolic ratio comparison, and symbolic math skills. Finally, we examined how children's ability to compare ratios non-symbolically related to their ANS acuity, and symbolic math skill. Table 1 displays the age and school location standardized correlations between non-symbolic and symbolic ratio comparison skill, ANS acuity, Key-Math-3 Numeration and Data Analysis and Probability sections, and the Reading Cluster score on the Woodcock-Johnson. The zero order correlations are reported in Appendix A1. Mean performance on the Key-Math-3 Numeration subtest was 9.05 (standard deviation 3.3), and mean performance on the Data Analysis & Probability subtest was 8.47 (standard deviation 3.1). The scale scores are constructed to have a mean of 10 (standard deviation 3).

We tested whether non-symbolic ratio comparison accuracy mediated the relation between ANS acuity and the two subtests of the Key-Math-3 test (Figure 6). We first ran the mediation analysis with the Numeration subtest as an outcome measure. ANS acuity was a significant predictor of the Key-Math 3 Numeration score (standardized $\beta = -.41$, $p < .001$) and was also a significant predictor of accuracy on the non-symbolic ratio comparison task (standardized β = -.27, p = .02). However, when ANS acuity and non-symbolic ratio comparison accuracy were both entered into the same model, non-symbolic ratio comparison accuracy was no longer a significant predictor of the Key-Math-3 Numeration score (β = .19, p = .10), while ANS acuity remained a significant predictor (β = -.36, p = .002). The direct effect was significant when tested with a bootstrap estimation approach with 5000 simulations (direct effect = -.36 95% CI = [-.54 -.18] p <.001), while the indirect effect was not significant (indirect effect = -.05 95% CI = [-.14 .001] p = .06). Thus, non-symbolic ratio comparison did not mediate the relation between ANS acuity and a child's score on the Key-Math-3 Numeration test. Instead, ANS acuity explained unique variance beyond non-symbolic ratio calculation accuracy in performance on the Key-Math-3 Numeration test.

We then ran the mediation analysis with the Data Analysis and Probability subtest as an outcome measure. ANS acuity was a significant predictor of a child's score on the Key-Math-3 Data Analysis and Probability section (standardized β = -.28, p = .03) and of accuracy on the non-symbolic ratio comparison task (standardized β = -.29[3], p = .02). In contrast to the previous analysis of the Numeration subtest, ANS acuity was no longer a significant predictor of the score on the Probability and Data Analysis subtest after controlling for the mediator, non-symbolic ratio comparison accuracy (ANS acuity standardized β = -.18, p = .13; non-symbolic ratio comparison accuracy standardized β = .32, p =.009). Non-symbolic ratio comparison fully mediates the relation between ANS acuity and performance on the Key-Math-3 Data Analysis and Probability subtest. The indirect effect was significant when tested with a bootstrap estimation approach (indirect effect = -.09, 95% CI = [-.21 -.01], p =.02). The direct effect was not significant (direct effect = -.18, 95% CI = [-.40 .02], p =.08). The proportion mediated was .34 (p = .03, 95% CI = [.04 1.17]). Thus, a higher Key-Math-3 Data Analysis and Probability score was associated with

[3] The correlation between ANS acuity and non-symbolic ratio accuracy is -.27 in the mediation model with the Numeration subtest, and this same correlation is -.29 in the mediation model with the Data Analysis & Probability subtest. This is because 5 fewer subjects completed the Data Analysis & Probability subtest.

.09 standard deviations sharper ANS acuity as mediated through non-symbolic ratio comparison accuracy.

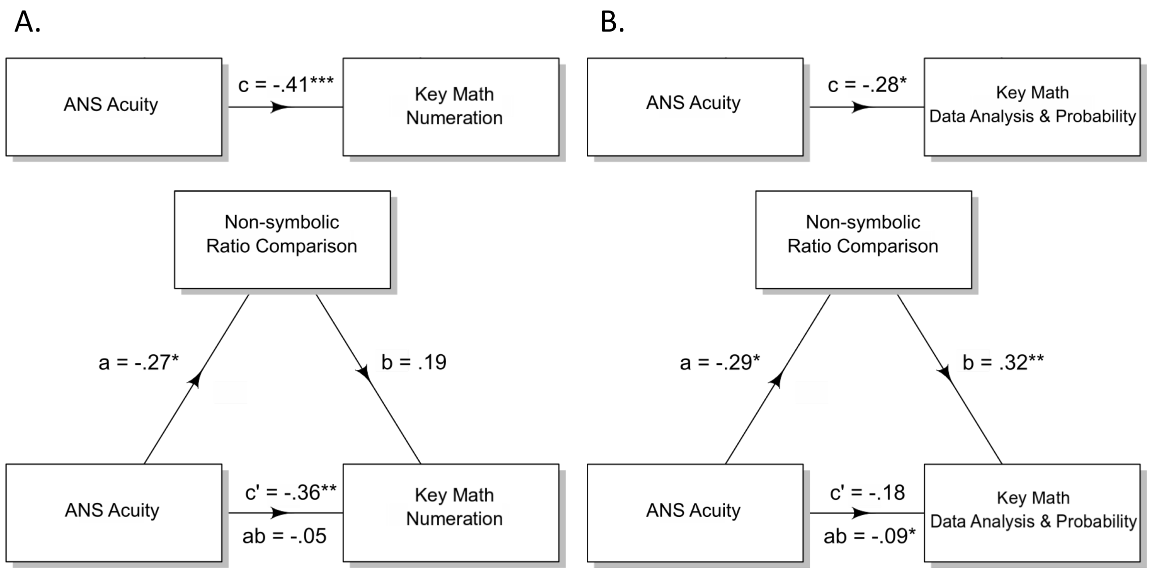

Figure 6. Mediation analyses test for a significant indirect effect (the product of the standardized coefficients a and b) that accounts for some portion of the original direct effect (c). The remaining direct effect is represented as c'. The models in this figure test whether non-symbolic ratio comparison performance mediates the relation between ANS acuity and a measure of general early math skills (Key-Math-3 Numeration) and math skills related to probability (Key-Math-3 Data Analysis & Probability). A) Non-symbolic ratio comparison accuracy does not mediate the relation between ANS acuity and a child's score on the Key-Math-3 Numeration section. The direct effect c' remains significant. B) Non-symbolic ratio comparison accuracy fully mediates the relation between ANS acuity and a child's score on the Key-Math-3 Data Analysis and Probability section. The remaining direct effect (c') is no longer significant, while the indirect effect (ab) is significant as tested with a bootstrap estimate approach

Discussion

The first goal of our study was to assess whether intuitive ratio reasoning can support both non-symbolic and symbolic ratio comparisons in elementary school children, and whether exposure to non-symbolic ratio reasoning might scaffold symbolic ratio reasoning. We found that children are adept at solving both non-symbolic and symbolic ratio comparisons before they receive formal education about fractions and ratios. Our results support previous work demonstrating non-symbolic intuitive ratio reasoning in young children (Boyer & Levine, 2015; Falk et al., 2012; Yost et al., 1962), and we for the first time show that children's intuitive ratio reasoning extends to symbolic numerals. While one previous study found that slightly older 7-10 year old children could compare ratios presented as a frequency distribution (e.g., 4 tokens out of 10 are gold), the stimuli in that study were constructed such that children could have attended only to the numerator rather than truly evaluating the ratio of the numerator to the denominator (Ruggeri et al., 2018). The current study is thus the first to demonstrate that 6-8 year-old children can perform ratio comparisons in both non-symbolic and symbolic format, and moreover, we provide evidence that children compare ratios without relying on unidimensional comparison heuristics.

Consistent with our hypothesis that engaging in non-symbolic ratio reasoning scaffolds symbolic ratio reasoning, we found a significant effect of task order on symbolic ratio comparison accuracy. Children who received the non-symbolic task first performed with higher accuracy on the symbolic task as compared to children who started the session with the symbolic task. One possibility is that engaging in the non-symbolic task first highlighted the conceptual link between the two tasks. The non-symbolic task offers a concrete representation of the ratio comparison problem, which may allow children to create a conceptual model of correct ratio comparison. This conceptual model may remain opaque when children encounter only symbolic representations of the numerical magnitudes involved in the computation. This interpretation is consistent with prior work where non-symbolic representations of a math problem boost symbolic calculation (for examples see Carbonneau, Marley, & Selig, 2013; Fujimura, 2001; Fyfe, McNeil, Son, & Goldstone, 2014; Park et al., 2016) An alternative possibility is that because overall performance is greater on the non-symbolic task, engaging in the non-symbolic ratio comparison task first boosted children’s domain general or domain specific confidence. Under this scenario, practice on the non-symbolic task may not have changed how children solved the symbolic task, but instead increased confidence in their ratio intuitions. Our finding that children did not change the strategies they used on the symbolic task after completion of the non-symbolic task, but still had an overall increase in accuracy, may be interpreted as support for this hypothesis. This type of confidence hysteresis effect, where children perform with higher accuracy when trials move from easy to hard, has been demonstrated in other domains (Hock & Schöner, 2010) . Ultimately, future work utilizing a pretest-training-posttest experimental design is necessary to disentangle the mechanism of this scaffolding effect and to test the robustness of the benefit of non-symbolic ratio practice for symbolic ratio understanding.

The second goal of our experiment was to examine the strategies children use in a binary choice ratio comparison task. First, we found evidence that children did indeed perform a true ratio comparison taking into account both the preferred and non-preferred color gumball in each machine. In other words, children's performance on the two ratio tasks could not be explained by any one incorrect heuristic. At the same time, children's errors were not random, but instead followed the known heuristics of picking the ratio with more of the preferred item, or with more total items. This finding is consistent with previous research on young children's intuitive ratio reasoning (Clarke & Roche, 2009; Falk et al., 2012; Jeong et al., 2007; Obersteiner et al., 2015; Shaklee & Paszek, 1985; Siegler et al., 1981). Taken together, our results suggest that children understand the true nature of the task, but also continue to exhibit a bias toward the ratio with the greater number of target items and the ratio with a greater number of total items. It is possible that children rely on these incorrect heuristics more when the task is difficult. Future work should test children's ratio intuitions with an even larger stimulus set to allow for an analysis of whether the use of incorrect unidimensional heuristics depends on the difficulty of the ratios being compared.

The third and final goal of this study was to examine whether the ANS serves as a cognitive foundation for intuitive ratio reasoning, and to test whether non-symbolic calculation may be a mechanism of the relation between ANS acuity and symbolic math. Our hypothesis was that sharper ANS acuity would allow children to create better intuitive ratio models. Specifically, sharper ANS acuity supports more accurate non-symbolic ratio calculation which would in turn provide a better conceptual scaffold for symbolic ratio comparison. Consistent with this hypothesis, non-symbolic ratio accuracy mediated the relation between ANS acuity and performance on the Key-Math-3 Data Analysis and Probability section. This Key-Math-3 section included questions such as "Which spinner gives you an equal chance of landing on green or white?” and “How many cats are in the pet store?” while displaying a pictograph. Thus, non-symbolic ratio comparison mediates the relation between ANS acuity and formal math questions about probability and counting objects. This finding is in line with work demonstrating that non-symbolic calculation is a better predictor of symbolic math skill than ANS acuity alone, and is more support for the existence of an intuitive Ratio Processing System (Matthews et al., 2016; Park & Brannon, 2014; Pinheiro-Chagas et al., 2014; Starr et al., 2016).

However, we found the reverse relation when ANS acuity and non-symbolic ratio accuracy were used to predict the Key-Math-3 Numeration section. In this case, ANS acuity was a better predictor of basic numeration concepts than non-symbolic ratio accuracy. This finding is consistent with work correlating sharper ANS acuity with better early symbolic math skills in children (Schneider et al., 2016). The Key-Math-3 Numeration section included questions such as “Read these numbers from least to greatest” and “Starting at forty-one, count up by tens”. Thus, the Key-Math-3 Numeration section tests general symbolic number skills, whereas the Data Analysis & Probability section tests concepts that are more similar to the non-symbolic ratio task. Taken together, our mediation analyses suggest that non-symbolic ratio calculations serve as a conceptual scaffold for symbolic math problems that conceptually overlap, but not when they are unrelated. Not surprisingly, non-symbolic calculation cannot function as a conceptual scaffold for symbolic calculation when the required calculations differ dramatically from each other. Future research should test whether a more general measure of non-symbolic approximate calculation, beyond ratio calculation alone, mediates the relation between ANS acuity and general math skill. Such work would identify the degree to which algorithmic overlap between the non-symbolic and symbolic math tested is required for mediation of the relation between ANS acuity and symbolic math.

It is our hope that the current experiment will inspire many future explorations of how non-symbolic numerical reasoning can scaffold symbolic numerical reasoning. One future direction concerns the specificity of how non-symbolic part-part ratio representations could provide a foundation for the subset of symbolic math skills that includes probability, fractions, and ratios. For example, in the current ratio comparison task design all ratios are represented in a part-part relation, but fractions are represented in a part-whole format (e.g. X out of X gumballs are blue). It is unknown if our facilitation and strategy analyses would change if the symbolic ratios were represented in a fraction format. It is certainly possible we could see an increased use of the More Items strategy if the total number of items were actually displayed during the task. Another future direction concerns the relation between ANS acuity, ratio comparisons, and symbolic number comparison ability. While the current experiment explored the relation between non-symbolic numerical comparison, symbolic math, and non-symbolic calculation, we did not include a measure of symbolic number comparison in the current study. The inclusion of symbolic numerical comparisons in a future study could determine if non-symbolic ratio comparison mediates the relation between symbolic magnitude comparison and symbolic math, or whether this relation is specific to non-symbolic magnitude comparison only. Finally, it would also be important for future work to explore the role of executive function and short-term memory in these relationships given the importance of these skills for general mathematics.

The results of the current experiment highlight that even before children begin to learn about fractions and ratios in school, they possess an intuitive ratio reasoning capacity. This intuitive ratio reasoning skill extends to numerals and allows them to make approximate ratio comparisons before they learn the algorithms to solve the same problems precisely. The results further suggest that grounding children's symbolic ratio reasoning in isomorphic non-symbolic problems may facilitate learning. Children have a notoriously difficult time understanding ratios in school (Siegler et al., 2013). This difficulty has sometimes been attributed to a "Whole Number Bias", the tendency for children to overgeneralize integer calculation to rational numbers (DeWolf & Vosniadou, 2015; Ni & Zhou, 2005). Emphasizing children's intuitions about ratios as well as whole numbers at an early age holds the potential to mitigate some of this conceptual confusion. A greater understanding of the intuitive mathematical knowledge children possess may help educators make abstract mathematical computations more accessible to young learners.

CHAPTER 2: INTUITIVE DIVISION IN NON-SYMBOLIC AND SYMBOLIC FORMAT WITHOUT INSTRUCTION

Introduction

Arithmetic skills underlie the entire elementary school math curriculum (Common Core State Standard Initiative 2019). Elementary math is particularly important because a mastery of early arithmetic begins a cascade that unlocks the opportunity to study more advanced branches of mathematics such as algebra, geometry and calculus. According to the US Common Core Standards children learn arithmetic operations in a sequence starting with addition and subtraction, then multiplication, and finally division beginning in grade 3. Division is commonly introduced as the inverse of multiplication, and children's early understanding of division is mediated via multiplication. Only later in more advanced math education do these representations diverge (Campbell, 1997; Mauro, LeFevre, & Morris, 2003). There is neural and behavioral evidence that division remains more effortful than the other basic arithmetic operations even into adulthood (Ischebeck, Zamarian, Schocke, & Delazer, 2009; Rosenberg-Lee, Chang, Young, Wu, & Menon, 2011). These findings suggest that division is the most difficult of the four basic arithmetic operations. However, this greater difficultly may be a function of how formal division is taught, and not a fundamental aspect of the operation.

Children have some basic intuitions about division before they formally learn how to divide. These basic intuitions may derive from insights into practical mathematics in the world around them. These insights are called intuitive action schemas (Correa, Bryant, & Nunes, 1998; Jitendra & Hoff, 1996; Riley, 1984). For example, an action schema for subtraction could be "You have a set of items, you take away a portion of those items, and you see how many you have left". One hypothesized action schema that supports division is children's knowledge of how to create a fair share of a set of items by dividing amongst people (Blake & McAuliffe, 2011; Shaw & Olson, 2012; Sheskin et al., 2016). For example, when children are asked to split up a set of items for themselves and another, children are able to split sets of items to create a numerically fair split by using an alternating handout strategy (Sheskin et al., 2016). But importantly, while the concept of equal sharing may foster a sense of proto-division, sharing tasks do not require performing a true division operation. True division requires using both the dividend and divisor to derive a quotient. In a sharing task, children do not always choose to share equally. Moreover, children do not have to understand that splitting a larger dividend will lead to a larger quotient, or that a smaller divisor will also lead to a larger quotient. For example, children may not realize that when dividing a fixed amount equally among a small group each recipient will get more than when dividing the same amount amongst a larger group (Correa et al., 1998). Thus, while children may create a rudimentary concept of division based on their understanding of sharing, evidence that children know how to share does not demonstrate that children understand the mathematical operation of division.

Another way in which children may begin to form a concept of division before formal instruction is through children's ability to use many-to-one counting to solve multiplication and division word problems. Many-to-one counting is the ability to link several objects to a set of target objects. For example, children can determine how many flowers they will need to put three flowers in each of four vases by counting 1-2-3, 4-5-6, 7-8-9, 10-11-12 and using the cardinal principle to deduce that the last number in the count list is how many flowers they will need (Sophian & Madrid, 2003). Four-year-olds can be taught to answer questions like "How many cookies are needed to give four dogs three cookies each" by having an experimenter model many-to-one counting with physical objects (Blöte, Lieffering, & Ouwehand, 2006). Children can use a version of this logic to solve division word problems. For example, when kindergarteners were presented with the problem "Tad has 15 guppies. He put 3 guppies in each jar. How many jars did Tad put guppies in?" children demonstrated a strategy of using many-to-one counting by counting out 15 guppies into groups of 3, and then counting the number of groups (Carpenter, Ansell, Franke, Fennema, & Weisbeck, 1993). However, these strategies usually require external support and may only work for small set sizes.

A third, currently untested, way in which children could develop an intuitive sense of division is through the Approximate Number System (ANS). The ANS allows children to approximately represent, compare, estimate, and calculate with large sets of objects (Feigenson et al., 2004). A substantial body of work has linked the acuity of the ANS to symbolic math abilities of both children and adults (Chen & Li, 2014; Fazio et al., 2014; Schneider et al., 2016). Importantly, the ANS allows adults, children, and even infants and non-human primates to manipulate non-symbolic numerical representations in mathematical operations. Adults, children, infants, and non-human primates can use their ANS representations to add and subtract arrays of objects (Barth et al., 2006, 2005; Cantlon et al., 2015; Gunderson, Ramirez, Beilock, et al., 2012; Knops et al., 2009; McCrink & Wynn, 2004; McNeil et al., 2011; Pica et al., 2004; Pinheiro-Chagas et al., 2014; Xenidou-Dervou et al., 2014). Young children are also capable of computing scaling operations on large arrays of objects, as well as performing multi-step operations (Barth et al., 2009; McCrink et al., 2016, 2013; McCrink & Spelke, 2010, 2016). Before formal math education, children can also solve addend unknown algebra problems (Kibbe & Feigenson, 2015, 2017), and compare ratios of discrete sets of items (Falk et al., 2012). This work indicates that ANS representations can be used in a variety of non-symbolic and approximate mathematical operations, however, to date no one has examined whether children can use the ANS to represent a true non-symbolic division operation with variable dividend, divisor, and quotient values.

Prior work demonstrates that children can compute a non-symbolic scaling operation across large quantities using the ANS (Barth et al., 2009; McCrink et al., 2016, 2013; McCrink & Spelke, 2016). In the non-symbolic, approximate scaling task used by McCrink and colleagues a child first sees a large set of items (ranging from 24 to 128). Then, the set of objects is hidden behind a white box. A 'dividing wand' appears on top of the box and the child is told "Look! They're getting divided". During training, the child watches this same dividing wand halve (or quarter in another version of the experiment) a set of objects. Then the child imagines the new amount and compares their imagined quotient to a target set of objects by picking the larger set. Children's accuracy is dependent on the ratio between the halved or quartered array and the target array shown, a hallmark of ANS representations. Scaling operations are a specific case of a division operation. They involve a constant divisor whereas a true division operation requires that both the dividend and divisor be allowed to hold multiple values. Thus, it is still an open question whether children can use their ANS to compute a true division operation.

The first major goal of the current experiment is to determine whether young children who have yet to learn about division in school have the ability to intuitively divide large quantities of objects with their ANS using a non-symbolic divisor that varies from trial to trial. To answer this question, we developed a novel non-symbolic division test (Figure 7). In our paradigm, children first see a large quantity of dots (ranging from 32 – 192) at the top of the screen and a flower with a particular number of petals at the bottom. Children are told that the same number of objects will fall onto each petal of the flower. Children watch this happen multiple times with flowers that vary in number of petals. Then, we tell the child that they will no longer be able to see how many 'pollen droplets' fall onto each petal, but that they have to imagine how many will land on each petal. Children watch the dots fall toward the flower, but this time they have to imagine how many dots landed on the petal. To respond, children then compare their imagined quotient to a new visible set of dots and pick the array that is greater in quantity. This paradigm provides a true test of the division operation where both the dividend (the set of objects at the top of the screen) and the divisor (the number of flower petals) vary from trial to trial. Subjects are not required to learn a symbol for halving or quartering as in the scaling paradigm, but instead flexibly divide by a different divisor ranging from 2-8 on every trial. Our divisor is inherently non-symbolic, because the number of petals represents the divisor in a concrete manner. Testing multiple divisors within one subject allows us to ask whether children truly have an intuitive sense of division, or whether children are limited to using an intuition about halving or quartering. There is some evidence to suggest that dividing by two is easier than dividing by any other quantity (Singer-Freeman & Goswami, 2001; Spinillo & Bryant, 1991). Previous work also demonstrates that performance is worse as a scaling factor increases on non-symbolic scaling and proportional reasoning tasks (Boyer & Levine, 2012; McCrink & Spelke, 2010, 2016). Our current design allows us to test if children can succeed with divisors other than 2 and 4, and moreover, how the magnitude of a divisor affects a child's intuition of division. Based on previous research, we hypothesized that subjects would perform with worse accuracy as the divisor increased, and that they would be particularly good at dividing by two.

Our second major goal was to determine whether intuitive division ability is related to and can be integrated with children's formal math abilities. Our sample includes children aged 6 to 9, which spans the age range before and during the beginning of formal division instruction. To measure children's formal division knowledge, we gave children a test that targeted their formal addition and division knowledge. Our aim was to find children who knew some math (addition) but little or no division. Our test asked children to solve addition and division equations and word problems (8 ÷ 4 = ?) and to identify the division symbol. By splitting up our sample based on performance on this test, we can determine whether children without any formal division knowledge can successfully compute a division calculation. The second part of this goal was to test whether children can integrate their division intuition within their symbolic math system. To test this, we created a parallel symbolic version of our intuitive division task where the dividend and the comparison target quantity were represented with Arabic numerals instead of arrays of dots (Figure 7). We then tested whether children who ostensibly cannot perform a division calculation were able to approximately use numerals to divide.

Finally, the third goal of the current experiment was to examine whether intuitive division skill is rooted in the ANS, and if this ability is related to general symbolic math skill. Prior work demonstrates that ANS acuity and symbolic math performance are correlated in children and adults, however, recent findings suggest that non-symbolic calculation tasks may be better predictors of symbolic math ability than ANS acuity (Matthews, Lewis, & Hubbard, 2016; Pinheiro-Chagas et al., 2014; Starr, Roberts, & Brannon, 2016, Szkudlarek & Brannon (under revision)). Here we test the possibility that non-symbolic calculation, in this case intuitive division, is a mechanism of the relation between ANS acuity and symbolic math. We hypothesize that sharper ANS acuity allows for better non-symbolic division calculation. In turn, better non-symbolic calculation allows children to conceptually ground their formal sense of division with their intuitions about division when they begin to learn arithmetic in school. This stronger concept of division may lead to a sturdier ability to tackle symbolic division calculation in the classroom. We test this hypothesis with a mediation analysis. If non-symbolic division accuracy fully or partially mediates the relation between ANS acuity and symbolic math ability, this is evidence for our hypothesized mechanistic pathway. To determine whether a sturdier concept of arithmetic continues to yield benefits for symbolic math ability into adulthood, we also tested a sample of college undergraduates on our approximate division tasks, a measure of ANS acuity, and a test of symbolic math and ran the same mediation analysis. Overall, this analysis tests whether enhanced non-symbolic calculation skill as a result of sharper ANS acuity will lead to better symbolic math test scores. Again, a full or partial mediation would support our hypothesis that non-symbolic calculation may be a mechanism of the established relation between ANS acuity and symbolic math skill.

In sum, the current experiments have three main goals. The first is to determine whether children can perform a true and entirely non-symbolic division operation before formal instruction with division. The second goal examines the relation between formal division knowledge and non-symbolic division performance, and tests whether children who do not yet exhibit any formal division knowledge are nevertheless capable of dividing approximately using Arabic numerals. Finally, the third goal is to test whether non-symbolic division performance mediates the relation between ANS acuity and symbolic math performance in both children and adults.

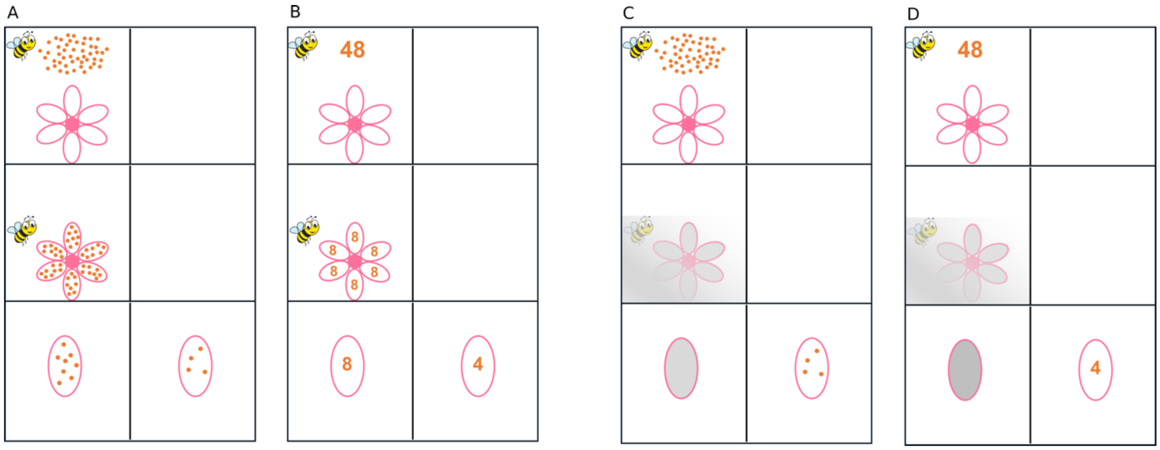

Figure 7. Schematic of the non-symbolic and symbolic division tasks. A) Demonstration trials for the non-symbolic task. Children and adults watched as the dots on the top of the screen fell onto the petals below. Then, one of the flower petals moved towards the center of the left side of the screen and the target petal appeared with a new quantity of dots to compare on the right side of the screen. Children responded by touching the petal with the greater quantity of dots. Children completed 8 demonstration trials. The demonstration trials are not included in any analysis. B) The demonstration trials for the symbolic version of the task. Children watched the numeral at the top of the screen split apart and change into the numerals that landed on the flower at the bottom of the screen. C) Experimental trials for the non-symbolic division task. The task was identical to the demo trials, except that as the dots fell to the bottom of the screen a cloud appeared that obscured how many dots fell onto each petal. Then the obscured petal moved to the middle of the left side of the screen and children had to imagine how many dots were on this petal and compare it to the visible target quantity. D) The experimental trials for the symbolic version of the task. Children watched the numeral move and disappear into the fog. Then children compared their imagined quantity to the new target number on the right side of the screen.

Materials and Methods

Child Experiment

Subjects. Eighty-nine 6-9 year old children participated in Experiment 1 (mean age = 7.9 years old, standard deviation = 1.1 years; 50 female, 39 male). Written parental consent was collected in accordance with a protocol accepted by the University of Pennsylvania's Institutional Review Board. Thirty-two additional children were consented but were excluded from the final sample because they did not complete both the non-symbolic and symbolic ratio comparison tasks due to early pick up or absence from their after-school program. The parents of 88 children in the sample completed a detailed demographics questionnaire. 86% identified as African American, 10% as Caucasian, 2% as Asian, and 2% as more than one race. Our sample included a large proportion of children from families with household incomes of 50,000 or less (7% $150,000+, 6% $100,000 - $150,000, 4% $50,000 - $75,000, 45% $25,000 - $50,000, 30% $0 - $25,0000, and

8% chose not to report). All subjects were recruited from six after school programs in the Philadelphia, PA area. A subset of the children who completed both the non-symbolic and symbolic ratio comparison tasks completed additional assessments (Dot comparison, n = 84; Key-Math Numeration subtest, n = 89; Division knowledge assessment, n = 82; the Woodcock-Johnson Basic Reading Skills cluster, n = 77; and a measure of numeral identification, n = 80). All participants received a small toy as a thank you gift after completion of the experiment.

Forty-two children (mean age = 8.03, standard deviation = 1.2 years) were tested in Experiment 2 to replicate Experiment 1 with a different stimulus set. These children were recruited from the same after school programs in the Philadelphia, PA area, but none of the children participated in Experiment 1. Children in Experiment 2 were tested on the non-symbolic and symbolic division tasks, a measure of numeral identification and the formal division test.

Procedure. Children completed all tasks individually with an experimenter in a quiet room in their after-school program. Children completed the non-symbolic and symbolic ratio comparison tasks first and the order of the two tasks were counterbalanced across children. The order in which all other tasks were administered was random across participants and was dependent on the duration of the task and the child's availability. Each participant was tested for a total of 45-60 minutes across 2-3 days. Children also completed a math anxiety questionnaire the results of which will be reported in a separate paper. Children received stickers throughout the session to maintain motivation. In Experiment 2, children completed all tasks in a single day.

Experiment Tasks.

The approximate division tasks and the dot comparison task were run in MATLAB and programmed using the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007). The programs were run on a 15-inch touch screen laptop computer.

Introduction to the division tasks. All children regardless of whether they were given the symbolic or non-symbolic version of the division task first were introduced to a bee named "Buzz" displayed on the computer screen. The children heard the following narrative as they watched the experimenter display pictures on the computer screen: “Buzz flies to flowers to find food to bring back to his hive. Buzz lands on the flower to get the food, and some of the food sticks to him. When Buzz flies away from the flower, some of the food falls down onto the flower.” A picture was displayed on the computer screen of Buzz above a flower with two petals. Four pollen droplets were shown falling towards the flower with two dots landing on each of the two petals. The children were told “The same amount of food falls on each petal of the flower”. The children were then shown Buzz above a flower with four petals and eight pollen droplets falling towards the flower. The children were then told “We can see the food falling down towards the flower. See how even if the flower looks different, the same amount of food falls onto each petal” Children were then shown Buzz above a flower with eight petals and eight pieces of food. The instructions were repeated one more time. The experimenter never mentioned the specific number of pollen droplets or petals, but only told the children that the same amount of food falls onto each petal. After these initial instructions the experimenter asked the child if they were ready to play the game.

Non-Symbolic Division. After the initial instructions the experimenter started the demonstration phase of the game. In this phase children watched an animated set of dots fall onto the petals of a flower. On demonstration trials, children could see how many pollen droplets fell onto each flower petal. In this task the initial number of pollen droplets is the dividend, the number of petals is the divisor, and the number of pollen droplets in one petal is the quotient. After the pollen droplets fell onto the flower petals the flower disappeared and one of the petals from the flower moved to the middle of the left side of the screen. A new flower petal with pollen droplets inside appeared on the right side of the screen. Then the experimenter asked “Which petal has more food?”. The child was told they should touch the petal to indicate their answer. The trial did not progress until the child made their response, but the experimenter encouraged the child to make their choice quickly. Once the child touched a petal, a happy bee with the words “Great job!” appeared for the correct response or a sad bee and the words “try again!” appeared for an incorrect response. Then, a screen appeared with Buzz in the center with a garden background.

The child was told to touch Buzz to continue playing the game. Once the child touched Buzz the next trial started. Children completed eight demonstration trials. The purpose of these trials was to make sure children understood that the same number of dots fall into each petal of the flower, and that their job was to pick which petal has more food. These trials were not used in the analysis of task performance. During the demonstration phase children only saw flowers with 2,5 or 8 petals. The experimenter did not allow the child to continue to the next phase of the game unless they were able to successfully pick which petal had more food.

After completing the demonstration phase children were told "Now it is a foggy day in the garden but Buzz still needs your help. You won't have enough time to count all the food Buzz is carrying, and because of the weather, you won't be able to see the amount of food he puts on each petal. Instead, you'll need to *imagine* how many pieces of food are on each petal." The experiment started the first trial. The array of dots and the empty flower appeared on the left side of the screen like before, but this time when the dots began to fall toward the flower a fog appeared over the flower. Child could still see the flower, but the pollen droplets were obscured before they were distributed onto the petals. Children could thus no longer see how many pollen droplets landed on each flower petal. Like before, the flower then disappeared and one of the petals moved up to the middle of the left side of the screen. This time the inside of the petal was grey so that the child could not see the number of dots inside the petal. Another flower petal appeared on the right side of the screen. In the petal on the right side children could see a set of pollen droplets. Then the experimenter said "Ok, which petal has more food? Try and imagine how many pieces of food are on this [gesture to left] petal even though you can't see them!" Children then responded by touching the petal they thought had the greater quantity of dots and received feedback. Children completed 32[4] trials with feedback. During these 32 trials children saw flowers with 2,5, or 8 petals. To test whether children could generalize to new divisors, after

---

[4] 11 subjects ran a longer version of the division tasks (53 trials with feedback and 40 trials without feedback) but the number of trials for the remaining subjects was reduced due to time constraints at the after-school programs where testing occurred.

the completion of the first 32 trials children completed 24[5] more trials with 3 or 6 petals without feedback. Throughout all trials the experimenter never mentioned any number words. Accuracy and reaction time were recorded for each trial.

Symbolic Division. The symbolic division task was identical to the non-symbolic version. The only difference between the two tasks was that during the symbolic task the dividend and quotient were Arabic numerals instead of sets of dots. The instructions and the stimulus set remained identical between the both versions of the division tasks.

Numerical values for the division tasks. We created a stimulus set (Figure 8) for the approximate division tasks such that children had to pay attention to all three numbers involved (dividend, divisor, target) to solve the task successfully. To accomplish this, we first required all of the target comparison numbers to be within the range of possible quotients. This constraint removes the possibility that a child would pick the target number simply because it is larger than any number they have seen in a petal before. We then minimized the correlation between the dividend and the divisor and whether the quotient or target was correct across the stimulus set. Specifically, this required including trials where the dividend was large relative to our range of potential dividend values (32 – 192) and the correct answer was the target, and trials where the divisor was small relative to the range of potential divisors (2-8) and the correct answer was the target. The quotients used (8,10,13,17, 22, 29, 37, 48) were picked to be approximately evenly spaced in log space. The targets used were the exact same values as the quotients. Because they are evenly spaced in log space, the same difference (one, two, three or four) between any two quotients will correspond to roughly the same change in ratio. This is important because we know the difficulty between comparing any two numerosities is dependent on the ratio between them (Feigenson et al., 2004). For example, a trial in which the quotient is 13 and the target is 22 should be just as challenging as a trial where the quotient is 17 and the target is 29 (both spaced

[5] One subject ran 9 extra trials without feedback due to a computer malfunction. Two subjects ran a majority (19 or 20 out of 24) but not all 24 trials due to an early pick up time. These subjects are included in the analysis.

two apart in our list of quotients). To be able to fairly compare accuracy for each divisor we ensured that at each divisor there were the same number of trials corresponding to one, two, three, and four difficulty levels. Taking all of this into consideration, during Experiment 1 with children and the adult experiment we excluded the divisors of 4 and 7 from our range of possible divisor values due to time constraints on testing. In Experiment 2 with the children we included a divisor of 4. The exact stimuli drawn from the stimulus space in Figure 8 for each experiment are presented in Appendix B.1 and B.2. Experiment 1 with children and the adult experiment included 32 trials with feedback testing divisors of 2 5 and 8, and 24 trials without feedback testing divisors of 3 and 6. In Experiment 2 children completed 32 trials with feedback testing divisors of 2,5 and 8, and 24 trials without feedback testing divisors of 3,4 and 6.

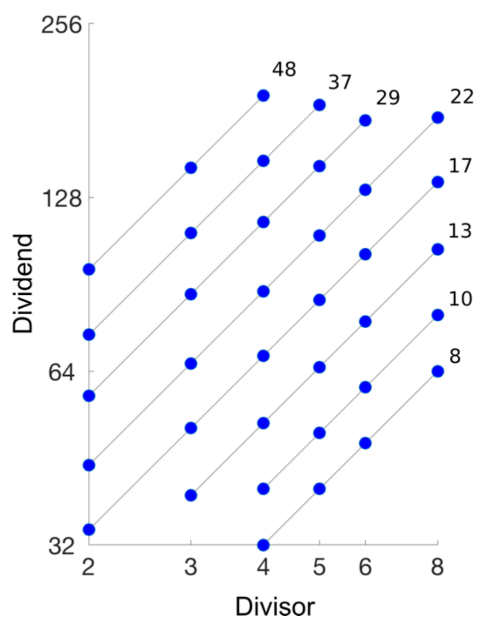

Figure 8. The stimulus space used in all experiments. The divisor is the number of petals on the flower and the dividend is the number of dots that appeared at the top of the screen at the beginning of a trial. Each diagonal line represents a given quotient (the quantity in one petal). The quotients were chosen to be approximately evenly separated in log space such that any vertical jump between two points corresponds to roughly the same ratio. The same values (8,10,13,17,22, 29, 37, 48) were used for the target comparison values that appeared on the right side of the screen. The quotient was the correct choice on 50% of trials.

Dot Comparison Task. Two dots arrays appeared on a black screen for 750ms. The arrays were then occluded and the task was to touch the location of the numerically larger array. Children completed 200 trials and were given feedback on every trial. The number of dots ranged from 8 to 32. The stimuli were created to evenly sample a stimulus space that varied by the ratio between the number, size, and the spacing of the dots. To encourage greater reliability of the measurement, trial level difficulty was titrated (Lindskog et al., 2013). The titration procedure calculated the percentage correct over the last 5 trials. The ratio between the two dot arrays moved one log level farther apart if the accuracy was less than 70% and moved one log level closer together if the accuracy was greater than 80%. A quantitative index of each child's ANS acuity was calculated as a Weber fraction (*w*) as specified in (DeWind et al., 2015). This model accounts for the effects of non-numerical features of dot arrays on numerical discrimination (DeWind et al., 2015).

Numeral Identification Task. The numerals 1-30 were printed and displayed individually on index cards. The numerals were displayed in random order to the child, and the child was asked "What number is this?" The accuracy of each child's response was recorded.

Key Math-3 Diagnostic Assessment. The Numeration section of the Key Math-3 Diagnostic Assessment Form B (Connolly, 2007) was administered. The Numeration section is a test of general basic math skills like place value, counting, the relative magnitude of numbers, and an understanding of fractions, decimals, and percentages. We used the age standardized scale score.

Woodcock-Johnson IV Test of Cognitive Abilities. Participants' reading abilities were assessed using the "Basic Reading Skills" cluster of the Woodcock-Johnson. This cluster is comprised of the "Letter-Word Identification" and "Word Attack" subtests. In the "Letter-Word Identification" subtest, participants named letters and read words aloud. In "Word Attack," participants read nonsense words and identified letter sounds. We used the age standardized Basic Reading Skills score.

Formal Division Test. This test consisted of 15 questions that examined children's addition and division knowledge. Six items were word addition and division word problems, eight items were symbolic arithmetic problems, and one item required the experimenter to show the child a picture of the division symbol (÷) and ask "Do you know what this symbol is?". For each symbolic arithmetic problem, the child was shown a flashcard with the arithmetic equation displayed as the experimenter read the problem aloud to the child. The test questions are reproduced in Appendix B.3.

Adult Experiment

Subjects. Participants were eighty-seven undergraduates (mean age 20.7 years old, 51 female). Written and informed consent was collected in accordance with a protocol accepted by the University of Pennsylvania's Institutional Review Board. Seven participants did not return to complete the second session which consisted of the dot comparison task, the fraction magnitude comparison task, and the addition verification task. The data from two dot comparison scores and two fraction magnitude comparison scores were lost due to a computer error.

Procedure. Subjects completed all experimental tasks in two sessions that took place on separate days no more than 3 days apart. Testing occurred in a quiet room on a touch screen desktop computer. During the first session adults completed the non-symbolic and symbolic division tasks in counterbalanced order, the vocabulary test, and the division strategy questionnaire. During the second session subjects completed an addition verification task, a dot comparison task, and a fraction magnitude comparison task in counterbalanced order. Subjects also completed a math anxiety questionnaire, but this data is not included in the current report. Participants received course credit as compensation.

Experiment Tasks.

Non-symbolic and symbolic division tasks. The tasks, instructions, and stimuli were identical as that described above for the children. The participants were told that this task was created for use with children to explain the presence of the cartoon bee and storyline. These tasks were run on a touch screen monitor on a desktop computer.

Dot comparison task. The task was the same as that described for the children.

Vocabulary Test. Subjects answered 42 multiple choice vocabulary questions in 5 minutes. The questions were taken from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman & Dermen, 1976). Performance was calculated as the number of problems answered correctly minus ¼ of the number incorrect to discourage guessing.

Addition Verification Test. In this task, 1 and 2-digit addition and subtraction problems were displayed horizontally with a proposed answer (e.g., 27 + 52 = 79). The task was to press one key if the statement was correct and another if it was incorrect and was modeled after (Klein et al., 2010). On incorrect trials (50% of all trials) the sum displayed were +/- 10 or +/- 2 from the correct sum. Participants had 10 seconds to make a response by pressing the F or J key on the keyboard. Subjects completed two blocks of 96 trials each. Performance was quantified as the median reaction time and accuracy on each trial.

Fraction Magnitude Comparison Task. Subjects viewed two fractions displayed in the middle of the screen in white on a black background. The goal of the task was to pick the fraction greater in magnitude by pressing the F key for the left fraction or the J key for the right fraction. The stimuli were the same as used in (Fazio, DeWolf, & Siegler, 2015). Performance was quantified as the median reaction time and accuracy on each trial.

Division Strategy Questionnaire. The goal of this questionnaire is to examine the strategies adults used to solve the non-symbolic and symbolic division tasks. It consists of 8 open ended and multiple choice questions. The full questionnaire is reproduced in Appendix B.4.Thirty subjects completed a shorter version of the questionnaire without questions 3 and 4. The analyses reported in this paper focus on question 3 “Which task was more difficult, the flowers & bees task with dots or with numbers?”, question 4 “How many petals appeared on a flower? (circle all that you remember)” and questions 7 and 8 “Which strategy best describes how you solved the task with the bee, flowers, and dots/numbers?”.

Results

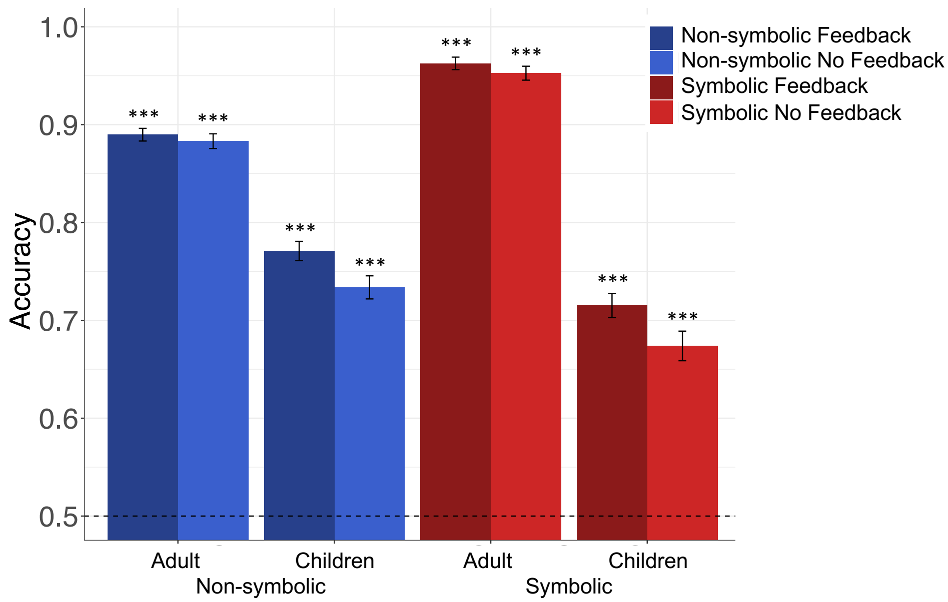

Figure 9. Accuracy on the non-symbolic and symbolic division tasks for both the children (Experiment 1) and adults. *** = p < .001. The dotted line depicts chance performance. Error bars represent standard error of the mean.

***Non-symbolic Division Performance**.* As shown in Figure 9, children and adults performed well above chance on both the feedback (children 77%, t_{88} = 27.4, p <.001; adults 89%, t_{86} = 60.1, p <.001) and no feedback (children 73, t_{88} = 19.8, p <.001; adults 88%, t_{86} = 51.2, p <.001) phases of the non-symbolic division task. These data indicate successful generalization of the division operation to novel divisors and demonstrate that both children and adults engaged in true division.

In Experiment 2, we replicated the finding that children can perform non-symbolic intuitive division and generalize to novel divisors with a different set of numerical values chosen from the same stimulus space (feedback 68% accuracy $t_{41} = 12.8$, $p < .001$; no feedback 74% accuracy $t_{41} = 15.9$, $p < .001$).

To test whether accuracy was dependent on the ratio between the quotient and the comparison target value, we ran a generalized linear mixed effects model (GLMM) following a binomial distribution predicting whether each trial was correct with the ratio level between the quotient and target as a fixed effect and a random effect of subject. As expected, this model indicated significant main effects of ratio for both adults and children (Figure 11A; adult $\beta = .80$, $z = 16.1$, $p < .001$; child $\beta = .36$, $z = 12.2$, $p < .001$). This analysis indicates that accuracy was dependent on ratio between the quotient and the target.

Symbolic Division Performance*.* As shown in Figure 9, children and adults performed well above chance on both the feedback (children 72%, $t_{88} = 17.4$, $p <.001$; adults 95%, $t_{86} = 77.8$, $p <.001$) and no feedback (children 67%, $t_{88} = 11.5$, $p <.001$; adults 95%, $t_{86} = 63.6$, $p <.001$) phases of the symbolic division task. Similar to the results on the non-symbolic version of the task, adults and children were able to generalize to novel divisors when the dividend and target were represented with numerals. We replicated this above chance performance with children in Experiment 2 with different numerical values drawn from the same stimulus space (feedback 62% accuracy $t_{41} = 5.99$, $p < .001$; no feedback 60% accuracy $t_{41} = 6.01$, $p < .001$).

To test whether accuracy was dependent on the ratio between the quotient and the comparison target value, we ran a generalized linear mixed effects model (GLMM) following a binomial distribution predicting whether each trial was correct with the ratio level between the quotient and target as a fixed effect and a random effect of subject. As expected, this model indicated significant main effects of ratio for both adults and children (Figure 11A; adult $\beta = .57$, $z = 7.69$, $p < .001$; child $\beta = .24$, $z = 8.69$, $p < .001$).

Adult and Child Division Format Effect*.* To compare the performance of adults and children, we ran a mixed effects ANOVA predicting overall performance on the division tasks with a main effect of task format (symbolic or non-symbolic) and age group (adult or child), an interaction between format and age, and a random effect of subject. Unsurprisingly, there was a main effect of age group on division performance, due to adults higher accuracy on the division tasks ($F_{1,174}$ = 331.8, $p < .001$). Additionally, there was a significant age by task format interaction ($F_{1,174}$ = 104.0, $p < .001$). Adults performed with higher accuracy on the symbolic version of the division task (paired t-test: t_{86} = -10.5, $p < .001$), whereas children showed the opposite effect. Children performed significantly better on the non-symbolic version of the task (paired t-test: t_{88} = 5.52, $p < .001$).

We ran the same analysis on median reaction time. Consistent with the accuracy data there was a main effect of age such that adults were faster than children ($F_{1,174}$ = 46.57, $p < .001$; mean RT adult non-symbolic = 1.02 seconds, mean RT adult symbolic = 1.20 seconds, mean RT non-symbolic child = 1.73 seconds, mean RT symbolic child = 1.94 seconds). Unlike the accuracy data, there was a main effect of format indicating that all subjects were faster to complete the non-symbolic task than the symbolic task ($F_{1,174}$ = 3.25, $p < .001$). There was no format by age interaction ($F_{1,174}$ = .027, p = .66).

Child Division and Symbolic Math Knowledge. There were 40 children who could not identify the division symbol in our sample. As shown in Figure 10, children who could not identify the division symbol successfully completed both the non-symbolic and symbolic division tasks with above chance accuracy (non-symbolic feedback, 76% t_{39} = 16.4, $p <.001$; no feedback 71% t_{39} = 11.5, $p <.001$; symbolic feedback, 68% t_{39} = 11.3, $p <.001$; no feedback 64% t_{39} = 7.78, $p <.001$). For the non-symbolic task, there was no significant difference in accuracy between children who could and could not recognize the division symbol (t_{76} = -1.83, p = .07), but children who could recognize the division symbol performed with higher accuracy on the symbolic division task (t_{76} = -3.23, p = .002).

There were 51 children who could not solve any of the four simple symbolic division problems on our formal division test (for example, 6 ÷ 3 = ?). As shown in Figure 10, children who could not solve symbolic division problems were nevertheless significantly above chance on both approximate division tasks (non-symbolic feedback, 77% t_{50} = 20.7, p <.001; no feedback 70% t_{50} = 12.8, p <.001; symbolic feedback, 69% t_{50} = 12.3 p <.001; no feedback 64% t_{50}= 8.20, p <.001). Children who could solve at least one division problem performed with significantly higher accuracy on the approximate division tasks than children who could not solve any division problems[6] (non-symbolic t_{79} = -2.30, p = .02; symbolic t_{79} = -2.60, p = .01). These data indicate that formal knowledge of division is not a precursor to solving the approximate division tasks, in either symbolic or non-symbolic format.

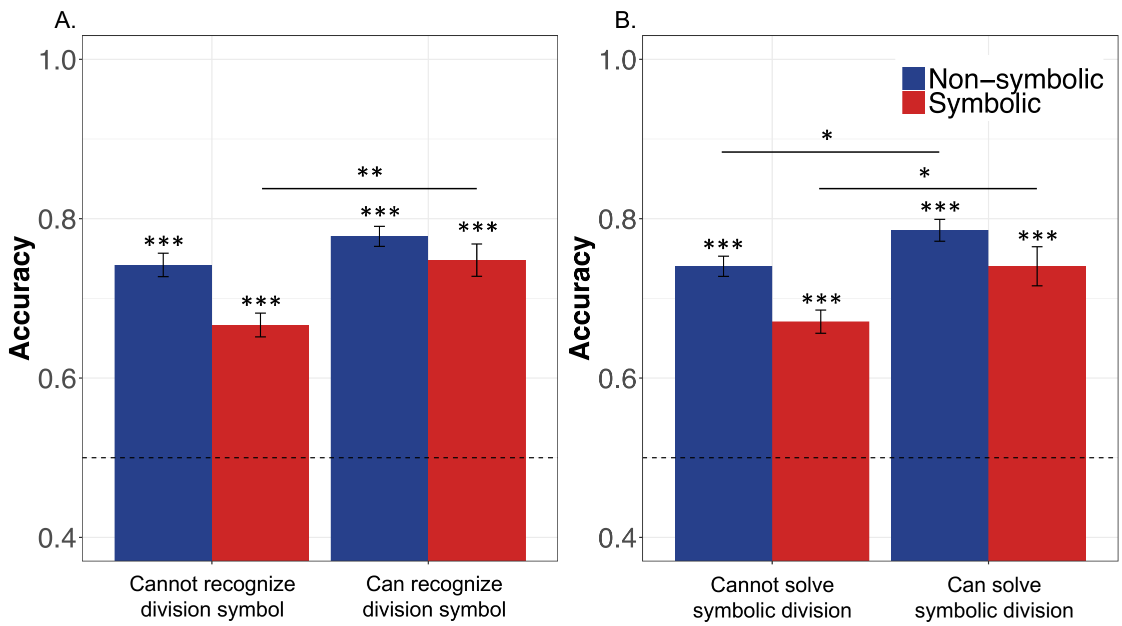

Figure 10. Children's performance on the non-symbolic and symbolic division tasks broken down by their performance on the formal division test. The feedback and no feedback phases are combined for each task. The dotted line represents chance performance. A) Children who cannot recognize the division symbol (÷) still perform at above chance levels on the non-symbolic and symbolic division test. There is no significant difference between performance on the non-symbolic task based on division sign knowledge, but children who can recognize the division symbol perform significantly better than children who cannot on the symbolic version of the division task. B) Children who cannot solve simple division equations (6 ÷ 3 = ?) perform significantly above chance on both approximate division tasks. Children who can solve a division equation perform significantly better than children who cannot.

---

[6] There are three more children included in the analysis of accuracy on the symbolic division problems than in the analysis of division sign knowledge. This is due to experimenter error; the experimenter forgot to mark down three children's answers to the division sign knowledge question during the formal division test.

Alternative Strategy Analysis. It is essential to rule out the possibility that participants were using an alternative heuristic instead of division to complete the division tasks. One possibility is that participants attempted to compare only the divisor to the target comparison number when making their response. Alternatively, participants could attempt to compare only the dividend to the target comparison number. However, the use of either of these strategies would yield chance performance because the target was greater than the divisor on all trials, and the target was less than the dividend on almost all trials (53/56) in Experiments 1 and 2.

We next examined whether participants used a heuristic where they constructed a mental model of the mean of the target value across all trials and evaluated whether the target on a given trial was more or less than the mean. The stimulus set used in Experiment 1 was not designed to rule out this alternative strategy, however, the stimulus set constructed for Experiment 2 ensured that participants could not score above chance if they relied on this strategy. As reported previously, children performed with above chance accuracy on both the symbolic and non-symbolic division tasks in Experiment 2. This was also true when the trials were split by whether the target was greater (non-symbolic 79% $t_{41} = 16.5$, $p < .001$; 62% symbolic $t_{41} = 3.86$ $p < .001$) or less than (non-symbolic 56% $t_{41} = 5.31$, $p < .01$; symbolic 61% $t_{41} = 5.78$, $p < 001$) the mean target value.

Division Strategy Questionnaire Analysis. Consistent with the accuracy data previously reported, the majority of adult subjects who completed question 3 indicated that the non-symbolic version of the division task was more difficult than the symbolic version (43/57). On the non-symbolic task, the majority of participants (47/85, 2 choose not to answer) reported using an approximate strategy by choosing the option "*I got a sense of the amount of dots on the top of the screen and the amount of petals below, and imagined approximately how many dots were on one petal*" on question 7. In contrast, on the symbolic task the majority of participants (50/85) reported using an exact calculation strategy by picking the choice "*I calculated the exact answer using the number up top and the exact number of petals*" on question 8. However, if subjects did use an exact calculation strategy to solve the symbolic division task, they would know the exact number of petals used during the task because they divided by these same five numbers repeatedly (2,3,5,6,8). Of the 34 participants who reported exact calculation on the symbolic task and also completed question 4, only about half could correctly identify the divisors used during the tasks (18/34).

Effect of Divisor on Division Performance*.* We tested the prediction that it is easier to divide by 2 compared to any other value. We ran a mixed effects ANOVA predicting accuracy with age (adult or child) and divisor (two or other) and error nested within subject. For the non-symbolic division task there were significant effects of age ($F_{1,174} = 92.67$, $p < .001$), divisor ($F_{1,174} = 51.62$, $p < .001$) and an interaction between age and divisor ($F_{1,174} = 18.30$, $p < .001$). The symbolic division task followed the same pattern of results with significant effects of age ($F_{1,174} = 280.4$, $p < .001$), divisor ($F_{1,174} = 34.7$, $p < .001$) and an interaction ($F_{1,174} = 7.07$, $p = .009$). Post-hoc tests on the non-symbolic data reveal a significant difference between accuracy on trials where the divisor was two and all other trials (adult mean difference = .026[7], $t_{86} = -2.36$, $p = .02$; child mean difference = .10, $t_{88} = -7.21$, $p < .001$). Similarly, post-hoc tests on the symbolic data reveal a significant difference between the accuracy on trials where the divisor was two and all other trials (adult mean difference = .028[8], $t_{86} = -4.60$, $p < .001$; child mean difference = .08, $t_{88} = -4.59$, $p < .001$). These results demonstrate that for both children and adults dividing by two is easier than dividing by other divisors (Figure 11B). The interaction indicates that this difference is greater for children than for adults.

Second, we tested whether there was a linear effect of divisor on division accuracy. To test whether performance varied as a function of divisor we ran a linear model following a binomial distribution predicting whether each trial was correct with fixed effects of divisor and age (adult or child, child coded with a 1), the interaction between divisor and age, and a random effect

---

[7] The data for the adults was not normally distributed. A Wilcoxcon signed rank t-test returns the same result (pseudo-median = .026, p = .02)
[8] The data for the adults was not normally distributed. A Wilcoxcon signed rank t-test returns the same result (pseudo-median = .04, p < .001).

of subject. We ran this model separately for the non-symbolic and symbolic division tasks. For the non-symbolic division task there were significant main effects of both divisor and age, and a non-significant interaction between the two predictors (divisor β = -.06 , z = -2.71 , p = .007 ; age β = -.71 , z = -4.60 , p < .001; interaction β = -.05 , z = -1.87 , p = .06). For the symbolic division task there was a significant interaction between age and divisor, as well as main effects of both divisor and age (divisor β = -.15 , z = -4.13 , p < .001; age β = -3.11 , z = -12.5 , p < .001; interaction β = .11 , z = 2.77 , p = .006). These results indicate a linear effect of divisor in the predicted direction for both the non-symbolic and symbolic division tasks. Accuracy was lower for both children and adults when the divisor was larger (Figure 11B). On the symbolic division task, the significant interaction indicated that this effect was larger for children than adults.

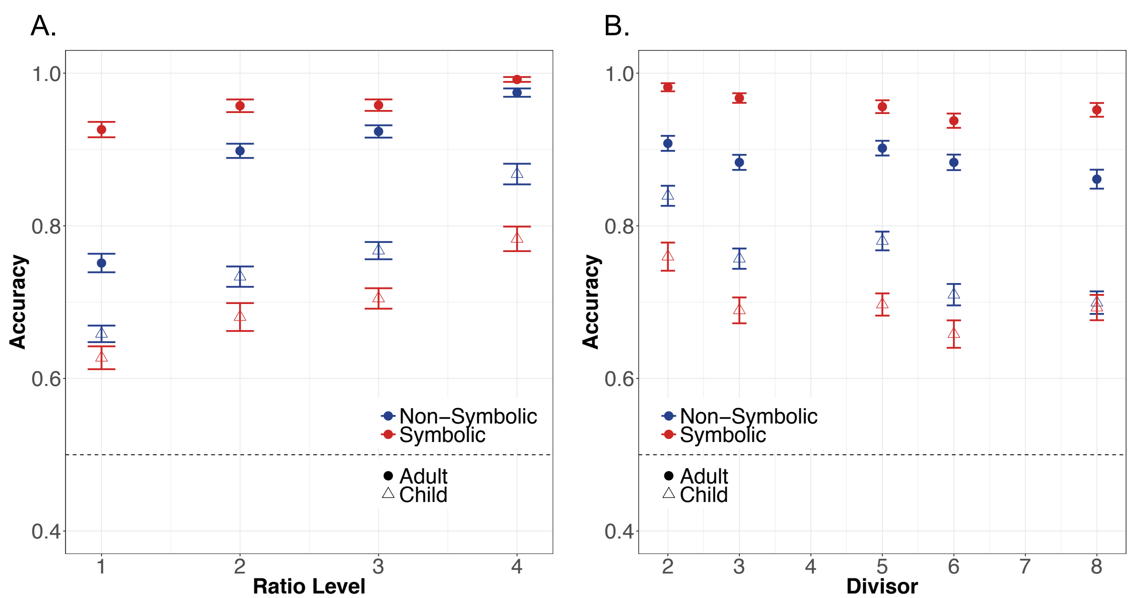

Figure 11. Performance of adults and children in experiment 1 on the non-symbolic and symbolic division tasks broken down A) by the ratio between the quotient and target and B) by the divisor. Error bars represent standard error of the mean. The dotted line represents chance. A) Ratio level 1 ≈ .8, Ratio level 2 ≈ .6, Ratio level 3 ≈ .45, Ratio level 4 ≈ .35. Performance is ratio dependent for the non-symbolic and symbolic division tasks for both the children and adults. B) Both children and adults are better at halving than dividing by any other quantity. There is a small linear effect of the divisor on approximate division accuracy.

The Relation between ANS acuity, Non-symbolic division, and Symbolic math. We first removed any outlier scores that were greater or less than three times the interquartile range for children and adults. This process removed four ANS acuity scores in the child data set. For the adult data set this process removed 3 symbolic division scores, 3 addition verification accuracy scores, and 1 addition verification median reaction time score. We transformed each measure to approach normality for both the child and adult data. We conducted the natural log transformation for child ANS acuity (Shapiro-Wilk W = .96), adult ANS acuity (W = .91), and fraction magnitude reaction time (W = .99), raised adult arithmetic verification accuracy to the ninth power (W = .97) and raised adult fraction magnitude accuracy to the fourth power (W = .98). The adult symbolic division data was highly skewed to the left and did not approach normality with strong transformations (W = .77). Zero order correlations and descriptive statistics are reported for children in Table 2 and adults in Table 3. To ensure that correlations between measures were not simply due to age in the children, we partialed out age from our measures of ANS acuity and symbolic and non-symbolic division.

Table 2. Descriptive statistics and bivariate correlation matrix of the child data

		M	***SD***	**1**	**2**	**3**	**4**
1	**Non-symbolic division**	.755	.08				
2	**Symbolic division**	.710	.12	.52***			
3	**ANS acuity**	.336	.15	-.30**	-.20		
4	**Key-Math-3 Numeration**	10.0	3.6	.37***	.32**	-.34**	
5	**Reading Cluster**	99.4	15	.32**	.22*	-.35**	.62***

Note. *M* = mean, *SD* = standard deviation. The bivariate correlations are controlling for age. ANS acuity is calculated from the dot comparison test. The non-symbolic and symbolic division measures is the total accuracy across the feedback and no feedback trials. Note: *** p<.001 ** p < .01 * p < .05

Table 3. Descriptive statistics and bivariate correlation matrix of the adult data

		M	*SD*	1	2	3	4	5	6	7
1	**Non-symbolic division**	.887	.05							
2	**Symbolic division**	.958	.04	.30**						
3	**ANS acuity**	.162	.04	-.50***	-.06					
4	**Fraction Comparison Acc**	166	17	.38***	.39***	-.32**				
5	**Fraction Comparison RT**	1.62	.49	-.04	.06	-.03	.44***			
6	**Addition Acc**	172	15	.40***	.27*	-.21	.62***	.41***		
7	**Addition RT**	2.00	.60	.09	-.13	-.20	.17	.52***	.18	
8	**Vocabulary**	14.8	7.3	.19	.12	-.33**	-.01	-.12	.00	-.08

Note. *M* = mean, *SD* = standard deviation. ANS acuity is calculated from the dot comparison test. The non-symbolic and symbolic division measures is the total accuracy across the feedback and no feedback trials. Note: *** p<.001 ** p < .01 * p < .05

We ran two mediation models: one for the Key-Math-3 Numeration subtest with children and one for the fraction magnitude comparison test with adults (Figure 12). We did not run the model using the arithmetic fluency measure in adults because this measure was not significantly correlated with ANS acuity ($r = -.21$, $p = .07$). Mediation analyses test for a significant indirect effect (the product of the standardized coefficients a and b) that accounts for some portion of the original direct effect (c). The remaining direct effect is represented as c'. The goal of this analysis was to examine whether non-symbolic division and ANS acuity account for the same or different variance in each symbolic math outcome measure. A full or partial mediation, indicated by a significant indirect effect, would be consistent with our hypothesis that non-symbolic division calculation is a mechanism of the relation between ANS acuity and symbolic math.

ANS acuity was a significant predictor of a child's score on the Key-Math-3 Numeration subtest (standardized $\beta = -.34$, $p = .002$) and of accuracy on the non-symbolic division task (standardized $\beta = -.29$, $p = .009$). ANS acuity continued to be a significant predictor of the score on the Numeration subtest after controlling for the mediator, non-symbolic division accuracy, however the strength of this relation was lessened (ANS acuity standardized β = -.27, p = .02; non-symbolic division accuracy standardized β = .23, p =.04). The significance of this reduction was confirmed when testing the indirect effect with a bootstrap estimation approach with 5000 simulations (indirect effect = -.07, 95% CI = [-.17 -.01], p =.03). The direct effect was also significant, indicating a partial mediation (direct effect = -.27, 95% CI = [-.48 .05], p =.02). The proportion mediated was .20 (p = .03, 95% CI = [.01 .69]). Thus, a higher Key-Math-3 Numeration score was associated with .07 standard deviations sharper ANS acuity as mediated through non-symbolic division accuracy. This finding is in line with our hypothesis, however, when we partialed out the relation between the Woodcock-Johnson Reading Cluster and the Numeration subtest, ANS acuity was no longer significantly correlated with scores on the Numeration subtest (ANS acuity standardized β = -.18, p = .13) suggesting that this relation is not specific to math skills but holds true when general academic performance is used as an outcome measure.

For adults, ANS acuity was a significant predictor of a subject's accuracy on the fraction magnitude comparison test (standardized β = -.32, p = .004) and of accuracy on the non-symbolic division task (standardized β = -.51, p < .001). ANS acuity was no longer a significant predictor of accuracy on the fraction magnitude test after controlling for the mediator, non-symbolic division accuracy (ANS acuity standardized β = -.16, p = .18; non-symbolic division accuracy standardized β = .31, p =.01). Non-symbolic division accuracy fully mediated the relation between ANS acuity and accuracy on the fraction magnitude comparison test. The indirect effect was significant when tested with a bootstrap estimation approach with 5000 simulations (indirect effect = -.16, 95% CI = [-.31 -.04], p =.01). The direct effect was not significant, indicating a full mediation (direct effect = -.16, 95% CI = [-.43 .09], p =.20). The proportion mediated was .49 (p = .01, 95% CI = [.10 1.7]). Thus, higher fraction magnitude comparison accuracy was associated with .16 standard deviations sharper ANS acuity as mediated through non-symbolic division accuracy.

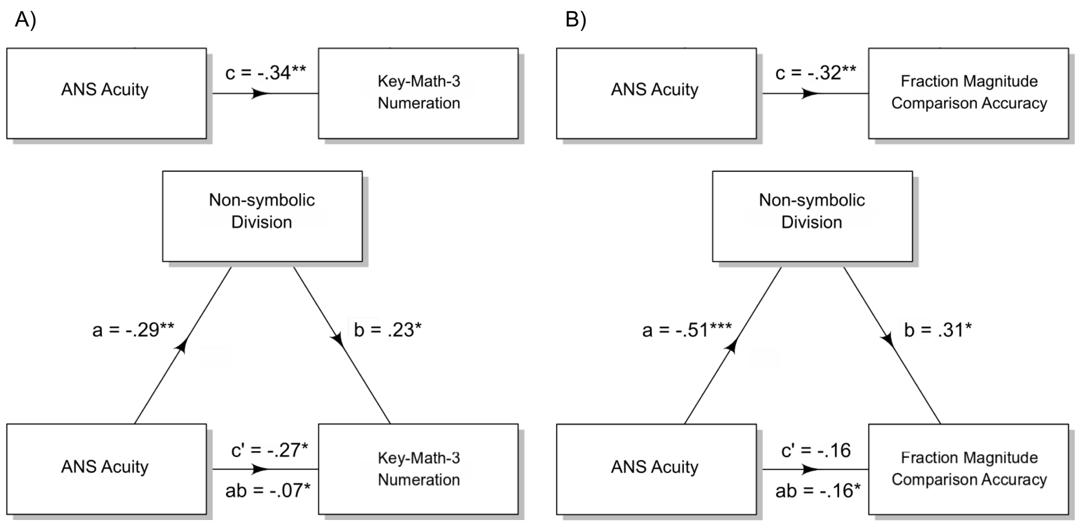

Figure 12. Mediation analyses test for a significant indirect effect (the product of the standardized coefficients a and b) that accounts for some portion of the original direct effect (c). The remaining direct effect is represented as c'. The models in this figure test whether non-symbolic division performance mediates the relation between ANS acuity and a measure of formal math skills in children (Key-Math-3 Numeration) and adults (Fraction Magnitude Comparison). A) Non-symbolic division accuracy partially mediates the relation between ANS acuity and a child's score on the Key-Math-3 Numeration section. Both the indirect (ab) and the direct path c' are significant. B) Non-symbolic division accuracy fully mediates the relation between ANS acuity and accuracy on the fraction magnitude comparison test. The remaining direct effect (c') is no longer significant, while the indirect effect (ab) is significant as tested with a bootstrap estimate approach.

Discussion

The current experiments are the first to demonstrate non-symbolic, approximate division ability in elementary school children and adults. Division requires participants to integrate the relations between a dividend, divisor, and quotient. Our results indicate that children can indeed perform a true non-symbolic, approximate division computation, and moreover, that they can use this informal understanding of division in concert with their growing knowledge of symbolic number to solve an approximate *symbolic* division task. Successful completion of the two division tasks was not dependent on formal knowledge of division. Children who could not recognize the division symbol nor solve simple division problems were nevertheless successful at performing non-symbolic division, and more surprisingly, were also able to complete the division task when the divided and target comparison number were represented symbolically with Arabic numerals.

This finding highlights the depth of intuitive math knowledge that children possess before formal education.

Previous research demonstrated that children can perform a scaling operation of halving or quartering (Barth et al., 2009; McCrink et al., 2016; McCrink & Spelke, 2016). Our study confirms children's ability to divide by 2 and 4. We extend these findings to conclude that children can flexibly switch between divisors from 2 to 8 from trial to trial to perform a true division operation. Children do not need to associate a symbol with one divisor to successfully divide a quantity but can perform a division operation integrating both the divisor and the dividend. We found that both children and adults were especially good at halving intuitively, and that there was a linear effect of divisor such that smaller divisors were easier than larger ones. This problem size effect is also found during exact, symbolic division calculation, suggesting a similar cognitive mechanism for non-symbolic and symbolic division (Mauro et al., 2003). While there is a preference for small number division, importantly for the current experiment both children and adults were well above chance when dividing by all divisors 2-8 as seen in Figure 11.

Whereas adults were significantly better at the approximate symbolic division task, children were significantly better at the non-symbolic task. However, all subjects exhibited faster reaction times on the non-symbolic task compared to the symbolic division task. This pattern suggests that all subjects were able to use their intuitive number sense to quickly compute non-symbolic division, but that increased experience with numerical symbols allows for increased computational accuracy. The point in development (or math education) where symbols facilitate more accurate arithmetic calculations may mark an important conceptual milestone in mathematical development. Future research can test whether the timing of this transition is predictive of later math achievement. One possibility is that making a switch to more accurate computation within the symbolic number system earlier in development is a better scaffold for increasingly complex computation. Alternatively, perhaps continuing to root a mathematical operation in its underlying concrete representation is a better foundation for understanding complex math concepts. These competing hypotheses should be explored in future work.

The novel nature of the approximate division tasks prompts an exploration of the strategies participants used to solve the approximate division problems. Importantly, participants did not use a heuristic to solve the approximate division tasks but did indeed integrate both the dividend and divisor to compute a quotient. Accuracy on the division tasks indicated that children and adults did not simply compare the dividend or the divisor to the target quantity without performing a division calculation. Experiment 2 demonstrated that children did not rely on a mean target strategy whereby they only considered the target value relative to the mean target value displayed over the course of the task. Instead of using an alternative heuristic, both child and adult participants engaged in an approximate division operation. Accuracy for all subjects was modulated by the ratio between the target and quotient in both non-symbolic and symbolic format, indicating use of an approximate strategy when making their choice. This result is somewhat surprising among adult subjects on the symbolic division task. University undergraduates have access to exact calculation strategies, especially when told the exact quantity of the dividend. Our analysis of the strategy questionnaire given to adult subjects corroborates the use of approximate calculation strategies. A majority of adult participants reported using an approximate strategy on the non-symbolic task, and only a slight majority of participants reported using an exact calculation strategy on the symbolic version of the task. Even among the subset of participants who reported using the strategy "*I calculated the exact answer using the number up top and the exact number of petals*" only about half could name the number of petals they actually saw. Overall accuracy on the division tasks, the significant effect of the ratio between the quotient and the target on accuracy, and the strategies reported by adults suggests that participants used an ANS based approximate division strategy to solve both formats of the task.

The third goal of these experiments was to examine the relation between ANS acuity, non-symbolic division, and symbolic math. In line with our hypothesis that non-symbolic calculation is a mechanism of the relation between ANS acuity and math, non-symbolic division mediated the relation between ANS acuity and symbolic math in both children and adults. Sharper ANS acuity may facilitate greater accuracy in a student's conceptual model of a division operation, and this conceptual model may function as a scaffold for formal symbolic computation. Thus, the mechanism for the established link between ANS acuity and symbolic math ability may be rooted in the computational abilities allowed by the ANS, and not in the acuity of the ANS per se. In this way ANS acuity may be indirectly linked to symbolic math by supporting access to mental models of computation. Having a strong mental model of what it means to divide (or engage in other operations such as subtraction or multiplication) may in turn create a strong foundation for the learning of abstract mathematical concepts. The significant mediation effect in adults suggests that adults can continue to use approximate mental models to calculate, even once they have access to exact calculation techniques. However, in children this mediation effect was not specific to the Key-Math-3 Numeration subtest. When children's scores on the Woodcock-Johnson Reading Cluster were partialed out of the model, ANS acuity was no longer correlated with scores on the Key-Math-3 Numeration subtest. This finding suggests that the correlation between ANS acuity and the Numeration test was driven by variance related to general academic skill, and not math skill specifically. This may be due to the very strong correlation between math and reading skills typical in children of this age because of extrinsic academic factors, and not because ANS acuity is meaningfully related to reading skill (Cantin, Gnaedinger, Gallaway, Hesson-McInnis, & Hund, 2016; Wang et al., 2015).

Future work should examine whether practice with approximate division computation can benefit formal arithmetic skills in elementary school children. The fact that children can perform non-symbolic division before learning division in school suggests arithmetic concepts can be introduced at an earlier age than previously thought. The theoretical framework of concreteness fading may be a particularly useful method for implementing such an intervention (Fyfe, McNeil, & Borjas, 2015; Fyfe & Nathan, 2018). A progression from practice with approximate non-symbolic division, to approximate symbolic division to exact symbolic division may be a way to link children's intuitions about division to formal division knowledge. Grounding abstract arithmetic concepts in children's intuitive understanding of arithmetic may boost not only children's conceptual understanding of arithmetic operations, but also their confidence that they have the skills to perform such calculations. Future work should explore whether incorporating approximate division early in math education is beneficial for later arithmetic knowledge.

Overall, the current work highlights the intuitive math abilities children bring to formal math education. Children can successfully compute an approximate division operation in both non-symbolic and symbolic format without formal knowledge of division. Children are remarkably good at dividing large numbers of objects, and this ability is not limited to simply halving or quartering. For both children and adults, non-symbolic division accuracy mediates the relation between their ANS acuity and symbolic math performance, suggesting non-symbolic calculation is a mechanism of the relation between ANS acuity and symbolic math. Children's extraordinary success at approximate division with large quantities suggests that introducing non-symbolic arithmetic calculation early in math education may be beneficial for formal arithmetic learning. Math learning that emphasizes how children's intuitive logic connects to highly abstract mathematical concepts may allow young math learners to utilize their own understanding of the world in a mathematical context.

CHAPTER 3: APPROXIMATE ARITHMETIC TRAINING YIELDS NO BENEFITS FOR SYMBOLIC ARITHMETIC FLUENCY IN ADULTS

Introduction

There is abundant evidence that training a particular cognitive skill improves performance on that very skill. However, it is less clear whether cognitive training benefits skills that are not directly trained (Sala & Gobet, 2017). Therefore, it is important to understand the conditions under which cognitive training promotes transfer to relevant skills, and to test the robustness of these training effects. One domain of cognitive training concerns the Approximate Number System (ANS). The ANS refers to the ability of humans and non-human animals to represent and manipulate non-symbolic, approximate quantities (Feigenson et al., 2004). It is hypothesized that training the ANS has the potential to improve the symbolic math abilities of both children and adults. There are multiple strands of evidence linking ANS acuity and symbolic math skills. ANS acuity is correlated with a variety of symbolic math skills across the lifespan (Chen & Li, 2014; Fazio et al., 2014; Schneider et al., 2016). ANS acuity is also longitudinally predictive of later math abilities (Y. He et al., 2016; Purpura & Logan, 2015; Soto-Calvo, Simmons, Willis, & Adams, 2015; Toll, Van Viersen, Kroesbergen, & Van Luit, 2015), and there is evidence that this relationship is bidirectional (Purpura & Simms, 2018; Shusterman et al., 2016). There is some evidence that children with a math specific learning disorder have a deficit in ANS acuity (Desoete, Ceulemans, De Weerdt, & Pieters, 2012; Olsson, Östergren, & Träff, 2016; Piazza et al., 2010). Thus, there is a large literature linking ANS acuity and mathematics, however, the majority of this evidence is correlational in nature. An ANS cognitive training paradigm that can improve math ability would provide causal evidence of the relation between the ANS and symbolic math.

Non-symbolic, approximate arithmetic training is an ANS cognitive training paradigm that has improved mathematical skill in both children (Hyde, Khanum, & Spelke, 2014; Khanum, Hanif, Spelke, Berteletti, & Hyde, 2016; Park, Bermudez, Roberts, & Brannon, 2016; Szkudlarek & Brannon, 2018) and adults (Au, Jaeggi, & Buschkuehl, 2018; Park & Brannon, 2013; Park &

Brannon, 2014). In the training paradigm used with adults, subjects watch large quantities of objects move behind or out from an occluder in an addition or subtraction operation. Then, subjects respond in one of two ways. In a matching trial subjects indicate which of two arrays of objects matches their imagined or sum or difference behind the occluder. In a comparison trial subjects indicate whether the quantity of a new array of objects is larger or smaller than their imagined sum or difference. As subjects progress through training difficulty is titrated by varying the ratio between the correct sum or difference and the number of objects in the comparison array. This training paradigm is non-symbolic because the arithmetic calculation is done across concrete depictions of number, and it is approximate because subjects do not need to provide an exact answer to the animated calculation, but instead compare their imagined solution to another quantity.

Specifically, training with non-symbolic, approximate addition and subtraction has yielded benefits in symbolic, exact arithmetic calculation in adult participants. In the original paper to discover this effect, Park & Brannon (2013) demonstrated over the course of two experiments that adult subjects who trained with non-symbolic, approximate arithmetic answered more double and triple digit addition and subtraction problems correctly at post-test compared to pretest than subjects who trained with a numeral ordering task, a knowledge training task, or a no contact control group. The training in both experiments was relatively brief, lasting for six or ten 25-minute sessions in total. A subsequent experiment by the same authors (Park & Brannon, 2014) replicated this effect, and found that non-symbolic, approximate arithmetic training was more effective at improving symbolic arithmetic fluency than training with a non-symbolic number comparison task, a visuo-spatial short-term memory task, or numeral ordering training task. A separate research group also replicated this effect with the finding that subjects trained on approximate arithmetic improved more on a symbolic arithmetic fluency test compared to subjects who spent the same amount of time answering general knowledge multiple choice questions (Au et al., 2018). Taken together, this work suggests that even for adult subjects skilled in symbolic addition and subtraction calculation, practice with non-symbolic, approximate arithmetic can increase the ability to solve symbolic addition and subtraction problems. These findings are interpreted as evidence for a causal relation between the Approximate Number System and symbolic mathematics.

As already described, the approximate arithmetic effect is one aspect of a larger literature that explores the relation between basic primitive, perceptual representations of number and symbolic math abilities. It is important to differentiate the evidence linking different aspects of the ANS and their relation to symbolic math skills. The finding that approximate, non-symbolic arithmetic training improves symbolic arithmetic skill appears to be more support for the relation between ANS acuity and symbolic math, however, there are important differences between the approximate arithmetic task and tasks used to measure ANS acuity. ANS acuity is typically measured with a dot comparison task. In this task, participants see two arrays of dots flash briefly on a screen, and then the participant identifies which of the two arrays was greater in quantity. The experimenter varies the ratio between the two quantities displayed, and models the ratio closest to one that a participant can reliably distinguish to get a measure of a participant's ANS acuity. Approximate arithmetic uses skills beyond what is required during a dot comparison task, including working memory, visual manipulation, and the concept of addition and subtraction. There is some evidence that these extra skills, beyond the practice with large number approximation, is the mechanism of the transfer to symbolic arithmetic. Adults who trained with approximate arithmetic improved significantly more on a symbolic arithmetic fluency task than subjects who trained on dot comparison alone (Park & Brannon, 2014). Moreover, there is no evidence that approximate arithmetic training improves ANS acuity (Au, Jaeggi, & Buschkuehl, 2018; Park & Brannon, 2014). This suggests that the mechanism of the approximate arithmetic training effect is not sharper ANS acuity, but a different aspect of the non-symbolic, approximate arithmetic task. Therefore, improved performance on symbolic math tasks after practice with non-symbolic, approximate arithmetic does not inform the relation between ANS acuity and symbolic math, but instead is one example of the way in which ANS representations can be used in a mathematical context to improve symbolic math skills.

The current study aims to further investigate the necessary parameters and the robustness of transfer from approximate arithmetic training to symbolic arithmetic calculation. We test the length of training necessary, potential mechanisms, and the effect size of the non-symbolic, approximate arithmetic training effect in college age adults across four independent experiments. Using the training and testing paradigms from the original experiments (Park & Brannon, 2013; Park & Brannon, 2014) we attempted four independent extensions and replications of the original effect. Experiment 1 attempts to determine if the approximate arithmetic training effect exists with a shorter training period than originally reported (2 vs 6 days of training). Experiment 2 compares approximate arithmetic training to a training condition matched in terms of the working memory load and numerical quantities manipulated in approximate arithmetic, but without addition or subtraction operations. Experiment 3 is an attempted replication of the approximate arithmetic effect found in Park & Brannon 2013, 2014. In this experiment the effectiveness of non-symbolic, approximate arithmetic training is compared to numeral ordering training at improving arithmetic fluency. Experiment 4 is another attempted replication of the approximate arithmetic effect, this time comparing the improvements in arithmetic fluency after approximate arithmetic or approximate number comparison training. Finally, to obtain the best power possible at detecting the transfer of non-symbolic, approximate arithmetic training to symbolic arithmetic fluency, we analyze the data from Experiments 1-4 combined with the original data from the Park & Brannon experiments to determine whether or not approximate arithmetic training is more effective at improving arithmetic fluency in adult subjects than a variety of numerical and non-numerical training conditions. Our results highlight the need for large-scale replications of established cognitive training paradigms.

Materials and Methods

Table 4. Methods for the original Park & Brannon 2013, 2014 experiments and the four experiments reported in Chapter 3.

Dataset	Training conditions	Pre and post tests	Days of Training
Park & Brannon 2013 Experiment 1	Approximate arithmetic No contact control	--Exact symbolic arithmetic --Vocabulary	10
Park & Brannon 2013 Experiment 2	Approximate arithmetic Numerical symbol ordering Knowledge training	--Exact symbolic arithmetic --Vocabulary --Numeral order judgement	6
Park & Brannon 2014 Experiment 1	Approximate arithmetic Approximate number comparison Visuo-spatial short term memory Numerical symbol ordering	--Exact symbolic arithmetic --Vocabulary --Non-symbolic numerical comparison --Spatial 2-back test --Numeral order judgement	6
Experiment 1	Approximate arithmetic Numerical symbol ordering	--Exact symbolic arithmetic --Numeral order judgement -- Symbolic addition verification	2
Experiment 2	Approximate arithmetic Numerical symbol ordering Approximate range	--Exact symbolic arithmetic --Non-symbolic numerical comparison --addition verification (post-test only) --mental rotation --Numeral order judgement --rhyming test	6
Experiment 3	Approximate arithmetic Numerical symbol ordering	--Exact symbolic arithmetic --vocabulary --Non-symbolic numerical comparison --Spatial 2-back test --Numeral order judgement	6
Experiment 4	Approximate arithmetic Approximate number comparison	--Exact symbolic arithmetic --vocabulary --Non-symbolic numerical comparison -- Spatial 2-back test --Numeral order judgement	6

Table 5. Subject demographics of the original Park & Brannon 2013, 2014 experiments and the four experiments reported in Chapter 3. AA = approximate arithmetic, NO = numeral symbol ordering, ANC = approximate number comparison, AR = approximate range, STM = visuo-spatial short term memory, KT = knowledge training, NC = no contact.

Dataset	Mean days between pre and post test	Mean age of subjects (range)	Subject gender	Number of subjects
Park & Brannon 2013 Experiment 1	AA 10.9	22.4 (18.8 - 31.4)	9 M,17 F	26
	NC 11.3	22.9 (18.6 – 33.4)	6 M 20 F	26
Park & Brannon 2013, Experiment 2	AA 8.9	20.9 (18.7 - 23.8)	3 M 13 F	16
	NO 9.3	22.85 (18.8 - 31.94)	5 M 9 F	14
	KT 9.1	22.17 (19 – 26.9)	6 M 10 F	15[9]
Park & Brannon 2014 Experiment 1	AA 9.2	21.4 (18.1 – 26.6)	7 M 11 F	18
	NO 9.1	21.4 (18.1 – 26.6)	6 M 11 F	17
	ANC 9.2	21.9 (18.6 – 31.2)	7 M 11 F	18
	STM 9.3	22.6 (18.1 – 34.3)	8 M 9 F	17
Experiment 1	AA 1.47	20.6 (18.1 - 24.9)	7 M 12 F	19
	NO 1.58	20.0 (18.6 - 22.4)	10 M 9 F	19
Experiment 2	AA 9.15	21.83 (18.8 - 27.8)	9 M 17 F 1 unreported	27
	NO 9.15	21.5 (18.5 – 34.1)	8 M 16 F 3 unreported	27
	AR 9.58	21.3 (18.7 – 30.4)	9 M 15 F	24
Experiment 3	AA 10.65	23.09 (18.4 - 31.9)	13 M 34 F 1 unreported	48
	NO 10.33	24.63 (18.2 - 34.7)	19 M 24 F	43
Experiment 4	AA 10.22	22.4 (18.0 – 30.7)	16 M 37 F 2 unreported	55
	ANC 10.18	22.6 (18.0 – 30.6)	13 M 36 F	56

Subjects. Participants in Experiment 1 were recruited through the University of Pennsylvania’s Psychology Subject pool and received course credit for participation. Participants in Experiments 2 – 4 were recruited with flyers distributed throughout the University of Pennsylvania’s campus. These flyers advertised for participation in a “Brain Exercise” psychology experiment to study

[9] One subject was removed from the knowledge training condition of Park & Brannon 2013 Experiment 2 due to the miscoding of this participant’s training condition.

adult cognition. Subjects in Experiments 2 and 3 were paid in a lump sum at the completion of the post-test. Subjects in Experiment 4 were paid after each session. Flyers contained the same language as the recruitment flyers used in (Park & Brannon, 2013; Park & Brannon, 2014). Participants recruited through flyers were largely students at the University of Pennsylvania. There were 38 subjects in Experiment 1, 78 subjects in Experiment 2, 91 subjects in Experiment 3, and 111 subjects in Experiment 4. Age and gender by experiment and training condition are reported in Table 5. Subjects were required to speak English and be under the age of 35.

Procedure.

Participants were randomly assigned to conditions. In all Experiments, subjects first completed a pretest battery consisting of the exact symbolic arithmetic test, and various other tests (see Table 4 for details). Experiment 1 consisted of only two sessions. In the first session, subjects completed the pretest battery followed by the first 25-minute training session. In session 2 participants completed the second training session followed by the post-test battery. Experiments 2-4 each consisted of 8 sessions. Participants completed six 25-minute training sessions on consecutive weekdays (pretest, 6 training sessions, post-test). All testing and training sessions took place in a quiet testing room with six computers. The average number of days between pre and post test for participants in the approximate arithmetic training condition is reported in Table 5. The order in which subjects completed the pre and post tests was counterbalanced across participants.

Training Conditions

Approximate Arithmetic. This condition was identical in Experiments 1- 4 and also identical to the approximate arithmetic training condition used in (Park & Brannon, 2013; Park & Brannon, 2014). Subjects mentally added or subtracted dot arrays ranging from 9 to 36 as the arrays moved behind or out from an occluder in the center of the screen. Subjects were then required to either compare their imagined sum or difference to a new target array, or to match their imagined sum or difference to one of two target arrays. In Experiments 1-3 subjects responded by touching a touch screen monitor to indicate their choice. In Experiment 4, subjects used the mouse to indicate their response. Dot arrays that represented the addends in the problem were visible for 1000ms. Target dot arrays to be compared or matched were visible for 1500ms before they were hidden behind a black circle. Both the matching and comparison trial types, and the addition and subtraction trial types were intermixed within 10 blocks of 20 trials each for each training session. Feedback was provided after each trial. The difficulty of the task was titrated to performance by decreasing the ratio between the imagined sum or difference and the target dot array. The numerical distance between the sum or difference and the target array varied in a log-base2 scale, the log difference level. All participants started training with a log difference level of 1.5. A log difference level of 1.5 is equivalent to a ratio of 2.83 (2^1.5 to 1) between the sum or difference and the alternative target array. For each 20-trial block, if performance was greater than 85% the log difference level decreased by one of the following randomly chosen values [.13, .14, .15, .16, .17]. If performance over a block of 20 trials was less than 70% the log difference level increased by one of the values randomly chosen from the following set [.08, .09, .10, .11, .12]. The log difference level achieved at the end of a training session was carried over into the next session.

Numeral Symbol Ordering. This training condition was used in Experiments 1-3, and is identical to the Number Symbol Ordering training condition used in (Park & Brannon, 2013; Park & Brannon, 2014). Subjects were required to order sets of three Arabic numerals before they moved off of the screen. A maximum of three triads appeared on the screen at the same time. If the triad was moving to the left on the screen the numerals needed to be in ascending order. If the triad was moving to the right on the screen the numerals needed to be in descending order. To reorder the triad a subject touched the triad of numerals, which rearranged themselves randomly with each touch. Subjects received feedback on whether the triad was in the right order as it entered a gray block at the edge of the screen. This grey block turned green if the triad was in the correct order, or red if the triad was in the incorrect order. Task difficulty was titrated by varying the speed in which the triads travelled across the screen. Triad speed started at 125 pixels/sec. If accuracy was greater than 90% over a 2.2 minute span, the speed increased by one of the values chosen randomly from the following [10,11,12,13,14]. If accuracy was less than 80%, the speed decreased by one of the following values [4, 5, 6, 7, 8]. The speed at the end of one training session was carried over into the next session.

Approximate Number Comparison. This training condition was used in Experiment 4, and is identical to the Approximate Number comparison training condition used in Park & Brannon (2014). The task was to identify which of two dot arrays contained the greater number of dots. There were two trial types. In the mixed trial type, white and black dots appeared in an intermixed array on a grey background for 750ms. Participants reported whether there were more black or white dots. In the other trial type participants saw two distinct black or white dot arrays on the screen at the same time, and chose which array was greater in numerosity. Participants responded with a mouse click. As in the approximate arithmetic training condition, difficulty was titrated using the log difference level. One of the dot arrays on each trial ranged from 16 to 32, and the other was determined by the log difference level. For example, if one array contained 16 dots, and the log difference level was 1.15, then the other array would contain either 16 X 2^1.15 or 16/2^1.15 dots. The titration procedure was the same as used in the approximate arithmetic training condition.

Approximate Range. This training condition was used in Experiment 2. It was designed to match the approximate arithmetic training condition in terms of working memory load and the manipulation of dot arrays without the arithmetic component. Subjects in this training condition were required to indicate whether or not the number of dots in a target dot array fell inside or outside the range of two previously viewed dot arrays. Subjects watched as one dot array appeared in the middle of the screen for one second before it moved behind either an occluder (a gray box) on the right or an occluder (an identical gray box) on the left, counterbalanced. A second dot array appeared for one second and moved behind the second occluder. On comparison trials, a target array appeared at the bottom of the screen for 1000ms and participants indicated by touching the screen whether or not this target was within the range of the previous two arrays. The target array was hidden by a black circle after 1000ms, but subjects had 3500ms more to respond to match the approximate arithmetic training condition. For example, if the first array contained 30 dots and the second array contained 120 dots and the target array had 15 dots the correct choice was “outside” the range. On matching trials, two arrays appeared at the bottom of the screen and subjects chose the array that was within the range of the previous two animated arrays. For example, if the first array contained 30 dots and the second array contained 120 dots the correct choice would be 90 rather than 180. The number of dots in one array ranged from 4 to 256. The two arrays that defined a range always had a ratio of 1:4. The difficulty of this task was titrated to match the approximate arithmetic training condition, by changing the ratio between one of the range defining arrays and the target dot array. The numerical distance between one of the

range defining arrays and the target array varied in a log-base2 scale, the log difference level. All participants started training with a log difference level of 1. A log difference level of 1 is equivalent to a ratio of 2 (2^1 to 1) between one of the range defining arrays and the alternative target array. If performance over a block of 20 trials was greater than 85% the log difference level decreased by .10. If performance over a block of 20 trials was less than 70% the log difference level increased by .05. The log difference level achieved at the end of a training session was carried over into the next session.

Pre and Post Tests

Exact Symbolic Arithmetic Test. The test of arithmetic fluency developed by Park & Brannon, (2013; 2014) was used in all four experiments. Subjects solved two and three digit addition and subtraction problems over two five minutes blocks. The operands of the problems ranged from 11 to 244. Problems were chosen randomly for each participant from a set of 800 potential problems for the pretest and a distinct set of 800 for the posttest (counterbalanced). The number of problems that required carrying and borrowing was matched for the two problem sets. Performance was quantified as the total number of correct problems solved over the 10 minute assessment. The analyses in the current paper focus on this pre and post test measure.

Numeral Order Judgement. This test was the same as used in Park & Brannon 2013 Experiment 2 and Park & Brannon 2014 Experiment 1, and was used in Experiments 1, 2, 3 and 4. Subjects saw a triad of Arabic numerals and had to indicate whether or not the triad was in the correct order or not. The numerals 1-9 were used. Trials were presented in ascending or descending order. The triads were presented for 1500ms without feedback. There were two blocks of 64 trials each. Performance was quantified as the mean accuracy and the median reaction time across correct trials.

Arithmetic Verification. This test was used in Experiment 1, and as a post-test only in Experiment 2. In this task, 1 and 2-digit addition and subtraction problems are displayed horizontally with a proposed answer (e.g., 27 + 52 = 79). The participant must press one key if the statement is correct and another if it is incorrect. On incorrect trials (50% of all trials) the sum displayed will be +/- 10 from the correct sum or +/- 2 from the correct sum. The problems are displayed for 10 seconds to allow subjects as much time as possible to respond. The stimuli were taken from (Klein et al., 2010). Performance was quantified as the median reaction time on trials where the sum was correct and the subject responded correctly following the analysis in (Klein et al., 2010).

Dot Comparison. This test was used in Experiment 2, 3. Subjects saw two white dot arrays presented on a black screen. The dot arrays are visible for 300ms. Subjects then indicate which dot array had a greater number of dots by pressing the arrow keys on the keyboard. There was a 2 second inter-trial interval after each response. The stimuli were created using the same technique as reported in (DeWind et al., 2015). Subjects completed 400 trials total with a break after 200 trials. Performance was quantified with the *Weber fraction* as calculated in (DeWind et al., 2015). A different version of the dot comparison test was used in Experiment 4 and Park & Brannon 2014 Experiment. In this version of the task, two white dot arrays were presented on a black screen for 750ms. Subjects indicated which dot array was greater in number. Subjects completed 264 trials. Performance was quantified with the *Weber fraction* as calculated in (Pica et al., 2004).

Mental Rotation. This test was used in Experiment 2. Subjects saw one criterion shape made up of blocks stacked together on the left side of the screen. On the right side of the screen were four choice shapes. Subjects indicated using the keys 1-4 which two of the four choice images are the same as the criterion shape but displayed at different angles. Subjects' choices were surrounded with as red box when indicated. Subjects completed two blocks of 10 problems each. Each block timed out after 3 minutes. Performance was quantified as the number of trials correct.

Rhyming test. This test was used in Experiment 2. Subjects saw a word at the top of the screen and a letter with dashes after it below. The dashes represented missing letters of a word that rhymed with the word above. For example subjects could see the word 'hat' and 'c - -' written below. Subjects should then type in the word 'cat' to solve the trial. The letters of the word a subject typed appeared below the displayed word. Subjects pressed 'enter' when they had finished typing their word. Subjects completed two blocks of five minutes each. Performance was quantified as the number of correct rhymes generated in the 10 minutes allowed for the task.

Spatial 2-Back. This is the same test as used in Park & Brannon 2014, and was used in Experiments 3 and 4. Subjects watched as a white circle appeared on a black screen in a particular location. Subjects has to remember this location and compare it to the location of the circle two back in the sequence. Subjects completed a total of 168 trials over two blocks. Performance was quantified as the z score of the false alarm rate subtracted from the z score of the hit rate.

Vocabulary test. This is the same test as used in Park & Brannon 2014, and was used in Experiments 3 and 4. Subjects has 5 minutes to answer 42 multiple choice questions taken from the Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976). Performance was quantified as the number of correctly answered questions minus ¼ of the problems marked incorrectly.

Analysis Plan.

Gain scores for each participant were calculated for the exact symbolic arithmetic assessment by subtracting the number of problems solved correctly in 10 minutes at post-test from the number of problems solved correctly in 10 minutes at pretest. In order to remove outlying patterns, any arithmetic gain score that was smaller than Q1 - 2 x IQR or larger than Q3 + 2 x IQR was removed, where Q1 is the first quartile, Q3 is the third quartile, and IQR is the interquartile range. In this analysis, quartiles were calculated with the data from each experiment separately. We then conducted a two-sample t-test or a one-way ANOVA to examine whether there was a significant difference between the average arithmetic fluency gain score by training condition. We also report gain scores for the other pre/post assessments using the same analytical technique. This analysis was preregistered for Experiment 4 with asPredicted.org. As a complement to the frequentist analysis of the training effect, we also report a Bayesian analysis of this effect for each experiment to examine the relative support for both our hypothesis of interest and the null hypothesis. We conducted a Bayesian t-test or ANOVA, dependent on the number of training conditions for each experiment. We set a non-informative Jeffreys prior width of .5 to correspond to a small effect (Lakens, 2015(Morey & Rouder, 2011)). These analyses result in a Bayes factor (BF_{10}), which can be interpreted as the likelihood ratio for the alternative hypothesis over the null. Given that the Bayes factor (BF_{10}) is a ratio of the likelihood for the alternative hypothesis over the null hypothesis, the inverse of the Bayes factor (BF_{01}) can be interpreted as the likelihood ratio for evidence of the null hypothesis over the alternative hypothesis. Following Jeffreys (1961) we use the following designations to interpret the strength of the Bayes factors: 0–3 offer anecdotal support for H1, 3–10 moderate support for the H1, 10 – 30 strong support for H1, 30 -100 very strong evidence for H1, and values greater than 100 offer decisive evidence for H1. To facilitate comparison with the present data we conducted a Bayesian analysis of Park and Brannon's (2013, 2014) previously reported data broken down by experiment.

Finally, we combined the data from all experiments, including the data from (Park & Brannon, 2013; Park & Brannon, 2014) to test the hypothesis that approximate arithmetic training improves exact symbolic arithmetic fluency more than any of the alternative training conditions: approximate number comparison, visuo-spatial short term memory, numerical symbol ordering, approximate range, knowledge training, and a no contact control. We again use the complementary frequentist and Bayesian approaches. To match the analytical technique used with each individual experiment, we first report a one way ANOVA with training condition as a factor and exact symbolic arithmetic gain score as the outcome. Then, we report a one way Bayesian ANOVA testing whether condition as a factor adds significant variance over the model with the mean intercept only. This analysis tests whether any of the training conditions results in differences in arithmetic fluency gain, however, we have a specific hypothesis that approximate arithmetic training improves arithmetic fluency more than any of the other training conditions. To examine this specific hypothesis, we also report a contrast between the average arithmetic fluency gain score for the approximate arithmetic condition and all other conditions. This analysis examines whether approximate arithmetic training results in an average arithmetic fluency gain score that is greater than every other individual training task. We also follow up this analysis with planned one sided t-tests between the average gain score for participants in the approximate arithmetic training condition and each other training condition. Finally, we compared the effect size of the arithmetic fluency gain found in the non-symbolic, approximate arithmetic condition across experiments to test the robustness of the transfer effect regardless of the control conditions.

Results

Analysis of Approximate Arithmetic Training Performance. To quantify training performance, we calculated the mean log difference score of the matching and comparison trial types at the end of each training session for each participant in the approximate arithmetic training condition to match the analysis in the original Park and Brannon experiments. A one-way ANOVA indicated there was no significant difference between the last log difference level reached on training day 6 by experiment across the three Park and Brannon studies and current experiments 2-4 (Figure 13; $F_{5,181} = 1.48$, $p = .20$). There were also no significant differences between mean log difference level by experiment on training session 2 for Experiments 1-4 and the three experiments conducted by Park and Brannon ($F_{6,201} = .800$, $p = .57$). These analyses suggest that motivation to complete the training task in the approximate arithmetic training condition was similar in the current experiments and the prior studies by Park and Brannon.

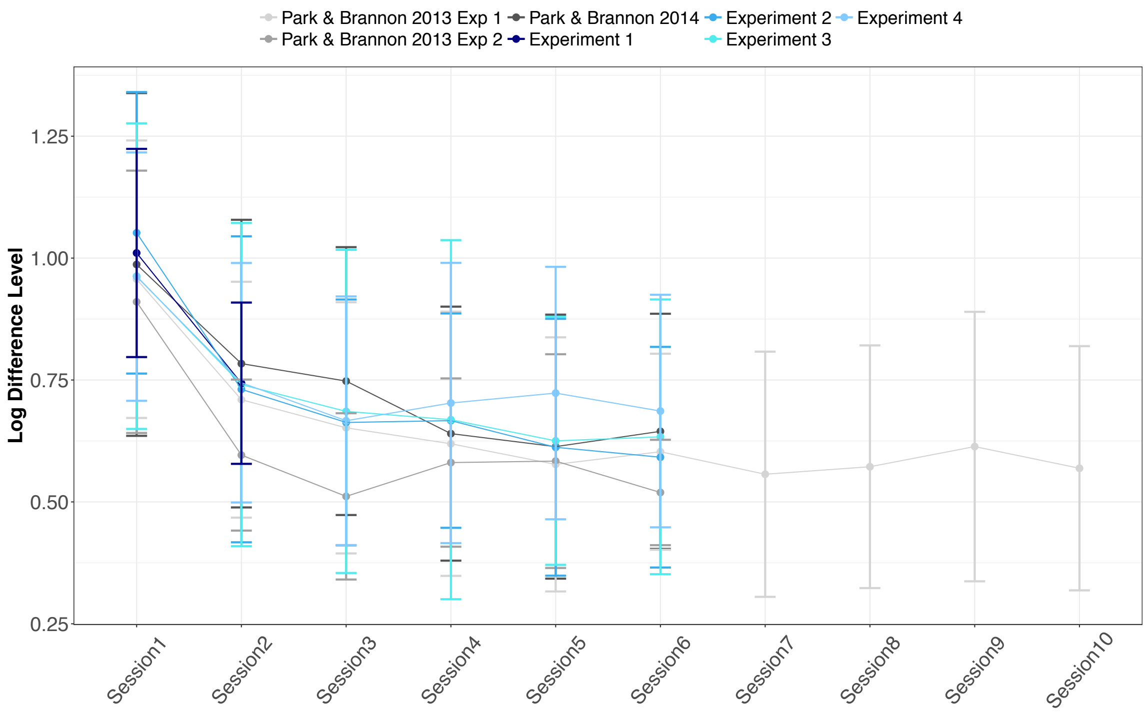

Figure 13. Training trajectories of the approximate arithmetic training condition by experiment. The blue lines represent Experiments 1-4. The grey lines represent the original experiments. The dots represent the mean log difference score at the end of each training session across participants. All participants started session 1 of training with a log difference level of 2 (1:4 ratio). The error bars represent the standard deviation. All experiments involved six days of training except for Experiment 1 (two days) and Park & Brannon 2013 Experiment 1 (ten days).

Experiment 1. A two-sample t-test indicated no significant differences in the number of arithmetic problems solved correctly at pretest by condition (Figure 14; $t_{36} = -1.20$, $p = .24$) suggesting that random assignment was effective. There was no significant difference between the average arithmetic gain score for the approximate arithmetic and numeral ordering training conditions (Figure 15; $t_{36} = -1.54$, $p = .13$). The complementary Bayesian t test indicated a BF_{10} = .98. A Bayes factor close to 1 suggests no evidence for either the alternative or the null hypothesis.

We also examined whether or not there was a significant difference in the gain scores of the numeral order judgment test or the arithmetic verification test by condition. Our outlier exclusion procedure removed one numeral ordering accuracy gain score, and one subject did not complete the arithmetic verification test. A t-test indicated no significant effect of condition on the change in accuracy (t_{35} = .953, p = .35) or RT (t_{36} = -1.28, p = .21) on the numeral order judgement test, or the RT on the arithmetic verification trials when the sum was correct and the subjects responded correctly (t_{35} = -0.30, p = .77).

Experiment 2. One participant in the approximate range condition was removed from the sample due to an outlier gain score. There were no significant differences in pretest score by condition (Figure 14; $F_{2,74}$ = .136, p = .87) again suggesting random assignment was effective. A one-way ANOVA with training condition as a factor indicated no differences in arithmetic gain score by training condition (Figure 15; $F_{2,74}$ = .257, p = .77). A Bayesian ANOVA resulted in a BF_{10} = .137, suggesting moderate evidence for the null hypothesis of no difference in arithmetic fluency gain score by condition.

We also examined whether or not there was a significant difference in the gain scores of the other pre and post tests. Two subjects did not complete the metal rotation post-test, one subject did not complete the rhyming post-test, and one subject's post-test numeral ordering data was not recorded accurately. Our outlier exclusion procedure removed 12 individual gain scores (7 numeral ordering accuracy gain scores, two weber fraction gain scores, one mental rotation gain score, and one rhyming gain score. One-way ANOVAs with condition as a factor indicated no significant difference between the change in accuracy ($F_{2,67}$ = .05, p = .95) on the numeral order judgement test, but indicated a significant difference in RT ($F_{2,74}$ = 4.72, p = .01). Pairwise t-tests indicated this difference was driven by a significant difference between the RT of the numeral ordering training condition and the approximate range condition (p = .009) when controlling for multiple comparisons with the Holm correction. One-way ANOVAs also indicated no significant difference in gain score by condition on the mental rotation test ($F_{2,72}$ = .62 p = .54), the rhyming test ($F_{2,73}$ = 1.4 p = .25), and the dot comparison test ($F_{2,73}$ = .54 p = .59). A one-way ANOVA predicting reaction time on the post-test only arithmetic verification test also indicated no significant difference by condition ($F_{2,75}$ = .86, p = .43).

Experiment 3. Two participants in the numeral ordering training condition were removed due to outlier gain scores. There was no significant difference in pretest arithmetic scores by condition (Figure 14; $t_{87} = .873$, $p = .38$). Again, there was no significant difference in arithmetic gain score by condition (Figure 15; $t_{87} = .873$, $p = .38$). The Bayesian t-test indicated a $BF_{10} = .60$, indicating anecdotal support for the null hypothesis that the difference between the mean arithmetic gain scores of each condition is zero.

We also examined whether or not there was a significant difference in the gain scores of the other pre and post tests by condition. One subject did not complete the numeral order judgment test post-test, three subjects did not complete both the pre and post test of the spatial 2-back, and two subjects did not complete both the pre and post test of the vocabulary test. Out outlier exclusion procedure removed 16 individual gain scores (four numeral ordering accuracy gain scores, six numeral ordering RT gain scores, one weber fraction gain score, and five spatial 2-Back gain scores). Two sample t-tests indicated no effect of condition on numeral ordering accuracy ($t_{84} = .072$, $p = .94$) or reaction time ($t_{82} = 1.24$, $p = .22$), or weber fraction ($t_{88} = -.260$, $p = .80$), or vocabulary ($t_{87} = 1.34$, $p = .18$), or spatial 2-Back ($t_{81} = -.657$, $p = .51$).

Experiment 4. There was a marginal difference between the arithmetic fluency scores at pretest by condition (Figure 14; $t_{109} = -1.92$, $p = .06$). This effect was driven by the pretest score of one subject in the approximate number comparison condition who scored over 6 standard deviations above the mean pretest score. With this outlier pretest score removed, there was no longer a marginal difference in pretest score by condition ($t_{108} = -1.66$, $p = .10$). However, this subject was not removed from our subsequent analyses because their gain score was within our outlier cutoffs. There was no significant difference between arithmetic gain score by training condition (Figure 15; $t_{109} = -1.35$, $p = .18$). The Bayesian t test indicated a $BF_{10} = .83$, indicating anecdotal support for the null hypothesis that the difference between the mean arithmetic gain scores of each condition is zero.

We also analyzed the effect of condition on the other pre and post tests. One subject did not complete the spatial 2-Back pretest, one subject did not complete the vocabulary post-test, one subject did not complete the dot comparison pretest, and one subject did not complete the dot comparison post-test. Additionally, there were four subjects for whom the model of Weber fraction used in Park & Brannon 2014 did not converge on either the post-test or pretest measure, and these subjects were excluded from the analysis. Our outlier exclusion procedure removed 14 individual gain scores (six numeral ordering accuracy gain scores, one numeral ordering RT gain scores, six weber fraction gain scores, and one spatial 2-Back gain score). Two sample t-tests indicated a significant effect of condition on numeral ordering accuracy ($t_{101} = -2.29$, $p = .02$) and numeral ordering RT ($t_{108} = -2.45$, $p = .02$), however, these effects were in opposite directions. Specifically, subjects in the approximate number comparison condition showed a significantly greater gain in numeral ordering accuracy as compared to the approximate arithmetic condition, but the approximate arithmetic condition showed a significantly greater decrease in RT than the approximate number comparison condition. Two sample t-tests indicated no significant effect of condition on weber fraction gain ($t_{95} = -.044$, $p = .97$), or vocabulary ($t_{108} = .815$, $p = .42$), or spatial 2-Back ($t_{107} = .742$, $p = .50$).

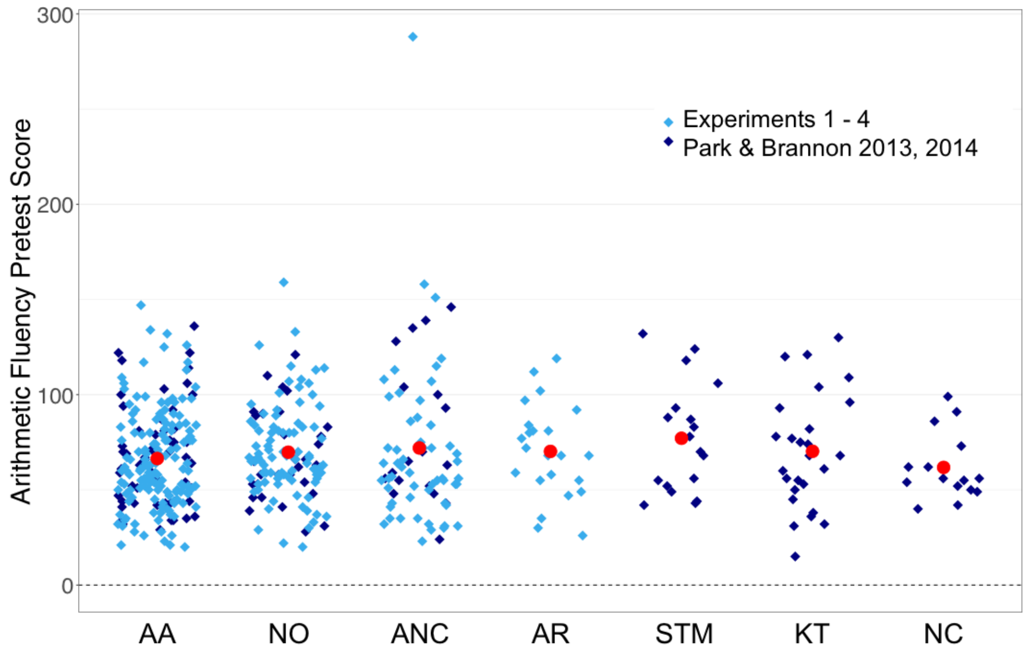

Figure 14. Pretest arithmetic fluency scores by condition and experiment. The dark blue points depict subjects from Experiments 1-4. The light blue data points depict data subjects from Park & Brannon 2013, 2014. The red points represent the mean pretest arithmetic fluency score by condition. There is no significant difference in pretest arithmetic fluency score by condition or experiment. AA = approximate arithmetic, NO = numeral symbol ordering, ANC = approximate number comparison, AR = approximate range, STM = visuo-spatial short term memory, KT = knowledge training, NC = no contact.

Bayesian Re-analysis of Park & Brannon 2013, 2014

Park & Brannon 2013 Experiment 1. A Bayesian t test yielded a BF_{10} = 3.27, suggesting moderate support for the alternative hypothesis of a significant difference between the arithmetic fluency gain scores for the approximate arithmetic and no contact control groups. This reanalysis is consistent with the conclusions reported in Park & Brannon 2013.

Park & Brannon 2013 Experiment 2. A Bayesian ANOVA indicated a BF_{10} = 1.7, indicating anecdotal evidence for the alternative hypothesis of a significant difference in arithmetic fluency gain score by training condition.

Park & Brannon 2014 Experiment 1. A Bayesian ANOVA indicated BF_{10} = 2.7, indicating anecdotal evidence for the alternative hypothesis of a significant difference in arithmetic fluency gain score by training condition.

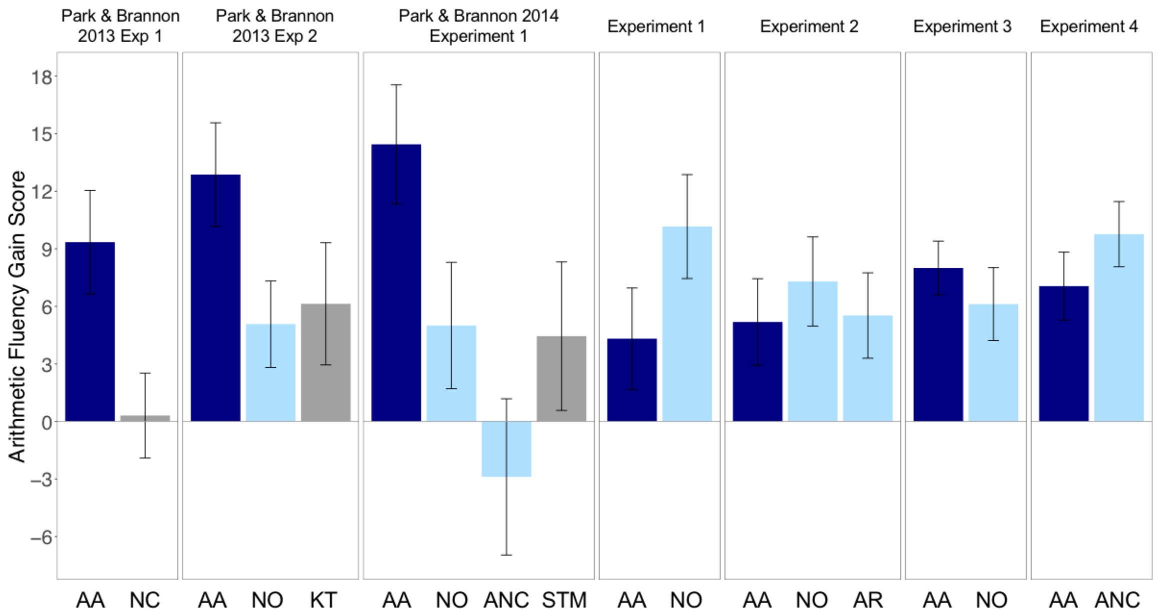

Figure 15. Bar plot of arithmetic fluency gain score by condition and experiment. AA = approximate arithmetic, NO = numeral symbol ordering, ANC = approximate number comparison, AR = approximate range, STM = visuo-spatial short term memory, KT = knowledge training, NC = no contact. The arithmetic fluency gain score is plotted in terms of the number of correct arithmetic questions (post-test minus pre-test). The bars colored in shades of blue are numerical training conditions. The grey bars represent non-numerical or no-contact control training conditions. The error bars represent standard error of the mean.

Combined Analysis. Combining all the data from the current study and the three previous experiments conducted by Park and Brannon (2013, 2014) created a dataset with 487 individual arithmetic fluency gain scores across seven training conditions. A new outlier analysis with the full data set resulted in five arithmetic gain score outliers across all seven experiments: one from Park & Brannon 2013 Experiment 2, two from Park & Brannon 2014 Experiment 1, one from Experiment 2, and one from Experiment 4. These subjects were excluded from the current analysis. A one-way ANOVA predicting pretest arithmetic fluency score by condition was not significant (Figure 14; $F_{6,475} = .855$, $p = .53$) indicating there were no significant differences in arithmetic fluency score by condition at pretest. Moreover, a one-way ANOVA predicting arithmetic fluency pretest score by experiment was also not significant (Figure 14; $F_{6,475} = 1.27$, $p = .27$) suggesting that the samples for each experiment had comparable initial arithmetic performance. Crucial to our main hypothesis, a one-way ANOVA predicting arithmetic fluency gain score with condition as a factor indicated no significant differences by condition (Figure 16; $F_{6,475}$ = 1.69, p = .12). The complementary Bayesian one-way ANOVA resulted in a BF_{10} = .14 for the condition factor. This is moderate evidence that the model with only the mean intercept is a better model of arithmetic gain score than a model with training condition as a factor. The current data is seven times (i.e., 1/0.14 = 7.14) more likely to occur under the null hypothesis that the intercept only model is a better model of the data than the alternative model that training condition explains variance.

To test the specific hypothesis that approximate arithmetic training improves arithmetic fluency more than any other training condition, we ran a contrast between the approximate arithmetic condition and all other conditions. This ANOVA indicated a marginally significant difference between the approximate condition and the other training conditions as a whole (Figure 16; $F_{1,480}$ = 3.93, p = .05, BF_{10} = .68). To follow up this analysis, we calculated a one sided t-test between the average gain score for the approximate arithmetic training condition and every other condition. There was no evidence that the average gain score for participants in the approximate arithmetic training condition was greater than the arithmetic fluency gain score of any other training condition (numeral ordering t_{325}= .987, p = .16, BF_{10} = .27; approximate number comparison t_{279}= .844, p = .20, BF_{10} = .28; approximate range t_{229}= .986, p = .16, BF_{10} = .44; visuo-spatial short term memory t_{224}= 1.20, p = .12, BF_{10} = .56; knowledge training t_{221}= .612, p = .27, BF_{10} = .41). However, there was evidence that approximate arithmetic training resulted in a greater arithmetic fluency gain score than the no contact control condition (t_{232}= .3.12, p = .001, BF_{10} = 16.0), suggesting that any significant difference between the approximate arithmetic training condition and the other experimental conditions was driven by this difference.

Finally, we compared the effect size of the gain scores within the non-symbolic, approximate arithmetic training condition across experiments. The original experiments reported effect sizes for the approximate arithmetic training condition of d = .68 (9.35 problems, Park & Brannon 2013 Experiment 1), d = 1.08 (15.4 problems, Park & Brannon 2013 Experiment 2), and d = 1.10 (14.4 problems, Park & Brannon 2014 Experiment 1). Experiments 2 – 4 of the present study used the same number of training sessions as Experiment 2 in Park & Brannon (2013) and Experiment 1 in Park & Brannon (2014). The effect sizes for Experiments 2-4 of the present study were d = .45. d = .82, and d = .54, corresponding to an increase of 5.19, 8.00, and 7.05 problems answered correctly respectively. A one-way ANOVA testing for a significant difference in arithmetic fluency gain score by experiment revealed a significant difference (Figure 15; $F_{4,159}$ = 3.10 p = .02). Pairwise tests revealed that the effect size for Experiments 2-4 were smaller than the effect size found in Park & Brannon 2013 Experiment 2 (Experiment 2 p = .01, Experiment 3 p = .04, Experiment 4 p = .02), and the effect sizes found in Experiments 2 and 4 were smaller than the effect size found in Park & Brannon 2014 (Experiment 2 p = .01, Experiment 4 p = .03). However, none of these pairwise comparisons survived the Holm correction for multiple comparisons.

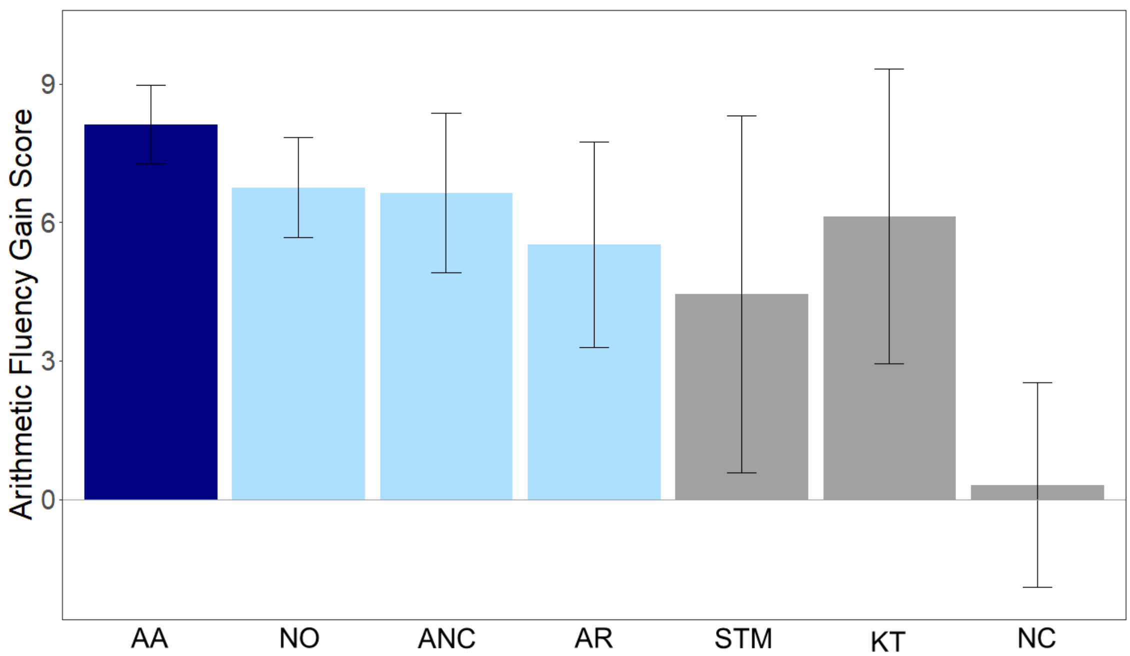

Figure 16. Arithmetic fluency gain score by condition collapsed across all experiments including the original Park and Brannon experiments. AA = approximate arithmetic, NO = numeral symbol ordering, ANC = approximate number comparison, AR = approximate range, STM = visuo-spatial short term memory, KT = knowledge training, NC = no contact. The gain score is plotted in terms of the number of arithmetic questions solved correctly (post-test minus pretest). The bars colored in shades of blue are numerical training conditions. The grey bars represent non-numerical or no-contact control training conditions. The error bars represent standard error of the mean.

Discussion

The original goal of the current study was to replicate the finding that non-symbolic approximate arithmetic training improves symbolic arithmetic fluency, and to build on this finding by subsequently investigating the amount of training required to yield the effect, and the longevity and specificity of the effect. However, we were unable to replicate the approximate arithmetic training effect across four independent experiments. Bayesian analyses for each of the four current experiments provided weak to moderate evidence in favor of the null hypothesis of no significant difference in arithmetic fluency gain score as a function of training condition. To increase the power to detect an approximate arithmetic training effect, we combined the data from all four experiments and the original data from Experiments 1 and 2 of Park & Brannon 2013, and Experiment 1 of Park & Brannon 2014. This was possible because all experiments used the exact same approximate arithmetic training condition and arithmetic fluency pre and posttest. With the large sample size of 487 individual gain scores (209 in the approximate arithmetic training condition alone) there was no significant difference in arithmetic fluency gain score by training condition. A Bayesian analysis of this combined data set indicated that the data is seven times more likely under the null hypothesis of no difference between training conditions than under the alternative hypothesis where training condition is predictive of arithmetic fluency gain. Importantly, this analysis included the data from the original experiments that individually found support for the idea that approximate arithmetic training improved arithmetic fluency. Even with these prior data included, there was no evidence that non-symbolic, approximate arithmetic training was more effective at improving arithmetic fluency than the other training conditions. Indeed, these data provide significant support for the null hypothesis of no effect of training condition.

All training conditions did, on average, improve participants scores on the arithmetic fluency test. However, without a significant difference in training effect by condition, this increase in performance is likely due to a test-retest effect. The arithmetic problems subjects solved at pre and posttest differed, but the testing environment, test instructions, and the method of response were identical during both testing sessions. The similarity of the testing environment alone could result in improved performance at post-test. At the same time, subjects in the no-contact control condition had a significantly smaller gain score than participants in any of the training conditions. We interpret this result with caution given the small sample in the no-contact condition, however, it suggests that improvements in arithmetic fluency gain scores were also influenced by the expectation that training would improve performance. Taken together, the positive gain scores in arithmetic fluency across all training conditions is likely due to a combination of the test-retest and a placebo effect.

Another way to assess the robustness of transfer from non-symbolic, approximate arithmetic to symbolic arithmetic fluency is to examine the effect size of the gain, rather than to compare gain scores across conditions. The effect sizes for the current experiments were statistically lower than the effect sizes found in Park & Brannon 2013 Experiment 2 and Park & Brannon 2014 Experiment 1. These comparisons indicate that our failure to replicate an effect of training condition on gain score was driven not only by the comparison to control training conditions, but also due to a smaller effect size within the approximate arithmetic training condition when compared to the original effect. This is a known statistical effect that occurs when attempting to replicate an original finding (Button et al., 2013; Ioannidis, 2008; Kahneman, 1971). The original experiment will tend to overestimate the size of the effect because, by chance, the original experiment found an estimate of the effect that was large enough to pass our statistical threshold for significance ($p < .05$). Consequently, subsequent experiments that attempt to replicate this effect will tend to find smaller effect sizes that are closer to the actual true effect size. The “Winner’s curse”, as this effect is sometimes known, may have played a role in our failure to replicate in the current experiments.

The results of the current experiments alone and in combination with the results of Park & Brannon (2013; 2014) indicate that it is highly unlikely that approximate arithmetic training is an effective way to improve arithmetic fluency in adults. Or at least, it is unlikely to improve the arithmetic fluency of adults highly capable and educated in mathematics. The subjects in

Experiments 1-4 and the original Park & Brannon studies were all college students at highly selective private universities who perhaps already employed highly algorithmic strategies to solve double digit addition and subtraction problems. This failure to find a benefit of approximate arithmetic in adults may not mean approximate arithmetic is ineffective for young children just beginning to learn arithmetic. The concrete and repetitive nature of approximate arithmetic may yield benefits in the conceptual understanding of arithmetic, working memory, and/or the ability to mentally manipulate quantity. Improvement in these skills may improve the symbolic math ability of young learners. Previous work has found a positive effect of non-symbolic, approximate arithmetic training for preschool and elementary school age children on tests of basic arithmetic calculation (Hyde et al., 2014), symbolic number line placement (Khanum et al., 2016) and standardized math tests like the TEMA-3 (Ginsburg and Baroody, 2003; Park et al., 2016) and a modified version of the Number Sense Screener (NSS; Research Edition: Glutting and Jordan, 2012; Szkudlarek & Brannon, 2018). Interventions that have incorporated an approximate arithmetic component into a larger intervention have also improved the symbolic math skill of children (Dillon et al., 2017; Käser et al., 2013; Obersteiner et al., 2013; Sella et al., 2016; Wilson et al., 2006). It is possible that children just beginning to learn arithmetic could benefit from non-symbolic and approximate arithmetic training, but that this positive effect does not extend to adults already expert in symbolic calculation. More work is needed to test the robustness and developmental trajectory of an approximate arithmetic training benefit for the symbolic math skills of children.

Finally, why did the original experiments replicate the approximate arithmetic training effect across three independent samples, and the current experiments did not? We can only speculate for answers to this question. One difference between the original effects and the current experiments is that while they were run by the same lab group, these experiments occurred at different universities. However, these universities are well matched in terms of academic performance, entrance selectivity and have similar populations of students. Moreover, the main way in which population differences could affect arithmetic fluency gain score would be through systematic differences in pretest arithmetic fluency scores. Specifically, if one population of students had lower pretest arithmetic scores, they may have had more room to improve through training. However, we found no significant differences in pretest arithmetic fluency score by testing location. We also found no significant difference in the training trajectory of approximate arithmetic participants by experiment. There was a small difference in the way conditions were run in the original experiments compared to the four reported in the current paper. In the original experiments training cohorts who were tested in the same room were largely made up of one training condition with the intention that each participant is not aware of different training conditions. In the current experiments subjects were assigned truly randomly, resulting in subjects from all training conditions tested together in the same room. Thus, it is possible that a cohort effect during the original experiments could have influenced the outcome. There was also variation in the method of response during training (mouse click vs. touch screen) between experiments. Experiments 1-3 all used touch screen responding instead of a mouse click during the training tasks. However, when we reverted to the mouse click response in Experiment 4, our results were the same as in Experiments 1-3, suggesting this was not a factor in the arithmetic gain outcome. Finally, the payment structure for participants varied by experiment between a payment after each session or a lump sum payment at the end of participation. In Experiments 2 and 3, subjects were paid the full amount of compensation after their eighth visit to the lab for their post-test. However, to make sure that this different payment structure did not influence their motivation to complete the training tasks, in Experiment 4 we reverted to daily payments as was done in the original experiments. Again, the payment structure did not appear to make a difference in the experimental result.

Overall, the current experiments provide evidence that approximate arithmetic training does not improve arithmetic fluency more than training with numeral ordering, approximate number comparison, approximate range calculation, visuo-spatial working memory and general knowledge facts. This result does not replicate the previous findings in Park & Brannon 2013, 2014. Indeed, when including the data from the original experiments, there is still support for the null hypothesis. For some readers, our failure to replicate Park and Brannon (2013, 2014) may be interpreted as evidence against the hypothesis that the approximate number system is an evolutionary precursor of the human mathematical mind. We note however, that there is a large amount of correlational evidence linking ANS acuity and symbolic math (Chen & Li, 2014; Fazio et al., 2014; Schneider et al., 2016), and interestingly, approximate arithmetic training did not change ANS acuity in our study or in prior studies (Au et al., 2018; Park & Brannon, 2014). Although a causal connection between a component of the ANS and symbolic math would have been strong evidence for the hypothesis that the human mathematical mind builds upon evolutionarily ancient adaptations for quantifying the world non-symbolically, more research is needed to establish this causal connection. The results of the current experiments highlight the need for rigorous testing and replication of potential cognitive training paradigms.

CHAPTER 4: MATH ANXIETY RELATES TO SYMBOLIC, BUT NOT NON-SYMBOLIC CALCULATION ACCURACY

Introduction

Math anxiety is a negative affective reaction specific to situations involving numbers or mathematics (Beilock & Maloney, 2015; Richardson & Suinn, 1972). Math anxiety occurs inside the classroom (Beilock, Gunderson, Ramirez, & Levine, 2010), in the professional context of teaching and nursing (McMullan, Jones, & Lea, 2012; Swars, Daane, & Giesen, 2006), and during everyday situations such as calculating a tip at a restaurant (Ashcraft & Moore, 2009). High math anxiety is correlated with lower levels of math performance from elementary school into adulthood (Ashcraft & Krause, 2007; Hembree, 1990; Lee, 2009; Ramirez et al., 2013). Competence with mathematics is important for successful academic performance and an increasing number of careers in the 21st century, but high levels of math anxiety deters students from pursuing careers in science, technology, and engineering (Hembree, 1990; Parsons & Bynner, 2005). Therefore, it is important to understand ways to alleviate math anxiety to improve daily experiences with math, and moreover, to expand the range of possible careers for math anxious students.

There are two dominant theories to explain why high math anxiety is correlated with poor mathematical performance. The first theory posits that when a person with math anxiety encounters a situation involving mathematics their anxiety levels restrict domain general abilities needed for calculation such as working memory or attention (Ashcraft & Kirk, 2001; Ramirez et al., 2016, 2013). Evidence for this theory comes from the finding that the relation between math anxiety and math performance differs as a function of a subject’s working memory skill. Adults who have greater working memory ability tend to use their working memory to solve math problems. When faced with a high pressure situation in which their working memory is taxed, subjects with high working memory skill actually perform worse under pressure when solving arithmetic problems than subjects with lower working memory ability (Beilock & Carr, 2005). Several studies have found that children and adults with higher working memory show a stronger correlation between math anxiety and math performance (Beilock & DeCaro, 2007; Ramirez et al., 2016, 2013). When children with math anxiety perform arithmetic in an MRI scanner they have reduced neural activity in posterior parietal and dorsolateral prefrontal cortex, brain regions associated with mathematical reasoning and executive control (Young, Wu, & Menon, 2012). Overall, these findings suggest that the reduction in available working memory due to math anxiety is a cause of worse math performance.

Alternatively, the association between high levels of math anxiety and lower mathematical skill may result from a poor representation of numerical magnitude that creates early difficulties with math (Beilock & Maloney, 2015). Weak fundamental math skills at a young age, could generate anxiety at the onset of formal math education. Evidence for this theory comes from correlations between basic numerical skills and math anxiety levels. For example, subjects with high levels of math anxiety are slower to respond when completing a single digit number comparison task (Dietrich, Huber, Moeller, & Klein, 2015; Maloney, Ansari, & Fugelsang, 2011; Núñez-Peña & Suárez-Pellicioni, 2014). In this task a subject sees two digits (e.g. 3 and 8) and has to identify which symbol represents the greater quantity. Typically developing subjects exhibit a distance effect whereby they take longer to respond when the numerals are close together and are quicker to respond when the numerals represent magnitudes that are farther apart (Moyer & Landauer, 1967). The size of a child's distance effect is predictive of later math skill such that a smaller distance effects predict better math performance (De Smedt, Verschaffel, & Ghesquière, 2009). College undergraduates with high levels of math anxiety have larger distance effects compared to students with low math anxiety (Dietrich, Huber, Moeller, & Klein, 2015; Maloney, Ansari, & Fugelsang, 2011; Núñez-Peña & Suárez-Pellicioni, 2014). Undergraduate students with high math anxiety are also slower to count sets of 5-9 items than low anxiety peers (Maloney, Risko, Ansari, & Fugelsang, 2010). These data suggest that math anxiety could arise from basic numerical processing differences in childhood that make learning mathematics more difficult from elementary math education onwards.

There is conflicting evidence as to whether the correlation between math anxiety and impaired numerical magnitude representations holds when numerical magnitude precision is measured with a non-symbolic, instead of a symbolic, numerical comparison task. In a non-symbolic numerical comparison task, subjects see two arrays of dots and pick the one larger in quantity. This ability to compare non-symbolic numerical magnitudes is supported by the Approximate Number System (ANS). The ANS allows for the representation of non-symbolic numerical quantity across the lifespan (Feigenson et al., 2004). ANS acuity, as measured with the non-symbolic comparison task, is correlated with symbolic math abilities in both children and adults (Chen & Li, 2014; Schneider et al., 2016). There is some suggestion that subjects with high levels of math anxiety have less precise ANS representations. Lindskog and colleagues found that math anxiety mediated the relation between ANS acuity and symbolic math performance in adult subjects, suggesting that the root of math anxiety may be a weaker sense of numerical magnitude (Lindskog, Winman, & Poom, 2017). However, other work has not found this relation between ANS acuity and math anxiety (Braham & Libertus, 2018; Dietrich et al., 2015).

One possible reason there is consistent evidence for the correlation between the symbolic number comparison distance effect and math anxiety is simply because the symbolic numerical comparison task uses numerals. Perhaps exposure to numerals alone provokes a subject's math anxiety, which in turn makes it harder to measure a subject's true distance effect. Math anxious participants may perform worse on symbolic number comparison tasks due to a reduction in domain general cognitive resources, rather than impairments in numerical magnitude representations. One way to test this hypothesis is to see whether math anxiety impacts performance on math tasks that do not involve numerals. Prior research demonstrates that children and adults can perform a variety of non-symbolic, approximate calculation tasks that rely on manipulating ANS representations. For example, young children can perform non-symbolic addition and subtraction (Barth et al., 2006, 2005; Pinheiro-Chagas et al., 2014), scaling operations (McCrink et al., 2016), simple algebra (Kibbe & Feigenson, 2015), ratio comparison (Matthews et al., 2016), and division (Szkudlarek & Brannon, in prep) before they learn the analogous symbolic operations. If math anxious children have a basic magnitude processing deficit, performance on these non-symbolic mathematical tasks will be correlated with math anxiety, because a non-symbolic calculation requires representation of numerical magnitude. Alternatively if math anxiety is triggered by exposure to numerals and causes a reduction in the domain general skills needed to perform mathematics then performance on non-symbolic math tasks should not correlate with math anxiety.

To test this hypothesis we gave elementary school aged children (6-9 years old) and college undergraduates a measure of math and reading anxiety, a non-symbolic calculation test, symbolic math tasks, a reading/vocabulary test, and a test of ANS acuity. We then examined whether performance on the non-symbolic and symbolic math tests correlated with math anxiety levels when controlling for reading anxiety levels to ensure the specificity of this relation. First, we expected to replicate the negative relation between symbolic math performance and math anxiety. We further predicted that accuracy on the non-symbolic calculation test would not be correlated with math anxiety level. Finally, we hypothesized that ANS acuity and math anxiety would each explain unique variance in symbolic math skill, but only ANS acuity would be a significant predictor of non-symbolic math skill.

The comparative nature of our adult and child samples also allowed us to explore when gender differences emerge in levels of math anxiety. Research suggests that women tend to experience higher levels of math anxiety than men (Betz, 1978; Hembree, 1990), and there is some suggestion that this difference in driven by differences in perceived confidence with mathematics (Goetz, Bieg, Lüdtke, Pekrun, & Hall, 2013). Children as young as second grade (age 7) have already acquired gendered stereotypes about math ability, and these stereotypes influence a child's perceived competence with math (Cvencek, Meltzoff, & Greenwald, 2011; Gunderson, Ramirez, Levine, & Beilock, 2012). Girls in their early teenage years (age 12 – 15) have higher levels of math anxiety than boys and the resulting negative correlation between math anxiety level and math performance (Devine, Fawcett, Szűcs, & Dowker, 2012; Meece, Eccles, & Wigfield, 1990; Wigfield & Meece, 1988). It is not yet known when exactly these gender differences emerge. In the current experiment, we compare levels of math anxiety by gender to levels of reading anxiety to control for general academic anxiety. We expect to replicate the higher levels of math anxiety found among women, and then test to see whether we also find this gender difference among elementary school girls and boys.

Overall, the current study was designed to examine three hypotheses. First, we predicted there would be a negative relation between symbolic math and math anxiety, but there will be no significant relation between non-symbolic math and math anxiety. This pattern of results would be evidence against the idea that a numerical magnitude impairment drives math anxiety. Our second hypothesis was that ANS acuity and math anxiety would predict unique variance in symbolic math ability. This finding would confirm the established relation between ANS acuity and symbolic math and indicate that both ANS acuity and math anxiety contribute to symbolic math performance. Third, we examined when in development gender related differences in math anxiety emerge. Based on the literature, we expected women to report a higher level of math anxiety than men, and we did not have a prediction as to whether elementary school aged girls would also show heightened math anxiety.

Materials and Methods

Subjects.

Adults. 311 college undergraduates (mean age = 20.6 years old, standard deviation = 2.0 years; 186 female, 121 male, 4 chose not to report) completed the short Mathematics Anxiety Rating Scale (sMARS; Alexander & Martray 1989). All subjects also completed both a symbolic and a non-symbolic version of an approximate division or ratio comparison task. A subset of this sample also completed additional assessments (dot comparison task n = 303; short Reading Anxiety Rating Scale n = 308; fraction magnitude comparison task n = 301; vocabulary test n = 310). Written consent was obtained in accordance with a protocol accepted by the University of Pennsylvania’s Institutional Review Board.

Children. 165 children (mean age = 7.73 years old, standard deviation = 1.0 years; 80 female, 83 male, 2 chose not to report) completed the Child Math Anxiety Questionnaire (C-MAQ; Ramirez et al., 2013; Suinn, Taylor, & Edwards, 1988; Maloney et al., 2015). All children also completed both a symbolic and a non-symbolic version of an approximate division or ratio comparison a task. A subset of this sample also completed additional assessments (dot comparison task n = 156; fraction magnitude comparison task n = 155; Woodcock Johnson Reading cluster n = 147). Written parental consent and children's verbal assent were obtained in accordance with a protocol accepted by the University of Pennsylvania's Institutional Review Board.

Procedure.

Adults. Subjects were recruited through a university psychology department subject pool. Participants received course credit as compensation for their time. All subjects completed the battery of tasks over two sessions that occurred on separate days in the laboratory. The two sessions occurred 1-3 days apart. The math and reading anxiety questionnaires were completed at the end of the second session. The non-symbolic and symbolic approximate calculation tests were completed during the first session, and the dot comparison task, fraction magnitude comparison, and vocabulary tasks were completed in counterbalanced order across participants during the second session. The majority of the participants (n =231) were given a ratio comparison task as their measure of approximate non-symbolic and symbolic calculation, and the remaining subjects were given a division task (n = 80). The math and reading anxiety questionnaires were administered with paper and pencil for the majority of participants (n = 230) and on a desktop computer for the remaining participants (n = 81).

Children. All subjects completed all tasks individually with an experimenter in a quiet room at their after school program. The dot comparison task and approximate non-symbolic and symbolic calculation tasks were completed on a 15 inch laptop computer. The Key-Math-3 Numeration and the C-MAQ were administered using paper and pencil. Children completed the approximate non-symbolic and symbolic calculation tasks first and the order of the other tasks was random across participants dependent on the child's availability. 80 subjects were given a ratio comparison task as their measure of approximate non-symbolic and symbolic calculation, and the remaining subjects were given a division task (n = 85). A subset of participants completed an extra section of the Key-Math-3 (N= 80) or a short division knowledge questionnaire (n = 85), but these measures are not included in this paper. Each participant was tested for a total of 45-60 minutes across 2-3 days. Children received stickers throughout the session to maintain motivation.

Adult Experimental Tasks.

Short Mathematics Anxiety Rating Scale (sMARS). This questionnaire is a 25 item version of the longer 98 item Mathematics Anxiety Rating Scale (Alexander & Martray, 1989; Ashcraft & Moore, 2009; Richardson & Suinn, 1972) Participants responded to questions about how anxious they would feel during particular situations involving math and numbers using a 5-point scale (1 = not at all, 2 = a little, 3 = a fair amount, 4 = much, 5 = very much). The questions involved focused mainly on using math in an educational setting (e.g. "Taking an examination (quiz) in a math course" and "Opening a math or statistics book and seeing a page full of problems"). Performance was quantified as the average of all 25 items[10].

Short Reading Anxiety Rating Scale (sRARS). This questionnaire is exactly parallel to the sMARS, but instead evaluates participant's anxiety in situations involving reading. The 25 items are parallel to the math items (e.g. "Taking an examination (quiz) in an English course" and "Picking up an English book to begin a difficult reading assignment"). This questionnaire also uses a 5-item scale (1 = not at all, 2 = a little, 3 = a fair amount, 4 = much, 5 = very much). Performance was quantified as the average of all 25 items. The order in which subjects completed the sRARs and the sMARS was counterbalanced across subjects.

Dot Comparison Task. Two dots arrays appeared on a black screen for 750ms. After this time, the arrays of dots were occluded and the subject indicated the array greater in quantity by touching the computer screen. Participants completed 200 trials of this task, with feedback on

[10] Five participants did not turn over the paper of the questionnaire and completed 17/25 questions. Their performance was the average of the first 17 items on the questionnaire.

every trial. The number of dots ranged from 8 to 32. The stimuli were created to evenly sample a stimulus space that varied by the ratio between the number, size, and the spacing of the dots. To encourage greater reliability of the measurement, trial level difficulty was titrated (Lindskog et al., 2013). The titration procedure calculated the percentage correct over the last 5 trials. The ratio between the two dot arrays moved one log level farther apart if the accuracy was less than 70% and moved one log level closer together if the accuracy was greater than 80%. A quantitative index of each subject's ANS acuity was calculated with the Weber fraction (*w*) as specified in (DeWind et al., 2015). This model accounts for the effects of non-numerical features of dot arrays on numerical discrimination (DeWind et al., 2015).

Vocabulary Test. This test was the same as used in (Park & Brannon, 2014). Subjects answered 42 multiple choice vocabulary questions in 5 minutes. The questions were taken from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman & Dermen, 1976). Performance was calculated as the number of problems answered correctly minus ¼ of the number incorrect to discourage guessing.

Fraction Magnitude Comparison Task. Subjects viewed two fractions displayed in the middle of the screen in white on a black background. The goal of the task was to pick the fraction greater in magnitude by pressing the F key for the left fraction or the J key for the right fraction. The stimuli were the same as used in (Fazio et al., 2015). Performance was quantified as the number of trials correct.

Non-symbolic, Approximate Calculation Tasks. Three-hundred and eleven participants completed either a ratio comparison calculation test (n = 231) or an approximate division task (n = 80). In the ratio comparison task two gumball machines were presented with green and orange gumballs and a picture of an alien character. Subjects were instructed to pick which gumball machine had the best chance of giving Mr. Alien an orange gumball on his first try. In one trial the correct choice had more orange gumballs, in one trial the correct choice had more orange gumballs and a greater number of total gumballs, in one trial the correct answer had fewer total items, and in the last trial the correct answer had fewer orange gumballs. Regardless of accuracy on this last trial, the subject was told “The machine with the *most* orange gumballs is not always the one with the *best chance* of an orange gumball.” After these practice trials, subjects were told that the alien wants a blue or white gumball, counterbalanced across participants. 72 subjects completed 60 trials, and 159 subjects completed 72 trials.

In the approximate, non-symbolic division task subjects were introduced to a bee and a set of orange colored dots ranging from 32 to 192 dots at the top of the screen. Subjects were told these dots are food for the bee. Then, the dots fell onto a set of 2 to 8 flower petals at the bottom of the screen. Subjects were told that the same amount of food falls onto each flower petal. One of the flower petals that now contained a set of dots moved to the middle of the left side of the screen and the flower disappeared. A new target flower with a quantity of dots appeared on the right side of the screen. Subjects were told to pick the flower petal that had more food. Subjects completed 8 of these trials while the experimenter repeated the instructions. Then, the experimenter told the subjects that in the next trials it is a “foggy day in the garden” and the bee will not be able to see how many pieces of food fall onto each petal of the flower, but that they would have to imagine how many pieces of food were on one petal of the flower and compare this imagined quantity to the target flower petal. In these trials when the dots fell from the top of the screen subjects could not see how many pieces fell onto each flower petal. Then this obscured flower petal moved up to the middle of the left side of the screen and subjects had to pick whether the imagined quantity on the obscured petal or the new target quantity was larger. Subjects received feedback for the first 32 trials. Subjects completed 24 more trials without feedback. Both the ratio and division tasks are non-symbolic because each numerical magnitude involved in the computation was depicted as a set of dots (a concrete representation of the numerical magnitude). Both tasks are approximate because subjects had to pick one of two potential answers and did not have to exactly calculate. To combine the data from both non-symbolic, approximate calculation tasks we Z-scored the data within each task (ratio or division) and used these Z scores in our analyses.

Symbolic, Approximate Calculation Test. The same 311 participants completed a parallel symbolic version of the approximate calculation task. The procedure and numerical values for the symbolic ratio task and the symbolic division task were identical to that described for the non-symbolic tasks except that they were represented with Arabic numerals instead of dots. Despite the use of Arabic numerals, the experimenter did not use number words to explain the symbolic tasks. The symbolic tasks were approximate because subjects had to pick one of two potential answers and did not have to exactly calculate. To combine the data from both symbolic, approximate calculation tasks we again Z-scored the data within each task (ratio or division) and used these Z scores in our analyses.

Child Experimental Tasks.

Child Math Anxiety Questionnaire (CMAQ-R). This measure is a 16-item revised version of the original Child Math Anxiety Questionnaire (C-MAQ; Ramirez et al., 2013; Suinn, Taylor, & Edwards, 1988). This revised version was designed for first and second grade children and has been used in previously (Maloney, Ramirez, Gunderson, Levine, & Beilock, 2015). Items ask children how nervous they would feel during various math related situations. Children respond by pointing to one of five faces that depict different states of nervousness on an emotional gradient (1 = not nervous at all, 2 = a little nervous, 3 = somewhat nervous, 4 = very nervous, 5 = very, very nervous). Before beginning the items, children were asked if they understood what it means to be nervous and were told “Sometimes people feel nervous when they are worried about something or are afraid they might not know the answer”. Then, to ensure that children understood how to use the five point smiley face scale, children were asked “How nervous do you feel when you’re looking down from the top of a building? (child points) “What if you were not nervous at all?” (child points) “What if you were very, very nervous?”) The experimenter did not proceed until children gave the correct responses to these questions. Three items including displaying a flashcard that depicted a bar graph, a clock, and a set of cubes to the child while reading the question. Items targeted math in an educational setting (“How do you feel when you have to sit down and start your math homework?”) and specific math problems (“How would you feel if you were given this problem: There are 13 ducks in the water. There are 6 ducks in the grass. How many ducks are there in all?") Children did not have to solve the math problems, only respond how nervous they would feel. Performance was quantified as the average of all 16 items.

Child Reading Anxiety Questionnaire. This questionnaire is exactly parallel to the CMAQ-R but asks questions relevant to English class and reading instead of mathematics (Ramirez et al., 2019). Example questions include "How do you feel when you have to sit down and start your reading homework?" and "How would you feel if you are asked to read this sentence (show child flashcard)". Six questions required the experimenter to show a flashcard that depicted words or a sentence while reading the item aloud. Children did not have to read the words or sentences but simply had to describe how nervous reading the words would make them feel. Child responded in the same way as the CMAQ-R, using the 5-point smiley face scale. The order in which children completed the CMAQ-R and the child reading anxiety questionnaire was counterbalanced across subjects. The instructions on how to use the smiley face scale described above were given only once before the first questionnaire. Performance was quantified as the average of all 16 items.

Dot Comparison Task. This task was exactly the same as the dot comparison task used with adults.

Key Math-3 Numeration subtest. The Numeration subtest of the Key Math-3 Diagnostic Assessment Form B (Connolly, 2007) was administered. The Numeration section is a test of general basic math skills such as place value, counting, the relative magnitude of numbers, and an understanding of fractions, decimals, and percentages. For example, children are presented with 4 numerals and told "Read the numbers in order, from least to greatest". Performance was quantified as the age standardized scale score for this subtest.

Woodcock-Johnson IV Test of Cognitive Abilities. Participants' reading abilities were assessed using the "Basic Reading Skills" cluster of the Woodcock-Johnson. This cluster is comprised of the "Letter-Word Identification" and "Word Attack" subtests. In the "Letter-Word Identification" subtest, participants named letters and read words aloud. In "Word Attack," participants read nonsense words and identified letter sounds. We used the age standardized Basic Reading Skills score.

Non-symbolic, Approximate Calculation Test. The non-symbolic ratio and division tasks were identical to the tasks used with adult subjects. 80 children completed the ratio task, and 85 completed the division task. All children who completed the ratio task did 60 trials.

Symbolic, Approximate Calculation Test. The symbolic ratio and division tasks were identical to the tasks used with adult subjects. 80 children completed the ratio task, and 85 completed the division task. All children who completed the ratio task did 60 trials.

Results

Descriptive statistics. Math anxiety scores were calculated by taking the average response on the 16 items of the CMAQ-M for children and the 25 items of the sMARS for adults (Table 6 & 7). Scores for both children and adults were within the range previously reported (e.g., Maloney et al., 2015). Reading anxiety scores were calculated by taking the average response on the 16 items of the CMAQ-R for children and the 25 items of the sRARS for adults. Tables 6 & 7 show the descriptive statistics for all measures collected for adults and children respectively.

The relation between math anxiety, ANS acuity, approximate calculation, mathematics performance and control measures. Before running zero order correlations we transformed outcome measures to approach normality separately for the children and adult data. For the adult data, we conducted the natural log transformation on ANS acuity (Shapiro-Wilk W = .97), took the square root of the sMARS score (W = .99) and the sRARS[11] (W = .98), raised the fraction magnitude comparison accuracy score to the third power (W = .99), raised approximate symbolic calculation accuracy to the fourth power (W = .97), and the non-symbolic calculation to the second power non-symbolic (W = .98). For the child data we conducted the

---

[11] A natural log transformation results in in a higher Shapiro-Wilk W (W = .99) for the adult reading anxiety scores, but we included the square root transformation to keep all the anxiety measures on the same scale. None of the results reported in this paper change as a result of this transformation.

square root transformation on both the CMAQ-R and child reading anxiety questionnaire scores (math W = .97; reading W = .98) and the Key-Math-3 Numeration scores (W = .98). Zero order correlations for the adults are reported in Table 6, and for children in Table 7.

In line with our hypotheses, adult math anxiety was significantly correlated with both symbolic math measures (symbolic calculation $t_{307} = -4.06$ $p < .001$, fraction comparison $t_{296} = -4.06$ $p < .001$), but was not significantly correlated with ANS acuity ($t_{298} = .957$ $p = .34$) or non-symbolic calculation ($t_{309} = -1.19$ $p = .23$). The correlations between math anxiety and the symbolic and non-symbolic calculation measures were significantly different from each other (Pearson and Filon's z = 2.91 p = .002). For children there were significant zero order correlations between math anxiety and all four measures (symbolic calculation $t_{163} = -3.04$ $p = .003$; Key-Math-3 Numeration $t_{151} = -4.28$ $p < .001$; non-symbolic calculation $t_{163} = -2.86$ $p = .005$; ANS acuity $t_{149} = 2.69$ $p = .008$). However, when controlling for reading anxiety, the correlations between math anxiety and the symbolic calculation measures were significantly stronger than the correlation between math anxiety and the non-symbolic measures (Pearson and Filon's z = 1.92 p = .02).

Table 6. Descriptive statistics and bivariate correlation matrix for adults

		M	*SD*	1	2	3	4	5	6
1	Math Anxiety	2.38	.72						
2	Reading Anxiety	2.10	.72	.23***					
3	ANS acuity	.154	.04	.06	.02				
4	Non-symbolic calculation	.850	.10	-.07	.04	-.16**			
5	Symbolic calculation	.914	.08	-.23***	.06	-.19**	.49***		
6	Fraction Comparison	155	.17	-.22***	-.02	-.24***	.21***	.35***	
7	Vocabulary	14.4	7.1	.07	-.25***	-.16**	.13*	.08	.09

Note. *M* = mean, *SD* = standard deviation. Math Anxiety (sMARS), Reading Anxiety (sRARS). ANS acuity is calculated from the dot comparison task. Note: *** p<.001 ** p < .01 * p < .05

Table 7. Descriptive statistics and bivariate correlation matrix for children

		M	*SD*	1	2	3	4	5	6
1	Math Anxiety	2.39	.85						
2	Reading Anxiety	2.34	.87	.75***					
3	ANS acuity	.329	.14	.22**	.21**				
4	Non-symbolic calculation	.724	.11	-.22**	-.29***	-.35***			
5	Symbolic calculation	.654	.13	-.23**	-.19*	-.31***	.59***		
6	Key-Math-3 Numeration	11.6	5.4	-.33***	-.37***	-.36***	.32***	.31***	
7	Reading Cluster	89.1	18.6	-.19*	-.33***	-.08	.15	.14	.60***

Note. *M* = mean, *SD* = standard deviation. Math Anxiety (CMAQ-R), Reading Anxiety. ANS acuity is calculated from the dot comparison task. Reading Cluster is the Reading cluster of the Woodcock-Johnson IV Note: *** p<.001 ** p < .01 * p < .05

Predicting symbolic and-non-symbolic math skills. We ran hierarchical regression analyses predicting non-symbolic and symbolic math separately. For the symbolic math outcome, we created a composite symbolic math score by averaging the already Z-scored approximate symbolic calculation measure with the Z-score of the fraction magnitude comparison test in adults, and the Z-score of the Key-Math-3 Numeration test in children. We then ran four hierarchal regression analyses: one each for the symbolic and non-symbolic math outcomes in both children and adults. For all models, in step 1 we entered math anxiety, reading anxiety, and the reading control measure to test for a negative relation between math anxiety and symbolic math when controlling for reading academic performance and reading anxiety levels. In step 2 we added ANS acuity to the model to replicate the relation between sharper ANS acuity and better symbolic math scores, and to determine whether math anxiety and ANS acuity predict the same or different variance in symbolic math ability. We predicted that ANS acuity but not math anxiety would be predictive of non-symbolic approximate calculation performance. We further predicted that math anxiety and ANS acuity would be predictive of our composite measure of symbolic approximate calculation performance and symbolic mathematics performance as measured by the fraction magnitude comparison task (adults) or Key-Math-3 (children).

Symbolic math outcome. Math anxiety was a significant predictor of symbolic math accuracy for both the adults (Table 8; β = -.13 p = .009) and the children (Table 10; β = -.24 p = .02). The negative sign indicates that high math anxiety was related to lower symbolic math skills for both age groups. This relation held even when controlling for reading anxiety and performance on a reading control measure (Vocabulary test in adults and the Woodcock-Johnson Reading Cluster in children). In step 2, math anxiety continued to be a significant predictor in both adults (β = -.11 p = .04) and children (β = -.21 p = .03). ANS acuity was also a unique predictor of symbolic math skills in both age groups (adults β = -.12 p = .02; children β = -.33 p < .001). The negative sign indicates that sharper ANS acuity was related to higher symbolic math accuracy.

Non-symbolic math outcome. As predicted, math anxiety was not a significant predictor of non-symbolic calculation accuracy for adults (Table 9; β = -.10 p = .10) or children (Table 11; β = -.01 p = .93). In step 2, math anxiety continued to not be a significant predictor for adults (β = -.07 p = .24) or children (β = -.02a p = .86), but ANS acuity was a significant predictor of non-symbolic math accuracy for both age groups (adults β = -.15 p = .01; children β = -.31 p < .001). The negative sign indicates that sharper ANS acuity was related to higher non-symbolic math accuracy.

Table 8. Multiple regression analyses predicting symbolic math in adults

		B	*SEβ*	*β*
Step 1	$R^2 = .03^*$			
	Intercept	494.0	.050	-.1115*
	Math Anxiety	-163.3**	.051	-.1339**
	Reading Anxiety	68.87	.053	.0583
	Vocabulary	3.168	.052	.0781
Step 2	$R^2 = .04^*$			
	Intercept	590.8	.050	-.1111*
	Math Anxiety	-133.5*	.052	-.1094*
	Reading Anxiety	70.14	.053	.0594
	Vocabulary	1.450	.054	.0357
	ANS acuity	-773.8*	.052	-.1205*

Table 9. Multiple regression analyses predicting non-symbolic math in adults

		B	*SEβ*	*β*
Step 1	$R^2 = .03^*$			
	Intercept	22.76***	.057	-.0186
	Math Anxiety	-3.900	.059	-.0970
	Reading Anxiety	4.120	.061	.1058
	Vocabulary	.2172**	.059	.1623**
Step 2	$R^2 = .05^*$			
	Intercept	26.73***	.057	-.0240
	Math Anxiety	-2.837	.060	-.0705
	Reading Anxiety	4.274	.061	.1097
	Vocabulary	.1521	.061	.1136
	ANS acuity	-31.98*	.059	-.1510*

Table 10. Multiple regression analyses predicting symbolic math in children

		B	*SEβ*	*β*
Step 1	*R^2 = .31****			
	Intercept	-.1794	.064	.0465
	Math Anxiety	-.7278*	.097	-.2374*
	Reading Anxiety	-.1549	.103	-.0512
	Reading Cluster	.0177***	.069	.3802***
Step 2	*R^2 = .43****			
	Intercept	-1.359**	.061	.0732
	Math Anxiety	-.6498*	.094	-.2120*
	Reading Anxiety	.0224	.100	.0074
	Reading Cluster	.0170***	.064	.3659***
	ANS acuity	-.7369***	.061	-.3261***

Table 11. Multiple regression analyses predicting non-symbolic math in children

		B	*SEβ*	*β*
Step 1	*R^2 = .10****			
	Intercept	1.191	.076	.0527
	Math Anxiety	-.0364	.114	-.0103
	Reading Anxiety	-.9262*	.121	-.2658*
	Reading Cluster	.0035	.081	.0647
Step 2	*R^2 = .20****			
	Intercept	-.10	.072	.0838
	Math Anxiety	-.0707	.111	-.0200
	Reading Anxiety	-.6022	.117	-.1728
	Reading Cluster	.0026	.075	.0479
	ANS acuity	-.8181***	.072	-.3143***

Gender differences in anxiety level. As shown in Figure 17, women reported higher levels of both math and reading anxiety than men. Children showed no effect of gender. A mixed effects ANOVA on adult anxiety scores showed a significant main effect of gender ($F_{1,302}$ = 16.48 $p < .001$), a significant main effect of topic ($F_{1,302}$ = 29.75 $p < .001$), and a significant topic by gender interaction ($F_{1,302}$ = 4.29 $p = .04$). For the children, the same mixed effects ANOVA revealed no significant effects (Gender $F_{1,161}$ = .163 $p = .73$; Topic = $F_{1,161}$ = 1.14 $p = .29$; Topic by gender interaction $F_{1,161}$ = .080 $p = .52$). Post-hoc tests on the data with adults revealed that women reported significantly higher math anxiety than men (Wilcoxcon rank sum test W = 7762, $p < .001$). There was no significant difference between men and women in reading anxiety (Wilcoxcon rank sum test W = 9705, $p = .08$).

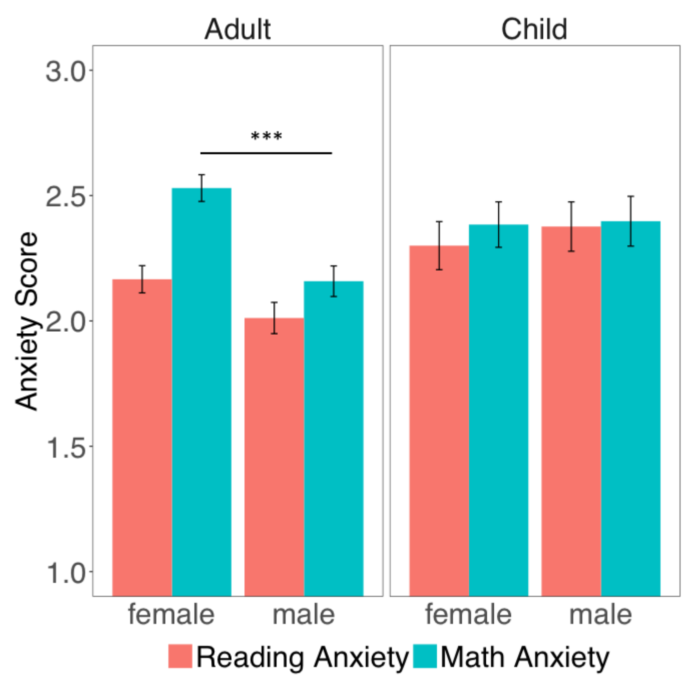

Figure 17. Math and reading anxiety levels plotted by gender and age. Women reported higher math and reading anxiety than men, but there was no significant gender difference among elementary school children.

Discussion

A growing body of evidence has linked high math anxiety to poor math performance in students around the world and in children by elementary school (Lee, 2009; Ramirez et al., 2013). An open question is whether one source of math anxiety is poor magnitude representation (e.g.

Beilock & Maloney, 2015; Dietrich et al., 2015; Maloney et al., 2011; Núñez-Peña & Suárez-Pellicioni, 2014). To address this question we examined the relationship between math anxiety and approximate calculation in both symbolic and non-symbolic formats. We found that math anxiety is a significant predictor of performance on approximate, symbolic calculation tasks but not approximate, non-symbolic calculation tasks in both elementary school aged children and adults. Moreover, the correlation between math anxiety and approximate symbolic calculation was significantly greater than and correlation between math anxiety and non-symbolic calculation in both adults and in children when controlling for reading anxiety. The lack of a significant correlation between accuracy on a non-symbolic calculation task and math anxiety suggests that math anxiety is unlikely to stem from inherently weak numerical magnitude representations. Instead, the presence of numerals during symbolic calculation may trigger math anxiety and resulting domain general impairments associated with anxiety (Ashcraft & Kirk, 2001; Ramirez et al., 2016, 2013). Previous work has found a relation between the distance effect on a single digit number comparison task and math anxiety, and this result has been interpreted as evidence for a deficit in basic magnitude processing among participants with high levels of math anxiety (Dietrich, Huber, Moeller, & Klein, 2015; Maloney, Ansari, & Fugelsang, 2011; Núñez-Peña & Suárez-Pellicioni, 2014). However, our findings suggest that the numerals in the symbolic number comparison task rather than magnitude processing per se may have induced math anxiety.

In children, although there was a zero-order correlation between non-symbolic calculation and math anxiety level, the correlation was no longer significant when controlling for reading anxiety. Furthermore, when controlling for reading anxiety the correlation between math anxiety and symbolic calculation was significantly greater than the correlation between math anxiety and non-symbolic calculation. It is important to note that almost all the zero order correlations were significant for children and are therefore likely due to something akin to the positive manifold found in research on intelligence and the high correlation between reading and math skills in elementary school (Cantin et al., 2016; Van Der Maas et al., 2006; Wang et al., 2015).

For adults, high levels of math anxiety predicted lower symbolic math performance when controlling for reading anxiety and vocabulary performance. Reading anxiety functions as a strong test of the specificity of the relation between math anxiety and math skill. With this control measure, the relation between math anxiety and math performance cannot be due to general academic anxiety, but is instead specific to how students feel about math in particular. Research on how reading anxiety impacts reading ability is in very early stages (Ramirez et al., 2019). In our child sample, reading anxiety was correlated with performance on a standardized test of reading ability (the Woodcock-Johnson Reading Cluster), and in adults reading anxiety was correlated with a measure of a vocabulary. Both reading and math anxiety may be important domain-specific influences on children's academic development.

For children and adults, math anxiety and ANS acuity explained unique variance in symbolic math performance. While there was a significant zero order correlation between ANS acuity and math anxiety in children, both factors explained unique variance in symbolic math in our regression model. Math anxiety and ANS acuity were not correlated in adults. Lindskog and colleagues reported that math anxiety mediated the relation between ANS acuity and symbolic math in their sample of adults, and suggested that math anxiety could explain the well-established correlation between ANS acuity and math performance (Schneider et al., 2016). However, in our larger sample of adults and children we find no evidence to support this claim. Instead our findings suggest that ANS acuity and math anxiety impact overall math ability via different mechanisms (Braham & Libertus, 2018; Dietrich et al., 2015).

A limitation of the current study is that we did not collect a measure of symbolic numerical comparison, and so we were unable to directly compare how numerical magnitude judgements relate to math anxiety when measured in non-symbolic and symbolic format. A previous experiment found that math anxiety correlated with symbolic, but not non-symbolic numerical magnitude comparison performance in adults (Dietrich et al., 2015). The results of the current experiment are consistent with this finding. Nevertheless, future work can confirm this pattern of results in children. We predict that the symbolic distance effect would correlate with math anxiety levels and the non-symbolic distance effect would not correlate.

An implication of our findings is that non-symbolic calculation tasks may be a beneficial addition to math curricula for students with math anxiety. Prior studies indicate that non-symbolic calculation tasks can improve the math skills of children, and especially children with low math skills (Dillon et al., 2017; Park et al., 2016; Szkudlarek & Brannon, 2018). Non-symbolic calculation tasks may be a way to teach calculation skills to math anxious students without provoking anxiety. While the current work is correlational in nature and cannot establish a causal link, future studies should test the hypothesis that introducing non-symbolic calculation operations before teaching the same operation in a symbolic form is beneficial for math learning. A potential benefit of practice with non-symbolic calculation could be a reduction in math specific anxiety. Calculation with numerical symbols is of course the goal of math education, but non-symbolic calculation may be a stepping stone for students, especially for those with high levels of math anxiety.

The third and last goal of the current experiment was to explore the emergence of gender differences in math anxiety levels. Consistent with previous work, we found higher self-reported math and reading anxiety among women compared to men (Hembree, 1990). In contrast, we found no significant effect of gender on math or reading anxiety levels in elementary school children. Our findings suggest that environmental factors lead to higher levels of math and reading anxiety among women. Interventions that begin in elementary school could be effective at reducing the increases in math anxiety reported among preteen and teenage girls (Devine et al., 2012; Wigfield & Meece, 1988).

Overall our data indicate that the negative relation between math anxiety and symbolic math does not extend to non-symbolic mathematical calculation. We interpret this result as evidence against the theory that the relation between math anxiety and symbolic math is due to a fundamental deficit in magnitude representation. Instead, we suggest that numerical symbols trigger anxiety related reductions in domain general skills necessary to perform mathematics. This work has important implications for potential math interventions among math anxious students. Non-symbolic calculation may be an effective way to introduce math concepts in an educational context without arousing anxiety. Children’s remarkable ability to intuitively perform abstract mathematical operations highlights the fact that math is a way to explain the fundamental logic of the world, and all children have the ability to understand this logic.

GENERAL DISCUSSION

Understanding the link between our basic intuitive sense of number and symbolic mathematics is important, because understanding the basis for how children think about number may help improve math instruction. Mathematical competence is important for success in the classroom, but also for the myriad of daily tasks that require basic numerical knowledge. Individuals with low numerical skill are more likely to default on a mortgage, have poorer long term health, and save less for retirement (Banks & Oldfield, 2007; Gerardi, Goette, & Meier, 2013; Parsons & Bynner, 2005). In school, math skill in kindergarten is the strongest predictor of later academic achievement when compared to literacy, attention, and socioemotional skills (Duncan et al., 2007). The importance of early math skills has been replicated in subsequent studies (Geary, Hoard, Nugent, & Bailey, 2013; Jordan, Kaplan, Ramineni, & Locuniak, 2009; Watts, Duncan, Siegler, & Davis-Kean, 2014). Despite the importance of strong mathematical skill, only 40% of 4th grade students and 34% of 8th grade students are at or above proficiency level in mathematics in the United States (National Center for Education Statistics, 2017). Can we leverage our intuitive sense of number to improve math skill? Across four chapters the current dissertation explored the relation between our intuitive ability to perform mathematical calculation and symbolic mathematics.

Chapter Summaries

The first chapter demonstrated that children can successfully compare ratios using their approximate number system in both non-symbolic and symbolic formats. This result is consistent with previous work demonstrating that children possess non-symbolic ratio reasoning ability (Boyer & Levine, 2015; Falk et al., 2012; Yost et al., 1962) and extends these findings to include intuitive ratio reasoning with Arabic numerals. Children’s ability to successfully compare ratios is not due to the application of an incorrect heuristic whereby they perform a unidimensional comparison across one element of the ratio or the total numbers of items in the ratio as a whole. Instead children are able to combine information about both the number of preferred and non-preferred items in each ratio to make a choice about which set of items is more likely to yield a preferred colored item. Children are more accurate comparing ratios of dots than comparing ratios depicted with Arabic numerals. Moreover, children who engaged in non-symbolic ratio reasoning first performed better on the symbolic task compared to children tested on the symbolic task first. This facilitation effect supports the idea that non-symbolic ratio reasoning can function as a scaffold for symbolic ratio reasoning. Providing further support for such scaffolding, non-symbolic ratio reasoning mediated the relation between children's ANS acuity and their symbolic mathematics performance in the domain of probabilities, but ANS acuity remained a better predictor of more general numeration skills. This pattern of results suggests that conceptual overlap between non-symbolic and symbolic calculation is necessary for non-symbolic calculation to function as a scaffold. This mediation provides support for the hypothesis that non-symbolic calculation is a mechanism of the relation between ANS acuity and symbolic math. Collectively, the findings of the first chapter suggest that practice with non-symbolic, approximate ratio comparison could be a beneficial tool for teaching children about ratios.

The second chapter presents the first evidence of non-symbolic division skill in both 6 to 9 year old children and university undergraduates. Both children and adults successfully performed true non-symbolic and symbolic approximate division both during a phase with feedback and during a generalization phase with the novel divisors of 3, 4 and 6. Previous work had shown that children can perform a non-symbolic, approximate scaling operation that increases or decreases dot arrays by a factor of 2 or 4, however it was unknown whether this ability can extend to other divisors (McCrink et al., 2016; McCrink & Spelke, 2010, 2016). This chapter demonstrates that all subjects were slightly better at dividing by 2 than any other divisor, however, subjects were well above chance when dividing by all divisors 2 to 8. This preference for dividing by small divisors mirrors the size effect found during adult symbolic division, suggesting a common cognitive mechanism for calculating symbolically and non-symbolically (Mauro et al., 2003). Successful approximate division calculation was not dependent on formal division knowledge. Children who were unable to identify the division symbol or unable to solve any simple division equations nevertheless solved both approximate division tasks significantly above chance. This result indicates that at the beginning of math education our understanding of division is not based in our symbolic mathematical system but is rather an affordance of the Approximate Number System. Both child and adult participants used approximate ANS based strategies to solve the non-symbolic and symbolic division tasks. Subjects accuracy was modulated by the ratio between the quotient and the comparison target array indicating use of an approximate strategy when making their choice. Moreover, many adult participants explicitly reported using approximate strategies on the non-symbolic, and even the symbolic task. Children were more accurate at non-symbolic division calculation but adults were more accurate during symbolic division. The ability to more accurately calculate with numerals than concrete stimuli may mark an important transition in math education. Future work can directly assess whether students initially use non-symbolic calculation as a model for symbolic calculation, and then build symbolic calculation concepts based in these intuitive arithmetic models. Finally, in both children and adults, the ability to divide non-symbolically mediated the relation between approximate number system acuity and symbolic math performance, suggesting that the ability to calculate non-symbolically may be a mechanism of the relation between ANS acuity and symbolic math. Encouraging a connection between children's correct division intuitions and symbolic division may be beneficial for early arithmetic skills.

The third chapter attempted to replicate and extend the finding that non-symbolic, approximate addition and subtraction training can improve the symbolic addition and subtraction fluency of university undergraduates (Au et al., 2018; Park & Brannon, 2013; Park & Brannon, 2014). Unexpectedly, I was unable to replicate this training effect across four independent replications and extensions. Subjects who trained with the approximate arithmetic training paradigm did not improve their arithmetic fluency more than subjects who trained with numeral ordering, numerical comparison, numerical range estimation, short term memory, or knowledge question training paradigms. Moreover, the effect sizes for improvement found in the original Park & Brannon experiments were significantly larger than the effect sizes reported in this chapter.

This failure to replicate indicates that non-symbolic arithmetic practice does not benefit the arithmetic fluency of adult subjects. This result is in contrast to findings that non-symbolic, approximate addition and subtraction training in children can improve basic arithmetic calculation (Hyde et al., 2014), symbolic number line placement (Khanum et al., 2016) and standardized math tests (Park et al., 2016; Szkudlarek & Brannon, 2018). Interventions that include an approximate arithmetic component into a larger intervention have also found positive effects on the symbolic math skills of children (Dillon et al., 2017; Käser et al., 2013; Obersteiner et al., 2013; Sella et al., 2016; Wilson et al., 2006). It is possible that for college undergraduates, practice with the concept of addition and subtraction does not benefit their arithmetic fluency, because college students are already adept at algorithmic strategies to solve addition and subtraction problems. These strategies are unlikely to be affected by non-symbolic, approximate arithmetic practice. Overall, the four experiments included in this chapter delineate the limits of potential benefits of approximate arithmetic training. Specifically, this chapter suggests that approximate arithmetic training may be the most useful during the initial acquisition and integration of arithmetic concepts and less useful once arithmetic calculation is mastered.

The fourth chapter examined how math anxiety relates to non-symbolic and symbolic calculation. There are two prominent theories for why high levels of math anxiety leads to weaker math performance. The first theory states that social learning may create math specific anxiety, which in turn depletes domain general cognitive skills necessary for mathematics. The second theory posits that poor numerical magnitude representation leads to math difficulties and subsequently math focused anxiety. Non-symbolic calculation offers a unique opportunity to test the second hypothesis, because these tasks involve calculating with non-symbolic numerical magnitudes without using mathematical symbols. If high math anxiety is due to deficits in magnitude processing, participants high in math anxiety should also perform poorly on tests of non-symbolic math calculation. Chapter 4 found that math anxiety predicted symbolic math accuracy even when controlling for reading anxiety, ANS acuity, and vocabulary knowledge among both university undergraduates and children. ANS acuity also explained unique variance in symbolic math ability in both children and adults. ANS acuity, and not math anxiety level, explained significant variance in non-symbolic calculation ability. This pattern of results supports the hypothesis that the negative relation between math anxiety and math skill is caused by a reduction in domain general cognitive skills when encountering numerical symbols, and not an underlying numerical magnitude processing deficit (Ashcraft & Kirk, 2001; Erin A. Maloney et al., 2011). Moreover, these results suggest that non-symbolic math tasks may be a way to introduce mathematical concepts to highly math anxious students without provoking anxiety. It may be beneficial for math anxious students to practice with non-symbolic calculation to build familiarity with math concepts before introducing these same concepts in a symbolic format. Building connections between a student's intuitions about math calculation and formal math education may help create confidence in a student's ability to perform highly abstract and formal mathematical operations.

Synthesis and future directions

The Approximate Number System supports the ability of humans and non-human animals to approximately compare, estimate, and manipulate large non-symbolic numerical quantities without language or symbols (Feigenson, Dehaene, & Spelke, 2004). Numerical cognition research has largely focused on how the acuity of this system relates to symbolic mathematical ability. Sharper ANS acuity is linked to better math skill across the lifespan (Chen & Li, 2014; Schneider et al., 2016). Less research has focused on the ability to perform approximate, non-symbolic calculation with ANS representations. There is growing evidence that a wide variety of math concepts can be represented using the ANS, from addition and subtraction (Barth et al., 2006, 2005; McCrink & Wynn, 2004; Pica et al., 2004) to ratio processing (Falk et al., 2012; Matthews et al., 2016; Ruggeri et al., 2018), basic algebra equations (Kibbe & Feigenson, 2017), number line placement (Honoré & Noël, 2016), and scaling operations (McCrink et al., 2016). Chapter 2 of the current dissertation has added to this list with the first evidence of non-symbolic, approximate division in children and adults. Moreover, Chapters 1 and 2 describe evidence that children can use intuitive models of arithmetic to perform the same mathematical operations with numerals. This is important evidence that approximate, non-symbolic calculation is not an isolated skill. The non-symbolic calculation concepts learned through the ANS can be applied to symbolic calculation during early math education.

The experiments in this dissertation add more evidence to the idea that approximate, non-symbolic calculation is a better predictor of symbolic math skill than ANS acuity alone (Matthews et al., 2016; Pinheiro-Chagas et al., 2014; Starr et al., 2016). One overarching goal of this dissertation was to examine whether non-symbolic calculation skill can function as a mechanism of the established relation between ANS acuity and symbolic math ability. According to this hypothesis, sharper ANS acuity provides a better sense of the numerical magnitudes involved in non-symbolic computation, which allows for better non-symbolic calculation accuracy. In turn better approximate, non-symbolic calculation allows for the creation of better conceptual models of arithmetic concepts. Consistently across all four chapters of this dissertation sharper ANS acuity was correlated with better non-symbolic calculation accuracy. This was true across multiple non-symbolic operations including ratio comparison in children and adults (Chapters 1 and 4), and division in both children and adults (Chapter 2). Moreover, Chapters 1 and 2 both provide direct evidence that non-symbolic calculation mediates the relation between ANS acuity and symbolic math. In addition, Chapter 1 identifies that non-symbolic calculation functions as a mediator only when the symbolic and non-symbolic calculations tested share conceptual overlap. The affordance of the ANS for calculation may be crucial to the link between the Approximate Number System and symbolic math.

More research is needed to determine whether approximate, non-symbolic calculation tasks are an effective way improve math skill in an educational context. Previous work has found evidence that non-symbolic calculation tasks may be an effective part of an intervention aimed at improving math ability (Dillon et al., 2017; Obersteiner et al., 2013, 2013; Park et al., 2016; Räsänen et al., 2009; Sella et al., 2016; Szkudlarek & Brannon, 2018; Wilson et al., 2009, 2006), but many open questions remain. What type of math problems are best suited to an approximate, non-symbolic representation? Can we use the ANS to represent more advanced math topics in algebra or calculus? How operationally specific are the potential benefits of practice with non-symbolic calculation? Does non-symbolic division practice improve symbolic division calculation? Does it also improve multiplication or subtraction ability? Can practice with one approximate, non-symbolic calculation task improve performance on another? Is there a general non-symbolic calculation skill independent of any individual operation? In addition to this long list of theoretical questions, direct translational work needs to be done in a classroom setting before it is known whether approximate, non-symbolic arithmetic is a useful addition to a math curriculum.

While a large amount of additional research still needs to be done, the experiments in this dissertation can offer suggestions for future experiments that explore the benefits of a non-symbolic, approximate arithmetic intervention for math ability. One theoretical framework from the education literature that may be particularly relevant is the concept of concreteness fading. Concreteness fading is a method of introducing a novel idea to students by beginning with a concrete representation of the idea and then explicitly fading this concept towards abstraction as practice progresses (Fyfe & Nathan, 2018). Chapters 1 and 2 indicate that children can perform approximate math operations in both non-symbolic and symbolic format. In a concreteness fading framework, children could begin practice with non-symbolic approximate tasks and then integrate symbolic, approximate tasks before finally moving toward the ultimate goal of symbolic, exact calculation. An implication of conceptual scaffolding is that making the conceptual connection between non-symbolic and symbolic calculation more explicit should enhance transfer between these skills. Interleaving non-symbolic and symbolic formats of the same approximate computation could make the shared operational context of the tasks more apparent for students.

Not all students will benefit from an approximate, non-symbolic math intervention. Chapter 3 indicates that there may be a particular level of math ability where non-symbolic intervention is no longer effective. Adults may not need conceptual scaffolding on what it means to add or subtract, but perhaps they could benefit from non-symbolic representations of math concepts where they have a weaker understanding, such as fraction processing (Fazio et al., 2015). Different non-symbolic calculation tasks may be best introduced at particular points in math education. For example, non-symbolic addition and subtraction may be most effective in preschool to 1st grade, but non-symbolic ratio and division may be more beneficial once children have some basic math knowledge. One benefit of non-symbolic calculation tasks is that they are designed to utilize a system of representing number that even newborn infants possess (Izard et al., 2009). By tapping into a child's underlying numerical competence, ANS calculation problems are accessible for children before they are developmentally capable of performing the same math operations symbolically. Earlier conceptual instruction could allow children to grasp abstract mathematical concepts before they begin learning algorithmic procedures to compute.

Non-symbolic calculation may be particularly helpful for groups of students who struggle with learning math under the current pedagogy. For example, utilizing a child's intuitive sense of number may be helpful for students with a math specific learning disability who nevertheless have intact ANS acuity (Skagerlund & Träff, 2016). Chapter 4 indicated that non-symbolic calculation may not provoke the domain general impairments related to math anxiety. Students high in math anxiety who struggle to use the symbolic math system may benefit from a non-symbolic introduction to math concepts. For all students, ANS calculation tasks can promote thinking about math problems in an alternative way. Students who are able to generate more self-explanations of a concept are better able to acquire and use this knowledge (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Rittle-Johnson, Loehr, & Durkin, 2017). Until the advent of computers and tablets in the classroom it was difficult to use ANS based tasks with groups of students in school. Approximate, non-symbolic calculation requires large numbers of objects that change on every trial. For example, it is impossible to recreate the over fifty trials of non-symbolic division that use upwards of 100 individual items across one trial with physical manipulatives. Due to this nature of ANS calculation tasks, non-symbolic and approximate calculation is a largely untested and unused way to teach children basic math operations. Future work should investigate whether children take away different ideas about a math concept when it is represented with small groups of physical manipulatives versus ANS based computerized tasks with large numbers of objects. Overall, approximate non-symbolic calculation may be most beneficial for young children who do not yet have symbolic math skills, special populations of students with learning disabilities, or as a way of encouraging alternative self-explanations of math concepts for all learners.

Finally, there is a large amount of variance in non-symbolic calculation skill that is not attributable to ANS acuity alone. Calculating non-symbolically requires other domain general cognitive skills, such as working memory and executive function, to successfully carry out a non-symbolic computation (Xenidou-Dervou et al., 2014). Perhaps, this affordance of non-symbolic computation to combine basic numerical intuitions with other cognitive skills known to be linked to math ability is a beneficial aspect of non-symbolic calculation. Combining working memory training with math tasks improves math skill more than working memory training alone (Nemmi et al., 2016). This use of domain general skills in addition to the ANS during non-symbolic calculation may be one reason why non-symbolic calculation is a better target as a math intervention than increasing ANS acuity alone (Cochrane et al., 2018; Maertens et al., 2016; Szűcs & Myers, 2017). Future work should explore how domain general cognitive skills impact non-symbolic math computation, and whether these aspects of the task are central to their potential ability to improve math skills.

Conclusions

The experiments described in this dissertation provide new insights into the relation between non-symbolic, approximate mathematical calculation and symbolic mathematics. This work provides evidence of novel approximate calculation abilities, a potential mechanism of the relation between the ANS and math, and suggestions for populations of students that could most benefit from non-symbolic arithmetic practice. Acquiring competence in mathematical calculation is frequently a difficult task for children in school, and math difficulties and anxiety can persist into adulthood. At the same time, students begin math education with the ability to intuitively perform complex arithmetic calculations across large quantities of objects. Encouraging students to harness the power of their intuitive sense of arithmetic can empower children to realize their own potential for mathematical thought.

APPENDIX

Appendix A.1. Zero order correlations of the measures collected in Chapter 1

	1	*2*	*3*	*4*	*5*
1. ANS acuity					
2. Non-symbolic Ratio Comparison	-0.32**				
3. Symbolic Ratio Comparison	-0.29*	0.62***			
4. Key-Math 3 Numeration	-0.54***	0.50***	0.55***		
5. Key-Math 3 Data Analysis & Probability	-0.36**	0.50***	0.45***	0.68***	
6. Woodcock-Johnson Reading Custer	-0.24	0.31**	0.37**	0.49***	0.44***

*Note: *** p<.001 ** p < .01 * p < .05*

Appendix A.2. Stimuli used during both the non-symbolic and symbolic ratio comparison tasks. A trial is congruent if the array with the greater number of total items is also the array with the better ratio of preferred to non-preferred items. A trial is unfavorable-favorable (UF) if one array has more preferred items and the other array has more non-preferred items. A trial is favorable-favorable (FF) if both arrays have more preferred items. A trial is unfavorable-unfavorable (UU) if both arrays have more non-preferred items. The ratios of ratios is defined as the ratio of preferred to non-preferred in machine 1/ratio of preferred to non-preferred in machine 2.

Left Preferred Color	Left non-preferred Color	Right preferred Color	Right non-preferred Color	Congruence	Favorability	Ratio of Ratios
12	10	4	5	1	UF	1.5
6	5	12	15	0	UF	1.5
15	6	5	3	1	FF	1.5
5	2	10	6	0	FF	1.5
6	16	2	8	1	UU	1.5
3	8	5	20	0	UU	1.5
25	15	12	18	1	UF	2.5
10	6	14	21	0	UF	2.5
20	2	4	1	1	FF	2.5
10	1	12	3	0	FF	2.5
10	22	2	11	1	UU	2.5
5	11	6	33	0	UU	2.5
20	12	5	9	1	UF	3
5	3	10	18	0	UF	3
14	4	7	6	1	FF	3
7	2	14	12	0	FF	3
9	21	1	7	1	UU	3
3	7	2	14	0	UU	3
21	15	4	10	1	UF	3.5
7	5	8	20	0	UF	3.5
28	6	4	3	1	FF	3.5
14	3	20	15	0	FF	3.5
14	22	2	11	1	UU	3.5
7	11	4	22	0	UU	3.5
24	6	2	5	1	UF	10
4	1	6	15	0	UF	10
24	2	6	5	1	FF	10
12	1	18	15	0	FF	10
15	21	1	14	1	UU	10
5	7	2	28	0	UU	10

Appendix B.1. Experiment 1 stimuli for both children and adults in Chapter 2.

Dividend	Divisor	Target	Ratio Level	Phase
14	2	25	N/A	Demo
45	5	3	N/A	Demo
24	2	3	N/A	Demo
75	5	4	N/A	Demo
40	8	18	N/A	Demo
24	8	11	N/A	Demo
25	5	18	N/A	Demo
48	8	2	N/A	Demo
44	2	29	1	feedback
85	5	22	1	feedback
145	5	37	1	feedback
80	8	13	1	feedback
44	2	17	1	feedback
110	5	17	1	feedback
65	5	10	1	feedback
136	8	13	1	feedback
44	2	37	2	feedback
65	5	22	2	feedback
104	8	22	2	feedback
136	8	29	2	feedback
44	2	13	2	feedback
110	5	13	2	feedback
136	8	10	2	feedback
176	8	13	2	feedback
34	2	37	3	feedback
44	2	48	3	feedback
40	5	17	3	feedback
136	8	37	3	feedback

74	2	17	3	feedback
58	2	13	3	feedback
110	5	10	3	feedback
136	8	8	3	feedback
34	2	48	4	feedback
65	5	37	4	feedback
64	8	22	4	feedback
136	8	48	4	feedback
74	2	13	4	feedback
185	5	13	4	feedback
145	5	10	4	feedback
176	8	8	4	feedback
66	3	29	1	no feedback
102	6	22	1	no feedback
60	6	13	1	no feedback
51	3	13	1	no feedback
144	3	37	1	no feedback
102	6	13	1	no feedback
39	3	22	2	no feedback
51	3	29	2	no feedback
102	6	29	2	no feedback
111	3	22	2	no feedback
78	6	8	2	no feedback
174	6	17	2	no feedback
66	3	48	3	no feedback
132	6	48	3	no feedback
78	6	29	3	no feedback
66	3	10	3	no feedback
51	3	8	3	no feedback
174	6	13	3	no feedback

39	3	37	4	no feedback
51	3	48	4	no feedback
60	6	29	4	no feedback
87	3	10	4	no feedback
174	6	10	4	no feedback

Appendix B.2. Stimuli for Experiment 2 with children in Chapter 2.

Dividend	Divisor	Target	Ratio Level	Phase
14	2	25	N/A	Demo
45	5	3	N/A	Demo
24	2	3	N/A	Demo
75	5	4	N/A	Demo
40	8	18	N/A	Demo
24	8	11	N/A	Demo
25	5	18	N/A	Demo
48	8	2	N/A	Demo
58	2	48	2	feedback
58	2	17	2	feedback
74	2	22	2	feedback
34	2	37	3	feedback
44	2	48	3	feedback
74	2	17	3	feedback
96	2	22	3	feedback
34	2	48	4	feedback
96	2	17	4	feedback
58	2	10	4	feedback
74	2	13	4	feedback
40	5	10	1	feedback
50	5	13	1	feedback
145	5	22	1	feedback
85	5	13	1	feedback
40	5	13	2	feedback
85	5	29	2	feedback
145	5	17	2	feedback
185	5	22	2	feedback
110	5	48	3	feedback

185	5	17	3	feedback
145	5	13	3	feedback
85	5	48	4	feedback
64	8	10	1	feedback
80	8	13	1	feedback
176	8	17	1	feedback
80	8	8	1	feedback
64	8	13	2	feedback
136	8	37	3	feedback
104	8	37	4	feedback
136	8	48	4	feedback
176	8	8	4	feedback
40	4	13	1	no feedback
60	6	13	1	no feedback
32	4	13	2	no feedback
48	6	13	2	no feedback
87	3	37	1	no feedback
39	3	10	1	no feedback
40	4	8	1	no feedback
87	3	48	2	no feedback
78	6	8	2	no feedback
66	3	48	3	no feedback
102	6	37	3	no feedback
132	6	48	3	no feedback
39	3	37	4	no feedback
51	3	48	4	no feedback
102	6	48	4	no feedback
132	6	8	4	no feedback
174	6	22	1	no feedback
111	3	22	2	no feedback

148	4	22	2	no feedback
144	3	22	3	no feedback
148	4	17	3	no feedback
192	4	22	3	no feedback
192	4	17	4	no feedback
192	4	17	4	no feedback

Appendix B.3. The questions from the formal division test given to children in chapter 2. This script was read to children and the experimenter marked down the answer given.

1. Do you know what 2 plus 3 is?

2. If Sam has four apples and Kate gives two more how many apples will he have?

3. Do you know what half of 4 is?

4. How about half of 18?

5. What if you ordered a pizza and it had 8 slices? If there were 4 of you who wanted to share the pizza how many slices would you each get?

6. The Football Factory makes 49 footballs per week. If the factory is open 7 days a week, how many footballs do they make per day?

7. Do you know what this symbol is? ÷

Can you solve any of these problems?

6 + 3 = ?

8 + 2 = ?

5 + 5 = ?

12 + 4 = ?

6÷3 = ?

8 ÷ 4 = ?

24÷3 = ?

45÷5 = ?

Appendix B.4. Adult division strategy questionnaire.

1. How did you solve the task with the bee, flowers, and DOTS? Did you use any specific strategies?
2. How did you solve the task with the bee, flowers, and NUMBERS? Did you use any specific strategies?
3. Which task was more difficult, the flowers & bees task with DOTS or with NUMBERS?
4. How many petals appeared on a flower? (circle all that you remember)

   1 2 3 4 5 6 7 8 9 10 11 12

5. What mathematical operation did the task with the flowers and bees test?
6. What differences were there between parts 1 and parts 2 of both flowers and bees games?
7. Which strategy best describes how you solved the task with the bee, flowers, and DOTS? (Circle the answer that best describes your strategy)

a) I got a sense of the amount of dots on the top of the screen and the amount of petals below, and imagined approximately how many dots were on one petal.

b) I got a sense of the amount of dots on the top of the screen and counted the number of petals and imagined approximately how many dots were on one petal.

c) I estimated the number of dots on the top of the screen and assigned a number to my estimate. I counted the number of flower petals. Then I calculated using those numbers.

d) Other, please explain

8. Which strategy best describes how you solved the task with the bee, flowers, and NUMBERS? (Circle the answer that best describes your strategy)

a) I looked at the number up top and estimated the number of petals below, and imagined approximately how many dots were on one petal.

b) I looked at the number up top and counted the number of petals and imagined approximately how many dots were on one petal.

c) I calculated the exact answer using the number up top and the exact number of petals.

d) Other, please explain

BIBLIOGRAPHY

Ahl, V. A., Moore, C. F., & Dixon, J. A. (1992). Development of intuitive and numerical proportional reasoning. *Cognitive Development*, *7*(1), 81–108.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*(2), 224–237. https://doi.org/10.1037/0096-3445.130.2.224

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*, *14*(2), 243–248. https://doi.org/10.3758/BF03194059

Ashcraft, M. H., & Moore, A. M. (2009). Mathematics Anxiety and the Affective Drop in Performance. *Journal of Psychoeducational Assessment*, *27*(3), 197–205. https://doi.org/10.1177/0734282908330580

Au, J., Jaeggi, S. M., & Buschkuehl, M. (2018). Effects of non-symbolic arithmetic training on symbolic arithmetic and the approximate number system. *Acta Psychologica*, *185*, 1–12. https://doi.org/10.1016/j.actpsy.2018.01.005

Banks, J., & Oldfield, Z. (2007). Understanding pensions: Cognitive function, numerical ability and retirement saving. *Fiscal Studies*, 143–170.

Barth, H., Baron, A., Spelke, E., & Carey, S. (2009). Children’s multiplicative transformations of discrete and continuous quantities. *Journal of Experimental Child Psychology*, *103*(4), 441–454. https://doi.org/10.1016/j.jecp.2009.01.014

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*(3), 199–222. https://doi.org/10.1016/j.cognition.2004.09.011

Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 14116–14121.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, *107*(5), 1860–1863. https://doi.org/10.1073/pnas.0910967107

Beilock, Sian L, & Carr, T. H. (2005). When High-Powered People Fail. *Psychological Science*, *16*(2), 5.

Beilock, Sian L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 983–998. https://doi.org/10.1037/0278-7393.33.6.983

Beilock, Sian L., & Maloney, E. A. (2015). Math Anxiety A Factor in Math Achievement Not to Be Ignored. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 4–12.

Betz, N. E. (1978). Prevalence, Distribution, and Correlates of Math Anxiety in College Students. *Journal of Counseling Psychology*, *25*, 441–448.

Blake, P. R., & McAuliffe, K. (2011). "I had so much it didn't seem fair": Eight-year-olds reject two forms of inequity. *Cognition*, *120*(2), 215–224. https://doi.org/10.1016/j.cognition.2011.04.006

Blöte, A. W., Lieffering, L. M., & Ouwehand, K. (2006). The development of many-to-one counting in 4-year-old children. *Cognitive Development*, *21*(3), 332–348. https://doi.org/10.1016/j.cogdev.2005.10.002

Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology*, *114*(3), 375–388. https://doi.org/10.1016/j.jecp.2012.09.015

Boyer, T. W., & Levine, S. C. (2012). Child proportional scaling: Is 1/3=2/6=3/9=4/12? *Journal of Experimental Child Psychology*, *111*(3), 516–533. https://doi.org/10.1016/j.jecp.2011.11.001

Boyer, T. W., & Levine, S. C. (2015). Prompting Children to Reason Proportionally: Processing Discrete Units as Continuous Amounts. *Developmental Psychology*. https://doi.org/10.1037/a0039010

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, *44*(5), 1478–1490. https://doi.org/10.1037/a0013110

Braham, E. J., & Libertus, M. E. (2018). When approximate number acuity predicts math performance: The moderating role of math anxiety. *PLOS ONE*, *13*(5), e0195696. https://doi.org/10.1371/journal.pone.0195696

Bugden, S., Woldorff, M. G., & Brannon, E. M. (2019). Shared and distinct neural circuitry for nonsymbolic and symbolic double-digit addition. *Human Brain Mapping*, *40*(4), 1328–1343. https://doi.org/10.1002/hbm.24452

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

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Campbell, J. I. (1997). On the relation between skilled performance of simple division and multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1140.

Cantin, R. H., Gnaedinger, E. K., Gallaway, K. C., Hesson-McInnis, M. S., & Hund, A. M. (2016). Executive functioning predicts reading, mathematics, and theory of mind during the elementary years. *Journal of Experimental Child Psychology*, *146*, 66–78. https://doi.org/10.1016/j.jecp.2016.01.014

Cantlon, J. F., Merritt, D. J., & Brannon, E. M. (2015). Monkeys display classic signatures of human symbolic arithmetic. *Animal Cognition*. https://doi.org/10.1007/s10071-015-0942-5

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, *105*(2), 380–400. https://doi.org/10.1037/a0031084

Carpenter, T. P., Ansell, E., Franke, M. L., Fennema, E., & Weisbeck, L. (1993). Models of Problem Solving: A Study of Kindergarten Children's Problem-Solving Processes. *Journal for Research in Mathematics Education*, *24*(5), 428. https://doi.org/10.2307/749152

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. https://doi.org/10.1016/j.actpsy.2014.01.016

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145–182.

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

Clarke, D. M., & Roche, A. (2009). Students' fraction comparison strategies as a window into robust understanding and possible pointers for instruction. *Educational Studies in Mathematics*, *72*(1), 127–138. https://doi.org/10.1007/s10649-009-9198-9

Cochrane, A., Cui, L., Hubbard, E. M., & Green, C. S. (2018). "Approximate number system" training: A perceptual learning approach. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-018-01636-w

Correa, J., Bryant, P., & Nunes, T. (1998). Young Children's Understanding of Division: The Relationship Between Division Terms in a Noncomputational Task. *Journal of Educational Psychology*, *90*(2), 321–329.

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-Gender Stereotypes in Elementary School Children: Gender Stereotypes. *Child Development*, *82*(3), 766–779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

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

Dehaene, S. (1997). The number sense: How the mind creates mathematics. Oxford University Press.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three Parietal Circuits for Number Processing. *Cognitive Neuropsychology*, *20*(3–6), 487–506. https://doi.org/10.1080/02643290244000239

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, *130*(3), 335–347. https://doi.org/10.1016/j.cognition.2013.12.001

Desoete, A., Ceulemans, A., De Weerdt, F., & Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study: Mathematical learning disabilities in kindergarten. *British Journal of Educational Psychology*, *82*(1), 64–81. https://doi.org/10.1348/2044-8279.002002

Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions*, *8*(1), 33. https://doi.org/10.1186/1744-9081-8-33

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247–265. https://doi.org/10.1016/j.cognition.2015.05.016

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00068

DeWolf, M., & Vosniadou, S. (2015). The representation of fraction magnitudes and the whole number bias reconsidered. *Learning and Instruction*, *37*, 39–49. https://doi.org/10.1016/j.learninstruc.2014.07.002

Dietrich, J. F., Huber, S., Moeller, K., & Klein, E. (2015). The influence of math anxiety on symbolic and non-symbolic magnitude processing. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01621

Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S., & Duflo, E. (2017). Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics. *Science*, *357*(6346), 47–55.

Drucker, C. B., Rossa, M. A., & Brannon, E. M. (2015). Comparison of discrete ratios by rhesus macaques (Macaca mulatta). *Animal Cognition*. https://doi.org/10.1007/s10071-015-0914-9

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Eckert, J., Call, J., Hermes, J., Herrmann, E., & Rakoczy, H. (2018). Intuitive statistical inferences in chimpanzees and humans follow Weber's law. *Cognition*, *180*, 99–107. https://doi.org/10.1016/j.cognition.2018.07.004

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. Education Testing Service: Princeton N.J.

Falk, R., & Wilkening, F. (1998). Children's construction of fair chances: Adjusting probabilities. *Developmental Psychology*, *34*(6), 1340.

Falk, R., Yudilevich-Assouline, P., & Elstein, A. (2012). Children's concept of probability as inferred from their binary choices—revisited. *Educational Studies in Mathematics*, *81*(2), 207–233. https://doi.org/10.1007/s10649-012-9402-1

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. https://doi.org/10.1016/j.jecp.2014.01.013

Fazio, L. K., DeWolf, M., & Siegler, R. S. (2015). Strategy Use and Strategy Choice in Fraction Magnitude Comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000153

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, *13*(2), 150–156.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links Between the Intuitive Sense of Number and Formal Mathematics Ability. *Child Development Perspectives*, *7*(2), 74–79. https://doi.org/10.1111/cdep.12019

Fujimura, N. (2001). Facilitating children's proportional reasoning: A model of reasoning processes and effects of intervention on strategy change. *Journal of Educational Psychology*, *93*(3), 589–603. https://doi.org/10.1037//0022-0663.93.3.589

Fyfe, E. R., McNeil, N. M., & Borjas, S. (2015). Benefits of "concreteness fading" for children's mathematics understanding. *Learning and Instruction*, *35*, 104–120. https://doi.org/10.1016/j.learninstruc.2014.10.004

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness Fading in Mathematics and Science Instruction: A Systematic Review. *Educational Psychology Review*, *26*(1), 9–25. https://doi.org/10.1007/s10648-014-9249-3

Fyfe, E. R., & Nathan, M. J. (2018). Making "concreteness fading" more concrete as a theory of instruction for promoting transfer. *Educational Review*, 1–20. https://doi.org/10.1080/00131911.2018.1424116

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. Cognition, 44(1), 43–74.

Garland, A., & Low, J. (2014). Addition and subtraction in wild New Zealand robins. *Behavioural Processes*, *109*, 103–110. https://doi.org/10.1016/j.beproc.2014.08.022

Garland, A., Low, J., & Burns, K. C. (2012). Large quantity discrimination by North Island robins (Petroica longipes). *Animal Cognition*, *15*(6), 1129–1140. https://doi.org/10.1007/s10071-012-0537-3

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' Functional Numeracy Is Predicted by Their School Entry Number System Knowledge. *PLoS ONE*, *8*(1), e54651. https://doi.org/10.1371/journal.pone.0054651

Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2017). Early Conceptual Understanding of Cardinality Predicts Superior School-Entry Number-System Knowledge. *Psychological Science*, 0956797617729817.

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences*, *110*(28), 11267–11271. https://doi.org/10.1073/pnas.1220568110

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., … Inglis, M. (2013). Individual Differences in Inhibitory Control, Not Non-Verbal Number Acuity, Correlate with Mathematics Achievement. *PLoS ONE*, *8*(6), e67374. https://doi.org/10.1371/journal.pone.0067374

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature*, *447*(7144), 589–591. https://doi.org/10.1038/nature05850

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*(3), 394–406. https://doi.org/10.1016/j.cognition.2010.02.002

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do Girls Really Experience More Anxiety in Mathematics? *Psychological Science*, *24*(10), 2079–2087. https://doi.org/10.1177/0956797613486989

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*, *48*(5), 1229–1241. https://doi.org/10.1037/a0027433

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles*, *66*(3–4), 153–166. https://doi.org/10.1007/s11199-011-9996-2

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences*, *109*(28), 11116–11120. https://doi.org/10.1073/pnas.1200196109

Halberda, Justin, Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature Letters*, *455*(2).

He, L., Zuo, Z., Chen, L., & Humphreys, G. (2014). Effects of Number Magnitude and Notation at 7T: Separating the Neural Response to Small and Large, Symbolic and Nonsymbolic Number. *Cerebral Cortex*, *24*(8), 2199–2209. https://doi.org/10.1093/cercor/bht074

He, Y., Zhou, X., Shi, D., Song, H., Zhang, H., & Shi, J. (2016). New Evidence on Causal Relationship between Approximate Number System (ANS) Acuity and Arithmetic Ability in Elementary-School Students: A Longitudinal Cross-Lagged Analysis. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01052

Hembree, R. (1990). The Nature, Effects, and Relief of Mathematics Anxiety. *Journal for Research in Mathematics Education*, *21*(1), 33. https://doi.org/10.2307/749455

Hock, H., & Schöner, G. (2010). Measuring Perceptual Hysteresis with the Modified Method of Limits: Dynamics at the Threshold. *Seeing and Perceiving*, *23*(2), 173–195. https://doi.org/10.1163/187847510X503597

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

Honoré, N., & Noël, M.-P. (2016). Improving Preschoolers' Arithmetic through Number Magnitude Training: The Impact of Non-Symbolic and Symbolic Training. *PLOS ONE*, *11*(11), e0166685. https://doi.org/10.1371/journal.pone.0166685

Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Greentree, A. D., & Dyer, A. G. (2019). Numerical cognition in honeybees enables addition and subtraction. *Science Advances*, *5*(2), eaav0961. https://doi.org/10.1126/sciadv.aav0961

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. https://doi.org/10.1016/j.cognition.2013.12.007

Ioannidis, J. P. A. (2008). Why Most Discovered True Associations Are Inflated: *Epidemiology*, *19*(5), 640–648. https://doi.org/10.1097/EDE.0b013e31818131e7

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

Iuculano, T., Tang, J., Hall, C. W. B., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science*, *11*(5), 669–680. https://doi.org/10.1111/j.1467-7687.2008.00716.x

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, *106*(25), 10382–10385.

Jeong, Y., Levine, S. C., & Huttenlocher, J. (2007). The Development of Proportional Reasoning: Effect of Continuous Versus Discrete Quantities. *Journal of Cognition and Development*, *8*(2), 237–256. https://doi.org/10.1080/15248370701202471

Jitendra, A. K., & Hoff, K. (1996). The Effects of Schema-Based Instruction on the Mathematical Word-Problem-Solving Performance of Students with Learning Disabilities. *Journal of Learning Disabilities*, *29*(4), 422–431. https://doi.org/10.1177/002221949602900410

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850–867. https://doi.org/10.1037/a0014939

Kahneman, A. T. A. (1971). Belief in the Law Of Small Numbers. *Psychological Bulletin*, *76*(2), 105–110.

Käser, T., Baschera, G.-M., Kohn, J., Kucian, K., Richtmann, V., Grond, U., … von Aster, M. (2013). Design and evaluation of the computer-based training program Calcularis for enhancing numerical cognition. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00489

Kayhan, E., Gredebäck, G., & Lindskog, M. (2017). Infants Distinguish Between Two Events Based on Their Relative Likelihood. *Child Development*. https://doi.org/10.1111/cdev.12970

Khanum, S., Hanif, R., Spelke, E. S., Berteletti, I., & Hyde, D. C. (2016). Effects of Non-Symbolic Approximate Number Practice on Symbolic Numerical Abilities in Pakistani Children. *PloS One*, *11*(10), e0164436.

Kibbe, M. M., & Feigenson, L. (2015). Young children 'solve for *x* ' using the Approximate Number System. *Developmental Science*, *18*(1), 38–49. https://doi.org/10.1111/desc.12177

Kibbe, M. M., & Feigenson, L. (2017). A dissociation between small and large numbers in young children's ability to "solve for x" in non-symbolic math problems. *Cognition*, *160*, 82–90. https://doi.org/10.1016/j.cognition.2016.12.006

Klein, E., Moeller, K., Dressel, K., Domahs, F., Wood, G., Willmes, K., & Nuerk, H.-C. (2010). To carry or not to carry — Is this the question? Disentangling the carry effect in multi-digit addition. *Acta Psychologica*, *135*(1), 67–76. https://doi.org/10.1016/j.actpsy.2010.06.002

Knops, A., Viarouge, A., & Dehaene, S. (2009). Dynamic representations underlying symbolic and nonsymbolic calculation: Evidence from the operational momentum effect. *Attention, Perception, & Psychophysics*, *71*(4), 803–821. https://doi.org/10.3758/APP.71.4.803

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

Kuhn, J.-T., & Holling, H. (2014). Number sense or working memory? The effect of two computer-based trainings on mathematical skills in elementary school. *Advances in Cognitive Psychology*, *10*(2), 59–67. https://doi.org/10.5709/acp-0157-2

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, *19*(3), 355–365. https://doi.org/10.1016/j.lindif.2008.10.009

Lindskog, M., Winman, A., Juslin, P., & Poom, L. (2013). Measuring acuity of the approximate number system reliably and validly: The evaluation of an adaptive test procedure. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00510

Lindskog, M., Winman, A., & Poom, L. (2017). Individual differences in nonverbal number skills predict math anxiety. *Cognition*, *159*, 156–162. https://doi.org/10.1016/j.cognition.2016.11.014

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences*, *109*(46), 18737–18742. https://doi.org/10.1073/pnas.1207212109

Lussier, C. A., & Cantlon, J. F. (2016). Developmental bias for number words in the intraparietal sulcus. *Developmental Science*, n/a-n/a. https://doi.org/10.1111/desc.12385

Lyons, I. M., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain: Neural Coding of Numbers. *Human Brain Mapping*, *36*(2), 475–488. https://doi.org/10.1002/hbm.22641

Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., & Reynvoet, B. (2016). Enhancing arithmetic in pre-schoolers with comparison or number line estimation training: Does it matter? *Learning and Instruction*, *46*, 1–11. https://doi.org/10.1016/j.learninstruc.2016.08.004

Maloney, E. A., Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2015). Intergenerational Effects of Parents’ Math Anxiety on Children’s Math Achievement and Anxiety. *Psychological Science*. https://doi.org/10.1177/0956797615592630

Maloney, Erin A., Ansari, D., & Fugelsang, J. A. (2011). Rapid Communication: The effect of mathematics anxiety on the processing of numerical magnitude. *Quarterly Journal of Experimental Psychology*, *64*(1), 10–16. https://doi.org/10.1080/17470218.2010.533278

Maloney, Erin A., Risko, E. F., Ansari, D., & Fugelsang, J. (2010). Mathematics anxiety affects counting but not subitizing during visual enumeration. *Cognition*, *114*(2), 293–297. https://doi.org/10.1016/j.cognition.2009.09.013

Matthews, P. G., Lewis, M. R., & Hubbard, E. M. (2016). Individual Differences in Nonsymbolic Ratio Processing Predict Symbolic Math Performance. *Psychological Science*, *27*(2), 191–202. https://doi.org/10.1177/0956797615617799

Mauro, D. G., LeFevre, J.-A., & Morris, J. (2003). Effects of problem format on division and multiplication performance: Division facts are mediated via multiplication-based representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 163–170. https://doi.org/10.1037/0278-7393.29.2.163

McCrink, K., Shafto, P., & Barth, H. (2016). The relationship between non-symbolic multiplication and division in childhood. *The Quarterly Journal of Experimental Psychology*, 1–17. https://doi.org/10.1080/17470218.2016.1151060

McCrink, K., & Spelke, E. S. (2010). Core multiplication in childhood. *Cognition*, *116*(2), 204–216. https://doi.org/10.1016/j.cognition.2010.05.003

McCrink, K., & Spelke, E. S. (2016). Non-symbolic division in childhood. *Journal of Experimental Child Psychology*, *142*, 66–82. https://doi.org/10.1016/j.jecp.2015.09.015

McCrink, K., Spelke, E. S., Dehaene, S., & Pica, P. (2013). Non-symbolic halving in an Amazonian indigene group. *Developmental Science*, *16*(3), 451–462. https://doi.org/10.1111/desc.12037

McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science*, *15*(11), 776–781.

McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science*, *18*(8), 740–745.

McMullan, M., Jones, R., & Lea, S. (2012). Math anxiety, self-efficacy, and ability in British undergraduate nursing students. *Research in Nursing & Health*, *35*(2), 178–186. https://doi.org/10.1002/nur.21460

McNeil, N. M., Fuhs, M. W., Keultjes, M. C., & Gibson, M. H. (2011). Influences of problem format and SES on preschoolers' understanding of approximate addition. *Cognitive Development*, *26*(1), 57–71. https://doi.org/10.1016/j.cogdev.2010.08.010

Meece, J. L., Eccles, J. S., & Wigfield, A. (1990). Predictors of Math Anxiety and Its Influence on Young Adolescents' Course Enrollment Intentions and Performance in Mathematics. *Journal of Educational Psychology*, *82*(1), 60–70.

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419. https://doi.org/10.1037/a0024377

Moyer, R. S., & Landauer, T. K. (1967). *Time required for judgements of numerical inequality*. *215*(5109), 1519.

Nemmi, F., Helander, E., Helenius, O., Almeida, R., Hassler, M., Räsänen, P., & Klingberg, T. (2016). Behavior and neuroimaging at baseline predict individual response to combined mathematical and working memory training in children. *Developmental Cognitive Neuroscience*, *20*, 43–51. https://doi.org/10.1016/j.dcn.2016.06.004

Ni, Y., & Zhou, Y.-D. (2005). Teaching and Learning Fraction and Rational Numbers: The Origins and Implications of Whole Number Bias. *Educational Psychologist*, *40*(1), 27–52. https://doi.org/10.1207/s15326985ep4001_3

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A Two-Minute Paper-and-Pencil Test of Symbolic and Nonsymbolic Numerical Magnitude Processing Explains Variability in Primary School Children’s Arithmetic Competence. *PLoS ONE*, *8*(7), e67918. https://doi.org/10.1371/journal.pone.0067918

Núñez-Peña, M. I., & Suárez-Pellicioni, M. (2014). Less precise representation of numerical magnitude in high math-anxious individuals: An ERP study of the size and distance effects. *Biological Psychology*, *103*, 176–183. https://doi.org/10.1016/j.biopsycho.2014.09.004

Obersteiner, A., Bernhard, M., & Reiss, K. (2015). Primary school children’s strategies in solving contingency table problems: The role of intuition and inhibition. *ZDM*, *47*(5), 825–836. https://doi.org/10.1007/s11858-015-0681-8

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students’ basic number processing and arithmetic skills. *Learning and Instruction*, *23*, 125–135. https://doi.org/10.1016/j.learninstruc.2012.08.004

Olsson, L., Östergren, R., & Träff, U. (2016). Developmental dyscalculia: A deficit in the approximate number system or an access deficit? *Cognitive Development*, *39*, 154–167. https://doi.org/10.1016/j.cogdev.2016.04.006

Park, J., & Brannon, E. M. (2013). Training the Approximate Number System Improves Math Proficiency. *Psychological Science*, *24*(10), 2013–2019. https://doi.org/10.1177/0956797613482944

Park, Joonkoo, Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of Experimental Child Psychology*, *152*, 278–293. https://doi.org/10.1016/j.jecp.2016.07.011

Park, Joonkoo, & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition*, *133*(1), 188–200. https://doi.org/10.1016/j.cognition.2014.06.011

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* London: National Research and Development Centre for Adult Literacy and Numeracy.

Perdue, B. M., Talbot, C. F., Stone, A. M., & Beran, M. J. (2012). Putting the elephant back in the herd: Elephant relative quantity judgments match those of other species. *Animal Cognition*, *15*(5), 955–961. https://doi.org/10.1007/s10071-012-0521-y

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., … Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33–41. https://doi.org/10.1016/j.cognition.2010.03.012

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex. *Neuron*, *53*(2), 293–305. https://doi.org/10.1016/j.neuron.2006.11.022

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503.

Pinheiro-Chagas, P., Wood, G., Knops, A., Krinzinger, H., Lonnemann, J., Starling-Alves, I., … Haase, V. G. (2014). In How Many Ways is the Approximate Number System Associated with Exact Calculation? *PLoS ONE*, *9*(11), e111155. https://doi.org/10.1371/journal.pone.0111155

Purpura, D. J., & Logan, J. A. R. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Developmental Psychology*, *51*(12), 1717–1724. https://doi.org/10.1037/dev0000055

Purpura, D. J., & Simms, V. (2018). Approximate number system development in preschool: What factors predict change? *Cognitive Development*, *45*, 31–39. https://doi.org/10.1016/j.cogdev.2017.11.001

Rakoczy, H., Clüver, A., Saucke, L., Stoffregen, N., Gräbener, A., Migura, J., & Call, J. (2014). Apes are intuitive statisticians. *Cognition*, *131*(1), 60–68. https://doi.org/10.1016/j.cognition.2013.12.011

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology*, *141*, 83–100. https://doi.org/10.1016/j.jecp.2015.07.014

Ramirez, G., Fries, L., Gunderson, E., Schaeffer, M. W., Maloney, E. A., Beilock, S. L., & Levine, S. C. (2019). Reading Anxiety: An Early Affective Impediment to Children's Success in Reading. *Journal of Cognition and Development*, *20*(1), 15–34. https://doi.org/10.1080/15248372.2018.1526175

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math Anxiety, Working Memory, and Math Achievement in Early Elementary School. *Journal of Cognition and Development*, *14*(2), 187–202. https://doi.org/10.1080/15248372.2012.664593

Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development*, *24*(4), 450–472. https://doi.org/10.1016/j.cogdev.2009.09.003

Richardson, F. C., & Suinn, R. M. (1972). The Mathematics Anxiety Rating Scale: Psychometric data. *Journal of Counseling Psychology*, *19*(6), 551–554. https://doi.org/10.1037/h0033456

Riley, M. (1984). *Schema Knowledge Structures for Representing and Understanding Arithmetic Story Problems.* 74.

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM*, 1–13.

Rosenberg-Lee, M., Chang, T. T., Young, C. B., Wu, S., & Menon, V. (2011). Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: A cytoarchitectonic mapping study. *Neuropsychologia*, *49*(9), 2592–2608. https://doi.org/10.1016/j.neuropsychologia.2011.04.035

Ruggeri, A., Vagharchakian, L., & Xu, F. (2018). Icon arrays help younger children’s proportional reasoning. *British Journal of Developmental Psychology*. https://doi.org/10.1111/bjdp.12233

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science*, *26*(6), 515–520.

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, *114*(3), 418–431. https://doi.org/10.1016/j.jecp.2012.10.012

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, n/a-n/a. https://doi.org/10.1111/desc.12372

Sella, F., Tressoldi, P., Lucangeli, D., & Zorzi, M. (2016). Training numerical skills with the adaptive videogame “The Number Race”: A randomized controlled trial on preschoolers. *Trends in Neuroscience and Education*, *5*(1), 20–29. https://doi.org/10.1016/j.tine.2016.02.002

Shaklee, H., & Paszek, D. (1985). Covariation Judgment: Systematic Rule Use in Middle Childhood. *Child Development*, *56*(5), 1229–1240.

Shaw, A., & Olson, K. R. (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General*, *141*(2), 382–395. https://doi.org/10.1037/a0025907

Sheskin, M., Nadal, A., Croom, A., Mayer, T., Nissel, J., & Bloom, P. (2016). Some Equalities Are More Equal Than Others: Quality Equality Emerges Later Than Numerical Equality. *Child Development*, *87*(5), 1520–1528. https://doi.org/10.1111/cdev.12544

Shusterman, A., Slusser, E., Halberda, J., & Odic, D. (2016). Acquisition of the Cardinal Principle Coincides with Improvement in Approximate Number System Acuity in Preschoolers. *PLOS ONE*, *11*(4), e0153072. https://doi.org/10.1371/journal.pone.0153072

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., … Chen, M. (2012). Early Predictors of High School Mathematics Achievement. *Psychological Science*, *23*(7), 691–697. https://doi.org/10.1177/0956797612440101

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: The new frontier for theories of numerical development. *Trends in Cognitive Sciences*, *17*(1), 13–19. https://doi.org/10.1016/j.tics.2012.11.004

Siegler, R. S., Strauss, S., & Levin, I. (1981). Developmental Sequences within and between Concepts. *Monographs of the Society for Research in Child Development*, *46*(2), 1. https://doi.org/10.2307/1165995

Singer-Freeman, K. E., & Goswami, U. (2001). Does half a pizza equal half a box of chocolates? Proportional matching in an analogy task. *Cognitive Development*, *16*, 811–829.

Skagerlund, K., & Träff, U. (2016). Number processing and heterogeneity of developmental dyscalculia: Subtypes with different cognitive profiles and deficits. *Journal of Learning Disabilities*, 0022219414522707.

Sophian, C., & Madrid, S. (2003). Young Children's Reasoning about Many-to-One Correspondences. *Child Development*, *74*(5), 1418–1432. https://doi.org/10.1111/1467-8624.00615

Soto-Calvo, E., Simmons, F. R., Willis, C., & Adams, A.-M. (2015). Identifying the cognitive predictors of early counting and calculation skills: Evidence from a longitudinal study. *Journal of Experimental Child Psychology*, *140*, 16–37. https://doi.org/10.1016/j.jecp.2015.06.011

Spinillo, A. G., & Bryant, P. (1991). Children's Proportional Judgments: The Importance of "Half." *Child Development*, *62*(3), 427. https://doi.org/10.2307/1131121

Starr, A., Roberts, R. C., & Brannon, E. M. (2016). Two Potential Mechanisms Underlying the Link between Approximate Number Representations and Symbolic Math in Preschool Children. *Proceedings of the Cognitive Science Society 2016*. Retrieved from https://www.researchgate.net/profile/Ariel_Starr/publication/308699729_Two_Potential_Mechanisms_Underlying_the_Link_between_Approximate_Number_Representations_and_Symbolic_Math_in_Preschool_Children/links/57ebf30408ae92eb4d263dec.pdf

Swars, S. L., Daane, C. J., & Giesen, J. (2006). Mathematics Anxiety and Mathematics Teacher Efficacy: What is the Relationship in Elementary Preservice Teachers? *School Science and Mathematics*, *106*(7), 306–315. https://doi.org/10.1111/j.1949-8594.2006.tb17921.x

Szkudlarek, E., & Brannon, E. M. (2018). Approximate Arithmetic Training Improves Informal Math Performance in Low Achieving Preschoolers. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00606

Szűcs, D., & Myers, T. (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: A systematic review. *Trends in Neuroscience and Education*, *6*, 187–203. https://doi.org/10.1016/j.tine.2016.11.002

Tecwyn, E. C., Denison, S., Messer, E. J. E., & Buchsbaum, D. (2016). Intuitive probabilistic inference in capuchin monkeys. *Animal Cognition*. https://doi.org/10.1007/s10071-016-1043-9

Toll, S. W., Van Viersen, S., Kroesbergen, E. H., & Van Luit, J. E. (2015). The development of (non-) symbolic comparison skills throughout kindergarten and their relations with basic mathematical skills. *Learning and Individual Differences*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1041608015000023

Vallentin, D., & Nieder, A. (2008). Behavioral and Prefrontal Representation of Spatial Proportions in the Monkey. *Current Biology*, *18*(18), 1420–1425. https://doi.org/10.1016/j.cub.2008.08.042

Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842

van Marle, K., Chu, F. W., Li, Y., & Geary, D. C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Developmental Science*, *17*(4), 492–505. https://doi.org/10.1111/desc.12143

Venkatraman, V., Ansari, D., & Chee, M. W. L. (2005). Neural correlates of symbolic and non-symbolic arithmetic. *Neuropsychologia*, *43*(5), 744–753. https://doi.org/10.1016/j.neuropsychologia.2004.08.005

Wang, Z., Soden, B., Deater-Deckard, K., Lukowski, S. L., Schenker, V. J., Willcutt, E. G., … Petrill, S. A. (2015). Development in reading and math in children from different SES backgrounds: The moderating role of child temperament. *Developmental Science*, n/a-n/a. https://doi.org/10.1111/desc.12380

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's Past Is Prologue: Relations Between Early Mathematics Knowledge and High School Achievement. *Educational Researcher*, *43*(7), 352–360. https://doi.org/10.3102/0013189X14553660

Wigfield, A., & Meece, J. L. (1988). Math Anxiety in Elementary and Secondary School Students. *Journal of Educational Psychology*, *80*, 210–216.

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

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*, *2*(20).

Xenidou-Dervou, I., van Lieshout, E. C. D. M., & van der Schoot, M. (2014). Working Memory in Nonsymbolic Approximate Arithmetic Processing: A Dual-Task Study With Preschoolers. *Cognitive Science*, *38*(1), 101–127. https://doi.org/10.1111/cogs.12053

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015.

Yost, P. A., Siegel, A. E., & Andrews, J. M. (1962). Nonverbal Probability Judgments by Young Children. *Child Development*, *33*(4), 769. https://doi.org/10.2307/1126888

Young, C. B., Wu, S. S., & Menon, V. (2012). The Neurodevelopmental Basis of Math Anxiety. *Psychological Science*, *23*(5), 492–501. https://doi.org/10.1177/0956797611429134