Investigating the Cognitive and Developmental Dynamics of Groupitizing

By

Gillian Sarah Starkey

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

December, 2014

Nashville, Tennessee

Approved:

Bruce D. McCandliss, Ph.D.

Gordon Logan, Ph.D.

Bethany Rittle-Johnson, Ph.D.

Geoffrey Woodman, Ph.D.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Bruce McCandliss, for teaching me how to ask good questions and turn statistics into stories. Over the last five years, you've helped me find my voice as a writer and researcher.

I am so grateful for my dissertation committee members, who have dedicated their time and energy over the last few years to contribute to my research training. Not many graduate students can say they have fond memories of committee meetings, but I do. Thank you for the support and encouragement.

I would also like to thank the many other professors that I've been lucky enough to know, who have influenced and inspired me - especially Jim Steiger, Dave Cordray, and Leslie Rescorla.

This project would not have happened without the help of my amazing Educational Cognitive Neuroscience Labmates. Liane Moneta-Koehler, thank you for always thinking of solutions when I thought there were none. It's been fun making the graduate school journey with you. I am hugely indebted to Chang Gu for the invaluable EEG advice and troubleshooting support, and to Madhu Govind for helping me with recruiting families and taking on REDCap duties. Hilda Fehd, your leadership and encouragement helped me stay organized and stick to my deadlines. I would also like to thank Ed Hubbard for being my role model for passion and dedication to science.

I also must acknowledge the many friends who have helped me stay grounded through the whole dissertation process. Stephen Killingsworth, you are my MATLAB savior. Tim Shaver and Scott Beeler, thank you for listening, for offering your support and perspective, and for always making me laugh. Jackie Fleming and Paige Walker, thank you for holding me to the high standards of Bryn Mawr alumnae, and for reminding me that there's nothing that can't be fixed by some Jane Austen.

I would not be at this point without the support of my incredible family – my mom and dad, distinguished scientists and teachers who have inspired me since day one. Thank you for teaching me the joy of learning, and for always doing whatever you could to support my love of science and math. To my brother-in-law, Christopher: you probably don't remember this, but you once told me, "there are few things I worry about less than you succeeding." That has stuck with me through times of anxiety and uncertainty. To Louis, who is the definition of unconditional love – thank you for keeping my spirits up.

Finally, my husband, Taylor Sundby, who is brilliant, driven, and most importantly, the kindest person I know: You inspire me. Thanks for being my teammate.

This work was supported in part by the Institute of Education Sciences, U.S. Department of Education, Grant #305B080025.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The importance of strong math skills to academic success should not be understated. A recent large-scale meta-analysis of multiple longitudinal data sets concluded that math ability at school entry is the strongest predictor of later school achievement (Duncan et al., 2007). However, the math performance of children and teens in the United States lags substantially behind that of children and teens in other industrialized countries (Mullis, Martin, Foy, & Arora, 2012). Further research has shown that American students' math test scores have not improved over the last 14 years, while several other countries around the world have made significant gains (Hanushek, Peterson, & Woessman, 2012). These poor rankings and lack of progress highlight the inadequacy of the United States education system in training children in critical math skills.

One reason for this is that despite the clear importance of math competence, the core abilities that are most crucial to math development are not well understood. Thus, it is imperative that research in psychology and education establish which early number skills are critical building blocks for higher-level math like arithmetic and algebra. Clarifying the core number skills and aspects of conceptual number knowledge that are essential for math development will contribute to creating school curricula that help teachers effectively teach math to young children.

Background

It is known that children's ability to enumerate, or to report the exact quantity of a set of items, is foundational for later math development. Within the scope of exact enumeration, several specific skills contribute to achievement in the math domain. For example, *subitizing*, the rapid and highly accurate enumeration of small quantities (1 through 3 or 4; Kaufman, Lord, Reese, & Volkmann, 1949), has been identified as a predictor of later math skill (Aunio & Niemivirta, 2010; Clements, 1999; Fischer, Gerhardt, & Hartnegg, 2008). Many studies have established that *counting*, the serial enumeration process required for sets that are larger than the subitizable range, also underlies and predicts higher-level math skills like arithmetic and algebra (Aunio & Niemivirta, 2010; Gallistel & Gelman, 1992; Geary, Bow-Thomas, & Yao, 1992; Kroesbergen, Van Luit, Van Lieshout, Loosbroek, & Van de Rijt, 2009).

More recent findings have suggested that a third distinct exact enumeration skill, referred to as *groupitizing*, may also be foundational for developing math abilities (Starkey & McCandliss, 2014). Groupitizing refers to the process of rapidly enumerating arrays that are spatially grouped into subitizable subgroups. This process is characterized by faster enumeration latencies than for arrays that are not arranged in a grouped structure, requiring a serial counting process (Freeman, 1912; Starkey & McCandliss, 2014; Wender & Rothkegel, 2000). Starkey and McCandliss showed that most early elementary school-aged children were able to groupitize when arrays were arranged into subitizable subgroups. They also demonstrated that individual differences in groupitizing skill accounted for significant and unique variance on a standardized symbolic arithmetic fluency test, suggesting that conceptual number knowledge recruited for groupitizing plays an important role in developing symbolic math skills.

It is evident that spatial grouping has a profound impact on the process of enumeration, and that the process of enumerating a grouped array reflects important core number skills, yet neither the processes that underlie groupitizing nor their development are well understood. Starkey and McCandliss (2014) proposed that groupitizing involves a two-step process of subitizing each subgroup, and then using the combination of subgroup quantities to access the overall quantity of the set, potentially by combining sets. Although Starkey and McCandliss suggested that the subgroups are subitized individually, one at a time, it is possible that subgroups could be subitized in parallel. The idea that adults can process the approximate quantity of multiple large sets in parallel, or simultaneously, has been suggested by Halberda, Sires, and Feigenson (2006) and Feigenson (2008), but it is unclear whether children have this ability, and whether it occurs in the domain of exact enumeration.

The second step of Starkey and McCandliss's proposed groupitizing model would require an awareness of number composition (e.g., that 4 and 2 or 3 and 3 can be combined to form 6, and conversely, that 6 can be broken down into 4 and 2, 3 and 3, etc.), which is a precursor to arithmetic and more complex math (Clements, 1999). This insight that sets can be combined and decomposed emerges when children begin to use "counting on" strategies, in which they start the successor function of counting at the cardinal value of one of the subsets (Siegler, 1987). Set combination skills have previously been linked to more complex math skills (Fuchs et al., 2006; Fuchs et al., 2010; Geary, Bailey & Hoard, 2009; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007), and may play a specific role in the mechanism underlying groupitizing. Groupitizing skills may emerge once children reach this stage, and then improve as children's concept of number develops to include information about number composition. However, further research is needed to further elucidate the conceptual number knowledge involved in groupitizing, and whether subgroups within an array are subitized individually or in parallel.

The experiments in this study aim to characterize the development of groupitizing and its relationship to symbolic arithmetic over the course of elementary school, and to explore the cognitive and neural mechanisms behind groupitizing that both distinguish it from other enumeration skills and support its importance to math development. This study will explore groupitizing as a potential reflection of the integration of children's exact enumeration skills with their insights into the composition of numbers. Specific hypotheses are based upon the proposal that groupitizing involves subitizing subgroups of items, and then accessing overall set size directly from the subgroup quantities.

Definitions

A common obstacle in the successful integration of various fields of research (in this case, math education, cognitive psychology, and cognitive neuroscience) stems from inconsistent terminology. Therefore, it is necessary to define several cognitive processes before reviewing previous research on these topics. This manuscript will use the terms *subitizing*, *counting*, and *groupitizing* in the following ways:

Subitizing refers to the parallel attentive process employed when enumerating a small number of items (Piazza, Giacomini, Le Bihan, & Dehaene, 2003; Vetter, Butterworth, & Bahrami, 2008; 2011). Subitizing is marked by a high level of accuracy, as well as fast reaction times, and a shallow or flat slope across these reaction times (relative to reaction times for enumerating non-subitizable quantities; Kaufman et al., 1949; Simon & Vaishhavi, 1996; Trick & Pylyshyn, 1994). Under conditions in which the array of items is structured in an ungrouped and non-canonical way, the range of subitizable quantities is 1-3 or 1-4 for children, and 1-4 for most adults (Chi & Klahr, 1975; Freeman, 1912; Svenson & Sjöberg, 1983). Because subitizing is a parallel attentive process, the range of quantities that an individual can subitize is determined by the capacity of what the individual can perceive in a single instance of attention and hold in working memory (Ester, Drew, Klee, Vogel, & Awh, 2012; Mazza & Caramazza, 2011; Pagano & Mazza, 2012).

Counting refers to the cognitive process employed when enumerating a number of items that is greater than an individual's range of subitizable quantities. Because not all items in the array can be perceived simultaneously or entirely held in working memory, counting requires serial shifts of attention and updating working memory (Piazza et al., 2003; Railo, Koivisto, Revonsuo, & Hannula, 2008). Compared to subitizing, counting is more error-prone and takes longer (presumably due to the additional cognitive demands required). Therefore, the use of counting instead of subitizing is reflected in higher error rates, longer reaction times, and a steeper slope across these reaction times (Kaufman et al., 1949; Trick & Pylyshyn, 1994).

Groupitizing refers to the enumeration of arrays that are spatially separated into subgroups that fall within an individual's range of subitizable quantities. Groupitizing is distinct from counting, and is marked by reaction times that are faster than the time required for counting an ungrouped array (Starkey & McCandliss, 2014; Wender & Rothkegel, 2000). If a grouped array is enumerated more quickly than an ungrouped array of the same set size, this is an indication that the array was groupitized, rather than counted. If a grouped array and an ungrouped array of equal set sizes require the same amount of time to enumerate, however, we cannot draw conclusions from these reaction times about whether groupitizing or counting was used to enumerate the grouped array (two different processes may take the same amount of time). Supported by previous findings (Wender & Rothkegel), Starkey and McCandliss (2014) suggested that groupitizing is a two-step process of (1) incrementally subitizing and arriving at a quantity for each subgroup, and (2) drawing on conceptual number knowledge to directly access the overall size of the array from the quantities of the subgroups.

In this proposed model, the term *direct access* does not refer to a specific mechanism, but rather to the notion that a connection is made between the subset quantities and the overall array size that does not require processes that are affected by changes in set size, like counting or addition. Thus, groupitizing bypasses the set size-sensitive cognitive mechanisms that are used to enumerate unstructured arrays of items. Instead, groupitizing may recruit conceptual insights into the semantics of numbers, such as the knowledge of the subsets that make up a number. Incorporating this knowledge of set composition into one’s concept of a number may enable this more efficient, direct connection in groupitizing, without relying on set size-sensitive processes.

Enumeration latency is a measure used to study the cognitive dynamics of enumeration skills, and reflects the speed of processing and identifying a quantity. Enumeration latency is measured from the onset of the array to the start of the verbal enumeration response. In the proposed model of groupitizing, enumeration latency reflects the time required to subitize each array, plus the time required to connect the combination of subgroup quantities to the total number of items in the array. Therefore, enumeration latencies for grouped arrays are especially informative when studying whether subitizing occurs individually or in parallel.

Fluency is a second measure used to study the cognitive dynamics of enumeration. Fluency is calculated from the *slope* across set sizes in a given range (Lassaline & Logan, 1993), and therefore controls for individual differences in factors that have the same impact across set sizes (such as visual processing speed and the time required to retrieve and verbally produce a number word). In the proposed model of groupitizing, when the number of subgroups is held constant, slope reflects the cognitive process used to connect subgroup quantities to total array size. Steep slopes indicate a process that is impacted by set size, such as serial counting or addition (Butterworth, Zorzi, Girelli, & Jonckheere, 2001). Flat slopes that approximate zero indicate that an individual can *directly* access total array size from the quantities of subgroups, reflecting that their concept of number is enriched to include information about the composition of number. Therefore, fluency is informative when studying the cognitive processes employed to derive overall quantity after subitizing subgroups. The distinction between enumeration latency and fluency is critical for Experiments 1 and 2.

Behavioral Characteristics of Subitizing

In studying the mechanism underlying groupitizing, it is important to establish that subitizing is a distinct cognitive process from counting. Evidence for the subitizing effect was first published by Stanley Jevons in 1871. Using a “bean counting” task, Jevons demonstrated that he could enumerate up through five beans nearly instantaneously, with “absolute freedom from error” (p. 281), reflecting a single act of perception. However, once he exceeded five beans, his accuracy decreased linearly with each increase in overall quantity. He concluded that, despite having a concept of the numerosities six, seven, and eight, he could not access those representations without relying on an additional process that required more time for each incremental increase in set size. Jevons was likely the first to experimentally demonstrate a difference in how small and large sets are enumerated, and to propose that separate cognitive processes account for each. In 1949, Kaufman et al. coined the term “subitizing” to refer to the enumeration of small sets that is characterized by near-perfect accuracy and reaction times that are significantly lower than for larger sets.

In nearly 150 years since Jevons published this experiment, many studies have investigated the cognitive processes underlying exact enumeration, and how these may be different for small and large quantities. All of these studies have replicated the finding that as set size increases, the time required to enumerate the set increases. Nearly all have also found that there is a discontinuity between the subitizing range and beyond (e.g., Chi & Klahr, 1975; Trick & Pylyshyn, 1994; for an exception, see Balakrishnan & Ashby, 1992). This discontinuity is marked by a shift from a shallow to a steep slope, indicating more time required for each additional item in the array. Reaction time graphs for enumeration appear to follow a bilinear function (Chi & Klahr, 1975; Klahr & Wallace, 1976; Trick & Pylyshyn, 1994) that can be separated into two sections: one line fitted to the slope of subitizable sets, and one fitted to the slope of set sizes beyond the subitizing range (see Figure 1). The upper limit of the subitizing range is typically defined as point at which the two lines of the bilinear function intersect (i.e., the point at which one is no longer able to subitize, and must recruit a serial counting process).

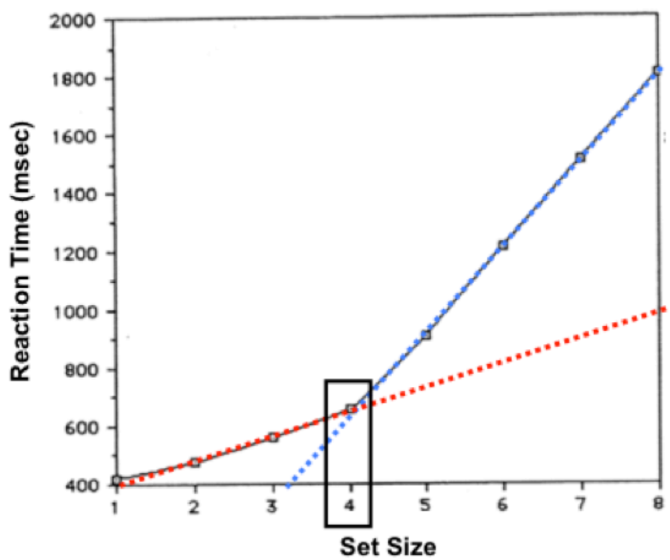

Figure 1. Approximation of upper limit of subitizing range as intersection between lines defined by subitizing range (red) and counting range (blue), adapted from Trick & Pylyshyn, 1994

Because subitizing has been shown to be a part of the mechanism underlying groupitizing (Starkey & McCandliss, 2014; Wender & Rothkegel, 2000), it follows that groupitizing should be affected by subitizing limits. Individuals who have a subitizing range of 1-3 would not be able to subitize a subgroup of 4 items in an array; serial counting would be required in order to enumerate correctly, which would increase enumeration latency. In contrast, individuals who have a subitizing range of 1-4 would be able to subitize all subgroups in this, resulting in a significantly faster enumeration latency relative to an ungrouped array of the same overall set size. The present study will investigate the impact of subitizing capacity on enumeration of arrays with subgroups of different sizes.

Neural Substrates of Subitizing

The differences between subitizing and counting that have been observed in behavioral studies are further supported by several neuroimaging studies that suggest that these two processes have different neural substrates (Ansari, Lyons, van Eimeren, & Xu, 2007; Feigenson, Dehaene, & Spelke, 2004; Starkey, Hubbard, & McCandliss, 2013). For example, an fMRI study conducted by Piazza et al. (2003) demonstrated that regions in the posterior intraparietal sulcus (IPS) typically linked to attentional shifts were activated when subjects enumerated more than 3 or 4 items, but not when they enumerated smaller numerosities. Piazza et al. were able to predict, at a single-trial level, whether the numerosity presented was in the subitizing or counting range, using posterior IPS activation as an index of attentional shift. Piazza et al. concluded that this supports a dual-process model in which subitizing and counting recruit distinct neural mechanisms.

This finding also suggests, more specifically, that subitizing is a parallel attentive process while counting is a serial attentive process. Several other studies, using a variety of behavioral and neuroimaging techniques, have also demonstrated that subitizing is a parallel attentive process, and therefore does not rely on a serial visual or attentive process (Dehaene & Cohen, 1994; Halberda et al., 2006; Pylyshyn & Storm, 1988). Overall, these studies overwhelmingly suggest that when subitizing occurs, all items in an array are perceived simultaneously, enabling direct access to quantity. This account explains the shallow slopes for reaction times across subitizable quantities; if items are perceived simultaneously, one additional item does not require extra time to process, as long as the total set size is still within the subitizable range of quantities.

A number of electrophysiology studies have also supported a dual-process model, demonstrating interesting links between enumeration and event-related potential components typically related to visual attention (Hyde & Spelke, 2011; Pagano & Mazza, 2012; Xu & Liu, 2008). These studies have been especially useful in understanding the basis of the upper limit of the subitizing range. For example, Mazza and Caramazza (2011) reported that the amplitude of the N2pc component, which is considered an index of selective attention in visual search paradigms (Drew & Vogel, 2008), scales with the number of items presented in an enumeration task. Ester et al. (2012) and Pagano and Mazza (2012) followed up on this, showing that individual differences in the asymptote of the N2pc amplitude correspond to behavioral measures of subitizing span (Figure 2). Both of these findings suggest that the limits of what an individual can grasp in an instance of attention may determine how many items can be subitized.

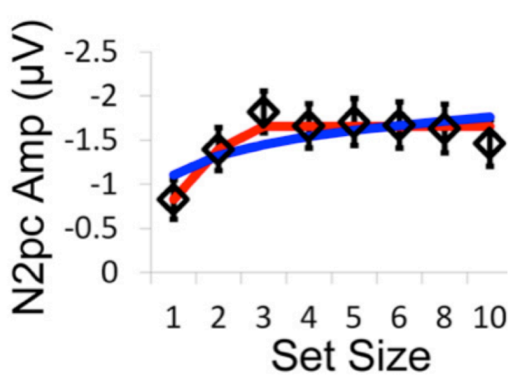

Figure 2. N2pc amplitudes across set size, fit with a bilinear function (red) and a continuous function (blue), from Ester et al., 2012

Additional electrophysiological studies have investigated the CDA component, which is considered an index of working memory capacity (Vogel & Machizawa, 2004), and have demonstrated that this component follows a similar pattern to the N2pc. CDA amplitudes reach a plateau at the point at which individuals' working memory precision begins to diminish (verified with behavioral data by Anderson, Vogel, & Awh, 2011). This suggests that the limits of attention and working memory capacity may both enable subitizing and determine an individual's subitizing slope.

Together, these findings lead to a "fixed-capacity model" of subitizing, in which an individual can access a fixed number of items at any given time (Piazza, Fumarola, Chinello, & Melcher, 2011; Revkin, Piazza, Izard, Cohen, & Dehaene, 2008). Small sets fit within this fixed number of items, and therefore can be represented simultaneously without shifting attention or updating working memory (Fukuda, Vogel, Mayr, & Awh, 2010; Mandler & Shebo, 1982; Pagano, Lombardi, & Mazza, 2013; Trick & Pylyshyn, 1994). Larger sets, in contrast, exceed this fixed number of items and require individuals to serially shift attention across the array and update working memory (Ester et al., 2012). The time required for these serial shifts of attention and working memory updates accounts for slower reaction times and steeper slopes for quantities above the subitizing range.

Importantly, these studies promote a promising way to evaluate the relationship between subitizing capacity and groupitizing fluency, because they offer an alternative method of deriving subitizing limits without relying on behavioral data. However, because of the high cognitive demands of the task typically used to evoke the N2pc, it is unclear whether it is feasible to conduct this type of experiment in developmental populations. In addition to using behavioral measures to study groupitizing, this study will evaluate the feasibility of measuring meaningful electrophysiological measures in children, using an N2pc enumeration task.

Developments in Subitizing and the Emergence of Groupitizing

In addition to behavioral and neural studies of enumeration in adults, studies of children's enumeration skills have identified developmental changes in subitizing ability, which presumably affect the emergence and improvement of groupitizing skills. Chi and Klahr (1975) studied reaction times when children and adults enumerated 1-8 dots, demonstrating that 5- and 6- year old children have slower reaction times for all set sizes compared to adults. Importantly, the range of numerosities that children can subitize also differs from that of adults. Children's subitizing limit is approximately three items, while adults can subitize four and, for some adults, five items. In a cross-sectional study, Freeman (1912) compared enumeration measures for young children (in early elementary school), older children (in late elementary school), and adults. He found that children's subitizing range was smaller than that of adults, and that this difference lay primarily between the younger and older children, while older children performed similarly to adults.

This indicates that some developmental change occurs between the start and end of elementary school that allows children to subitize higher quantities, although their enumeration latencies and fluency may not change substantially. According to the fixed-capacity model of subitizing, an increase in attention and working memory capacity could account for the increase in subitizing between early and late elementary school.

Younger children may only be able to grasp three objects simultaneously, while older children can grasp four due to increases in attention and working memory capacity. This finding has important implications for groupitizing. Individuals with different subitizing ranges may perform differently on groupitizing tasks, regardless of their fluency in accessing number composition, and this could potentially be traced back to individual differences in attention and working memory capacity.

Interestingly, Freeman (1912) found that while adults and older children were able to groupitize arrays that were structured into subgroups, younger children did not demonstrate any groupitizing ability. Their enumeration speed for grouped arrays did not differ from their speed for unstructured arrays. Starkey and McCandliss (2014) conducted a cross-sectional study of groupitizing skills with children who were enrolled in kindergarten through third grade. The results reported in this study support Freeman's findings: while first, second, and third graders demonstrated significantly faster reaction times for grouped than unstructured arrays, kindergarten children did not (Figure 3).

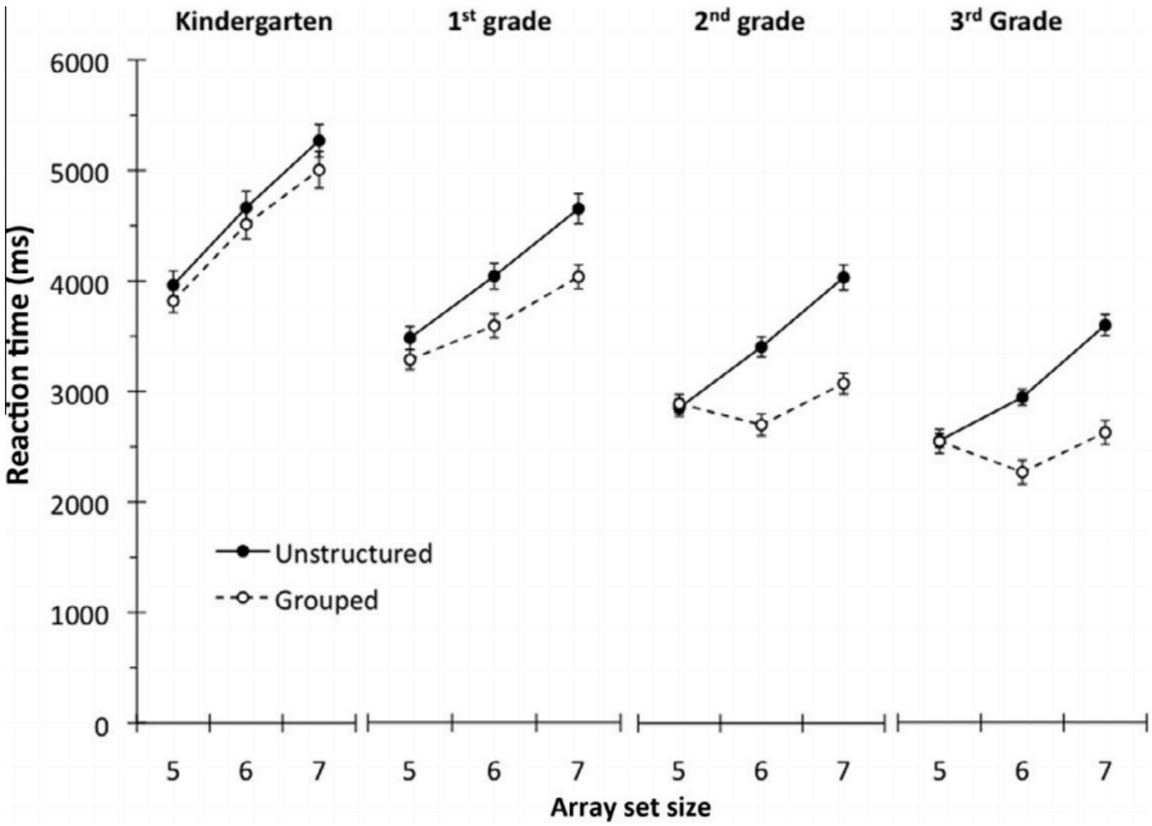

Figure 3. Enumeration latencies for unstructured and grouped arrays for kindergarten, first, second, and third grade children, from Starkey and McCandliss, 2014

A longitudinal study is needed in order to examine the development of groupitizing skills more conclusively. However, the findings discussed above provide initial evidence that some cognitive or perceptual change during early elementary school enables children to shift from serially counting every non-subitizable array to taking advantage of grouped structure for faster enumeration. This emergence of a reaction time benefit for grouped arrays could be accounted for if children lack the perceptual abilities necessary in order to notice the pop-out effect of subgroups in an array. However, multiple studies have demonstrated that children possess these perceptual abilities (Gerhardstein & Rovee-Collier, 2002), and are able to group information based on perceptual properties (Feigenson, Carey, & Hauser, 2002; Feigenson & Halberda, 2008). This suggests that conceptual, rather than perceptual, developments are responsible for the emergence of groupitizing.

Importantly, Starkey and McCandliss (2014) found that children in second and third grade did not demonstrate a positive slope in reaction times across set size for grouped arrays (Figure 3). Slopes for arrays in the grouped condition were not significantly different from zero, while positive linear slopes were found for arrays in the unstructured (counting) condition. This result is critical to understanding the mechanism behind groupitizing, because it demonstrates that no serial, set size-sensitive processes are recruited by children who have well-developed groupitizing skills. The involvement of a serial process, like counting the items in a subgroup, would produce a positive slope, with reaction times increasing with set size (Trick & Pylyshyn, 1994). Similarly, the involvement of adding (such as adding the quantities of the subgroups) would produce a positive slope, as reaction times for addition have been shown to increase with the magnitude of the addends (Butterworth et al., 2001).

This finding from Starkey and McCandliss (2014) leads to the idea that grouped arrays enable *direct* access to the overall quantity of the array from the quantities of the subgroups, meaning no impact of set size. As children develop and gain exposure to the properties of number, their concept of number develops to include the combinations of sets that make up a number, or the sets into which a number can be decomposed. This enables children to connect the subgroup quantities and the total quantity of the array using a process (such as incremental set combination; Geary et al., 2009; Geary et al., 2007) that is not influenced by changes in array quantity while the number of subgroups is held constant. This would produce a flat slope with no set size effect, as was observed for second and third graders.

Starkey and McCandliss (2014) also found a strong relationship between groupitizing fluency and symbolic arithmetic skills. This link provides further support for the idea that when children do acquire the ability to groupitize, this shift reflects important developments in their concept of number and insights into number composition. Therefore, it is important to characterize how groupitizing skills in a longitudinal sample of elementary school children, to clarify when groupitizing slopes become flat and to identify other variables that predict this development. Therefore, Experiment 1 of this study assessed the development of groupitizing skills in a longitudinal sample, and examined the association between development in groupitizing fluency and other enumeration skills.

Experiment 2 applied the proposed model of groupitizing to an enumeration paradigm in which the number and magnitude of subgroups are manipulated. This provided the opportunity to compare the impact of these manipulations on enumeration performance for children with different subitizing ranges. In 2000, Wender and Rothkegel conducted a study that provided the basis for this experiment. Inspired by Mandler and Shebo (1982), who investigated whether it is possible to access quantities beyond the subitizing range without engaging in a serial counting process, Wender and Rothkegel (2000) evaluated adults' enumeration speed for grouped arrays. Importantly, they varied the number of subgroups and size of subgroups in their grouped arrays. Wender and Rothkegel found that adults' enumeration latencies for grouped arrays were not influenced by the total array size, as long as the subsets were subitizable and the number of subsets remained fixed. However, enumeration latencies increased as the number of subgroups increased from two to three. Taken together, these findings provide additional evidence that the subgroups within an array are subitized in the process of groupitizing, because serially counting each subgroup would yield longer reaction times with increased array size, regardless of the grouping structure. In addition, the observation that reaction times increase as the number of subgroups (but not the total set size of the array) increases suggests that the subgroups are subitized one at a time, rather than in parallel.

Experiment 2 further explored this effect in a sample of children, some of whom could subitize four and some of whom could not. Children enumerated grouped arrays that had been designed to manipulate two parameters – magnitude of the subgroups, and number of subgroups – in both non-symbolic and symbolic formats. This experiment evaluated the effect of these manipulations on enumeration latency and fluency to test whether subitizing is involved in groupitizing, and to further explore the dynamics of accessing overall array size from subgroup quantities. The symbolic component of Experiment 2 assessed individuals' performance when the subitizing component was removed from the task, isolating the effect of these manipulations on children's ability to apply insights into number composition.

Goals of the Study

The three experiments included in this study investigated the development of groupitizing, and mechanisms underlying this ability. These experiments were motivated by the hypothesis that groupitizing recruits (and therefore reflects) important developments in children's understanding of number, and that it is a unique predictor of math development. Experiments evaluated the theory that the process of groupitizing consists of subitizing subgroups, and applying insights into the composition of numbers in order to access the total numerosity. Specific goals of each experiment were as follows:

1. Experiment 1 modeled developmental changes in groupitizing performance, and modeled developmental changes in associations between groupitizing and improvements in higher-level math skills.
2. Experiment 2 examined how two manipulated parameters – number of subgroups, and maximum size of the subgroups – impacted enumeration performance in the non-symbolic and symbolic domains, comparing these effects for children who could and could not subitize four.
3. Experiment 3 attempted to derive electrophysiological measures of attention (N2pc) and behavioral measures of visual working memory in children to investigate the neural and cognitive underpinnings of developments in subitizing range.

The primary goal of this study was to elucidate the cognitive basis of groupitizing ability. A secondary goal was to evaluate the feasibility of conducting an ERP study of the N2pc component in children despite the high cognitive demand of N2pc tasks. If successful, the N2pc could provide a preliminary window into the neural substrates of groupitizing.

CHAPTER II

EXPERIMENT 1: ASSESSING DEVELOPMENTAL CHANGE IN GROUPITIZING ABILITY IN A LONGITUDINAL SAMPLE OF CHILDREN

Starkey and McCandliss (2014) conducted a cross-sectional study of the enumeration skills of children in kindergarten through third grade. Comparisons of mean enumeration latencies for each grade suggested that groupitizing skills undergo significant development over the course of elementary school. Furthermore, groupitizing slopes were a significant predictor of symbolic math scores. Therefore, this experiment sought to explore these effects in a longitudinal sample, enabling conclusions about the relationships between developing enumeration skills and improvements in symbolic math.

Distinguishing fluency from speed

In Experiment 1, slope was used as a measure of enumeration fluency. Steep positive slopes indicate a strong effect of set size, such that it takes much longer to enumerate 7 dots than 5 dots. Flat slopes approximating zero, on the other hand, suggest that there is no effect of set size; it takes no additional time to enumerate 7 dots compared to 5 dots. There is an inverse relationship between slope and fluency: steep slopes equate to low fluency, and flat slopes close to zero indicate greater fluency. Because slope controls for factors that are consistent across set sizes, this can be attributed to the strategy children use to determine the overall array size once they know how many dots are in each subgroup. A steep slope indicates a process that is influenced by set size; this could mean that children are serially counting the array, adding subgroup quantities together, or retrieving arithmetic facts from memory (Butterworth et al., 2001; Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). A flat slope that is not significantly greater than zero indicates that children are gaining direct access to overall quantity, which is not influenced by set size; children with flat groupitizing slopes are able to connect subgroup quantities to total numerosity without engaging addition or math fact retrieval. This is significant, because it demonstrates that direct access to numerosity is possible for quantities beyond the subitizing range when arrays are structured into subitizable subgroups (Starkey & McCandliss, 2014; Wender & Rothkegel, 2000).

Starkey and McCandliss showed that older children had flatter slopes (greater fluency) for three-subgroup arrays of 5, 6, and 7 dots than younger children did. Moreover, some second and third graders had slopes of zero, suggesting that they had achieved a level of groupitizing fluency that enabled direct apprehension of quantity. Regression analyses showed that groupitizing slopes predicted significant and unique variance in scores on a test of symbolic arithmetic. This result suggests an important role for groupitizing as a skill that reflects number concepts that are foundational for math development. However, this study used a cross-sectional design, introducing the possibility that grade-related effects are due to group differences.

Experiment 1, therefore, followed a subset of the cross-sectional sample from Starkey and McCandliss over the course of three years. The goal of Experiment 1A was to describe the development of groupitizing fluency as children progress through elementary school, and to identify other enumeration skills that may predict this growth.

Multilevel modeling was used to describe the development of individuals' groupitizing fluency over elementary school.

Experiment 1B built upon regression results from Starkey and McCandliss. The goal was to establish whether change in groupitizing fluency predicts individual differences in symbolic arithmetic skills and rate of improvement over elementary school, and to establish whether this contribution was distinct from that of other enumeration skills.

Experiment 1 Method

Participants

The sample of participants for Experiment 1 was drawn from a large sample of children who participated in Starkey and McCandliss's (2014) cross-sectional study in the fall of 2010. Fall of 2010 served as time point (TP) 1. Four cohorts of children (kindergartners, first graders, second graders, and third graders) participated in TP1. Some of these children returned for follow-up sessions: TP2 occurred in spring 2011, TP3 occurred in spring of 2012, and TP4 occurred in spring of 2013. The sample of participants included in Experiment 1 analyses consists of children who have data from two to four time points, spanning three years of school. Most children were tracked from kindergarten through second grade, or first through third grade (Table 1).

Table 1. Number of participants from each cohort who participated in follow-up testing sessions.

	Time point			
Grade at TP1	TP1	TP2	TP3	TP4
Kindergarten	23	41	30	26
1	17	33	22	18
2	0	17	8	5
3	1	19	10	5

All children who participated in more than one time point and produced high-quality data at these time points were included in analyses for Experiment 1, yielding a sample of 113 participants. Note that most children participated in two or three of the four time points, resulting in missing data with uneven spacing between time points. To address this, hierarchical linear modeling was used for analyses of longitudinal growth, because this methodology and software (unlike most other approaches to growth modeling) do not require participants to have data for all time points, and do not require equal temporal spacing of time points.

Measures

Experiment 1 used data from a computer-based enumeration task, named Feeding Nemo due to the “fish food store” game setting. This task measured children’s exact enumeration skills for sets of 1 through 8 dots. Stimuli consisted of black dots on a gray screen. For all set sizes, two levels of dot size and two levels of dot density were used in order to control for the potential cueing effects of these perceptual characteristics (Gebuis & Reynvoet, 2011). Thus, each set size was presented in four combinations of dot size and density: small dots with low density, small dots with high density, large dots with low density, and large dots with high density.

Unstructured stimuli were created using a MATLAB 7 (MathWorks, Natick, MA, USA) program that randomly generated locations for dots. All unstructured arrays with a set size of greater than three dots were not arranged in canonical patterns, in order to avoid the influence of canonicity on reaction time (Logan & Zbrodoff, 2003Mandler & Shebo, 1982). For arrays with a set size of greater than one, the distance between each dot and the closest neighboring dot was calculated. To ensure that each unstructured stimulus had very weak perceptual grouping cues, inter-dot distances did not vary by more than 3 millimeters across all unstructured arrays. To control for the impact of number of subgroups on enumeration speed (Wender & Rothkegel, 2000), each grouped stimulus for Feeding Nemo contained three subgroups. Additionally, all grouped stimuli within a set size were split into subgroups of the same numerosities (e.g., Grouped 5 always appeared as 2, 2, and 1; Grouped 7 always appeared as 3, 3, and 1). Table 2 shows sample stimuli for target set sizes in grouped and unstructured conditions.

Table 2. Sample stimuli for set sizes 5-7 in both grouping conditions from the original Feeding Nemo task (Starkey & McCandliss, 2014).

	Stimulus Structure	
Set Size	Unstructured	Grouped
5		
6		
7		

Grouped stimuli were created in two steps using a MATLAB 7 program. Individual subgroups were generated first, and then locations for these subgroups were generated separately. Subgroup images were superimposed on locations by matching the center pixel of each subgroup with the exact coordinates of the location using digital image manipulation software. A large number of stimuli were generated, and the distance between dots in each subgroup, and between subgroup locations, were calculated. Grouped stimuli with a ratio of inter-dot distance to inter-location distance that was between than 1:4 and 1:5 had strong perceptual grouping cues, and were selected for use in the task. This mathematical approach to creating stimuli helped to ensure that arrays were distinctly unstructured or grouped.

The Feeding Nemo task was administered using Paradigm (Perception Research Systems) on a Dell Latitude E6500 laptop computer with a 15.4-inch display. To motivate participants to enumerate exactly, they were told to pretend that they were helping the experimenter make labels for fish food containers, and that the number therefore had to be exactly correct. When a tray of fish food (dots) appeared on the computer screen, their job was to say exactly how many pieces of fish food they saw. They were urged to enumerate as quickly and accurately as possible, and to remain quiet until saying their final answer into a handheld Audio-Technica ATR20 microphone that recorded their vocalizations. Words like “count” and “group” were explicitly avoided in the instructions, so as not to prompt any particular enumeration strategy. This ensured that if participants were groupitizing the grouped arrays, this was a spontaneous rather than cued strategy.

Participants were given fourteen practice trials before starting the Feeding Nemo task. Stimuli were presented in random order, with unstructured and grouped arrays intermixed. Before each stimulus onset, participants saw a fixation asterisk in the center of the screen. The experimenter pressed the spacebar to cue the onset of the trial, and the stimulus appeared after 500 milliseconds. The stimulus remained on the screen for as long as it took the child to enumerate. After each vocalization, the experimenter entered the child’s answer, and pressed the spacebar to cue the next trial.

Each set size was presented eight times for a total of 64 trials: set sizes 1-3 each occurred eight times in the unstructured condition, and set sizes 4-8 each occurred four times in the unstructured condition and four times in the grouped condition. Trials were split into three blocks with 21-22 stimuli per block to allow participants to take short breaks. No specific trial-by-trial feedback was given regarding accuracy or speed, but after the practice trials and between each block, the experimenter reminded children to enumerate as quickly and accurately as possible.

Microphone recordings of participants' responses began at stimulus onset and ended after each vocalization, when the experimenter cued the next trial. This resulted in one sound file for each completed trial. Digital recordings of each trial were evaluated by one of four independent raters, all trained experimenters, who were blind to grouping condition. Raters marked enumeration latency for each trial using a semi-automated voice onset latency estimation software package (Utterance 1.0, Perception Research Systems, Inc.). Enumeration latency started at the onset of the visual stimulus, and ended at the onset of the participant's vocalization. Inter-rater reliability was calculated using twenty subjects' sound files (1,280 files total) that were coded by all raters and compared to a single 'master rater.' Inter-rater reliability was very high; for all raters, Kappa > 0.80, $p < 0.001$, with the coding of the master rater.

Individual trials were excluded for one or more of the following reasons: inaccurate response, accurate initial response that was changed within the recording epoch to an inaccurate response, or a response that was unclear. For each stimulus type (e.g., Unstructured 3, Grouped 5, Unstructured 6), one median reaction time was calculated from the remaining trials. Medians were used instead of means because they are more robust to outliers (e.g., the occasional trial in which a participant was not paying attention, resulting in an atypically long enumeration latency). Participants with fewer than two valid trials per stimulus type were reported as having missing data for that stimulus type.

Set sizes 1, 2, and 3 provided measures of subitizing speed. Set sizes 4 and 8 were excluded by design from analyses, in order to avoid boundary effects induced by the smallest and largest set sizes for arrays that appear in both grouping conditions. Unstructured arrays of 5-7 dots provided measures of counting speed, and grouped arrays of 5-7 dots provided a measure of groupitizing speed. Table 3 displays participants' median enumeration latencies for unstructured and grouped arrays of target set sizes, averaged separately for all kindergarteners, first, second, and third graders.

Table 3. Enumeration latencies in milliseconds *(standard deviations)* for target counting and groupitizing arrays, averaged across each grade level.

Grade	Unstructured dots			Grouped dots		
	5	6	7	5	6	7
K	3963.72 *(1171.83)*	4664.06 *(1342.77)*	5271.10 *(1341.09)*	3819.30 *(979.09)*	4513.57 *(1222.16)*	5004.89 *(1502.34)*
1	3479.88 *(943.53)*	4039.56 *(1080.84)*	4649.95 *(1226.45)*	3283.81 *(812.79)*	3590.57 *(982.52)*	4033.60 *(1002.38)*
2	2835.32 *(657.40)*	3383.29 *(769.45)*	4007.36 *(980.51)*	2873.22 *(744.10)*	2681.85 *(841.47)*	3053.99 *(790.12)*
3	2461.72 *(576.35)*	2839.67 *(671.24)*	3470.15 *(872.27)*	2454.82 *(689.02)*	2185.69 *(753.53)*	2532.38 *(640.42)*

In addition to enumeration latencies, reaction time slopes were calculated as indices of fluency. Slopes for unstructured 5-7 and grouped 5-7 were calculated as measures of counting and groupitizing fluency, respectively. Slope calculations were only performed on participant datasets that contained no missing values for any of these six critical conditions.

The Woodcock-Johnson III Math Fluency (MF) subtest was also administered; this served as the measure of symbolic arithmetic fluency for Experiment 1. The Woodcock-Johnson III is a widely used standardized test of academic achievement, and is normed for use with subjects 5 years of age and older. Test-retest reliability for the MF subtest has been calculated to be 0.95 (Schrank, McGrew, & Woodcock, 2001). The MF subtest is a timed pencil-and-paper test of basic arithmetic facts. Raw scores reflect how many problems are answered correctly in three minutes. Problems were presented using symbolic digits in a vertical format. Addition and subtraction problems were intermixed, and multiplication was introduced after item 60. This assessment provided a measure of school-relevant symbolic arithmetic skills, which have been shown to linked to the Feeding Nemo groupitizing measures (Starkey & McCandliss, 2014).

For longitudinal data analysis in Experiment 1, slopes were used instead of enumeration latencies, because hypotheses were concerned with change in fluency, rather than speed. Therefore, the variables used in the following analyses were subitizing slope, counting slope, groupitizing slope, and MF raw scores.

Grades were coded as 0, 0.5, 1, 1.5, etc., to accurately reflect spacing between testing sessions, because the first time point occurred in the fall, and the remaining time points were the spring. For example, for a child who participated in all four testing sessions and was enrolled in kindergarten at the start of the study, grades would be coded as 0, 0.5, 1.5, and 2.5.

Approach to building multilevel models

To model growth in groupitizing fluency and identify other skills that predict variance in groupitizing development, a hierarchical linear modeling (HLM) approach was taken using HLM software (Raudenbush, Bryk, & Congdon, 2011). HLM is often used for data that has a nested structure (such as students nested within a classroom), which can result in non-independence at the first level of measurement (student). In longitudinal data for developing skills, observations at each time point are naturally nested within individuals. In this case, groupitizing fluency at TP2 depends to some extent on groupitizing fluency at TP1. Multilevel modeling approaches, such as HLM, should be used to model this type of data due to this violation of the assumption of independence.

HLM offers the advantage of modeling individual change over time as well as between- and within-individual variation in starting point and rate of change. Furthermore, HLM allows for predictor variables to be time-varying, such that *developments* in enumeration fluency could be used as predictors, rather than simply using a measure of fluency from one time point (Raudenbush & Bryk, 2002). In addition to these advantages, HLM software is flexible with missing data, and can handle these inconsistencies without using interpolation to estimate values for missing data points, which can be inaccurate. Consequently, HLM is often used to model students' growth in math, reading, and other cognitive abilities as they progress through school (for examples, see Ding & Davidson, 2005, and Duckworth, Tsukamaya, & May, 2010).

For Experiments 1A and 1B, two-level models were built in HLM with the use of restricted maximum likelihood estimation (Raudenbush, Bryk, Cheong, Congdon, & du

Toit, 2011). Level 1 (grade) measured within-child variability across time. Level 2 (individual) measured between-child variability in change over time.

HLM is sometimes used to specify models with many predictors, and to compare these models using various tests of goodness of fit. Building large comprehensive models to characterize an outcome, like groupitizing fluency, should be addressed in future studies, but was beyond the scope of this experiment. Instead, Experiment 1 analyses focused on a specific theory- and data-driven hypotheses that were directed toward identifying enumeration variables that predict change in groupitizing fluency and MF beyond what is accounted for by individuals' starting point and the process of going through elementary school. Therefore, analyses and results will focus on p-values for added variables, which indicate whether these variables predicted a significant amount of unique variance at Levels 1 and 2.

Experiment 1A: Describing and predicting change in groupitizing fluency over grade

Hypothesis

Based on findings from cross-sectional data (Starkey & McCandliss, 2014), I hypothesized that children become more fluent at groupitizing as they progress through elementary school due to developments in their concept of numbers. Therefore, groupitizing fluency should decrease linearly across grades. I did not anticipate quadratic or exponential curvature in the growth of groupitizing slopes. Curvature in the trajectory of a decreasing variable (for example, in an exponential decay function) would suggest that rate of change accelerates or decelerates over time. The proposed mechanism of groupitizing involves two steps: subitizing subgroups, and using knowledge of the combinations of sets that comprise a number to access the quantity of an array. Neither of these processes have been shown to develop exponentially over time, potentially causing rate of groupitizing fluency growth to accelerate or decelerate. Therefore, there was no theoretical reason that a non-linear term would improve a model predicting change in groupitizing fluency.

Next, subitizing fluency (calculated across enumeration latencies for unstructured arrays of 1-3 dots) and counting fluency (calculated across enumeration latencies for unstructured arrays of 5-7 dots) were added to the growth model as potential predictors of change in groupitizing fluency. Developments in counting may predict developments in groupitizing, particularly for children with low groupitizing fluency who were not able to directly access the numerosity of an array from subgroup quantities, and instead were using a serial cognitive process. Therefore, changes in counting fluency might be associated with changes in groupitizing fluency more for some (low-fluency) individuals than for others (high-fluency). If this is the case, counting fluency should be a significant predictor when added to the model *with random effects* (indicating that the influence of the predictor variable varies across individuals).

Finally, even though subitizing is involved in the process of groupitizing, subitizing fluency does not change substantially over elementary school. Subitizing is characterized by relatively flat slopes, even for children (Starkey & McCandliss). Therefore, I hypothesize that changes in subitizing fluency will not significantly predict developments in groupitizing fluency.

Results

Before testing growth models, a plot of groupitizing slopes across grades was created (Figure 4). Singer and Willett (2003) recommend that longitudinal data be visually inspected before building models for an initial assessment of the shape of trajectories. Groupitizing slopes did appear to decrease linearly across grades, providing preliminary support for the predicted effect. There did appear to be some curvature among the oldest children, as slopes approach zero, so an exponential decay function of grade (in which rate of change would slow down as grade increases) was assessed to check for this.

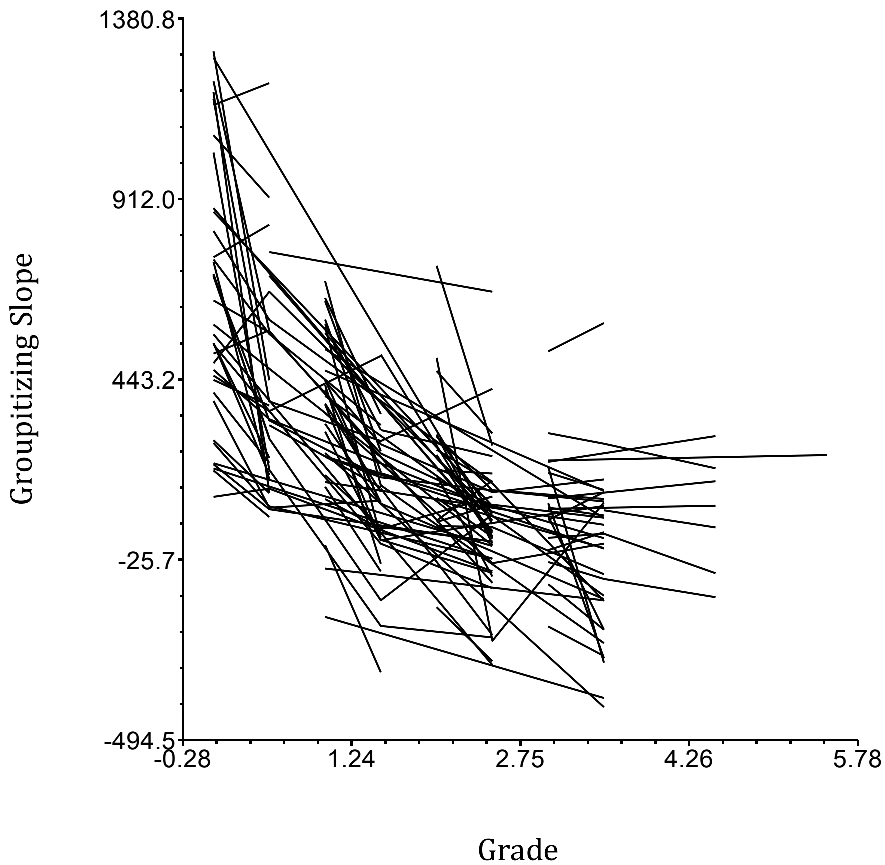

Figure 4. Change in groupitizing slopes over grades

The first step of the HLM approach is to build an unconditional model in which variables representing time are the only predictors of the outcome variable. Once an unconditional model is specified, other predictors of interest may be added in to test whether they predict additional variance in growth (O'Connell & McCoach, 2008). For Experiment 1A, groupitizing slope was set as the outcome variable. Models including grade and an exponential decay function of grade were both assessed, with the exponential decay function added as a third predictor after intercept and grade. Grade was a significant predictor of change in groupitizing fluency, but when the exponential decay function was added to this model to test for the presence of non-linear curvature, the model improvement was negligible.

Therefore, the unconditional model for groupitizing slope included intercept with random effects and grade as an uncentered predictor with random effects (uncentered because the intercept was intended to reflect starting point at grade 0, or kindergarten). In two-level longitudinal models, intercepts indicate initial status for individuals at time 0 (in this case, groupitizing fluency at kindergarten). In this model, the intercept term with random effects indicates that there was significant variance in individual's initial groupitizing slope at grade 0 (kindergarten). The grade term with random effects indicates that the linear rate of growth also varied across individuals. Figure 4 supports this model because individual's average groupitizing fluency, and the slopes of individuals' changes in groupitizing fluency across grade, both appear to vary. A plot of residuals for this model showed that they were evenly distributed around zero, indicating that the model was a good fit for the data.

Next, other enumeration fluency measures were added to the unconditional model to examine whether they explained additional Level 1 and Level 2 variance even after accounting for variance due to grade and intercept. Because subitizing and counting fluency are time-varying covariates, meaning that they may change with grade, steps were taken to avoid problems of multicollinearity that can occur when two variables vary together. Therefore, subitizing slope and counting slope were added to the unconditional model separately as individual-mean-centered variables. Mean-centering substantially reduces any bias that could be introduced by the relationship between these predictors and the outcome measure (Raudenbush & Bryk, 2002).

As expected, subitizing slope did not predict significant variance at Level 1 or Level 2, and was therefore excluded from the model. However, counting slope with random effects predicted significant variance at Level 1 (change across time within each individual) and Level 2 (between-individuals differences in rate of change). Table 4 displays Level 1 coefficients and results of significance tests for this model. The positive coefficient for counting slope at Level 1 indicates that decreases in counting slope predict decreases in groupitizing slope (and, thus, that improvements in counting slope predict improvements in groupitizing slope). This result supports my hypothesis that changes in counting fluency are significantly related to changes in groupitizing fluency. Specifically, this result demonstrates that counting fluency predicts changes in groupitizing fluency across grade within each student, and affect individual differences in rate of groupitizing development.

Table 4. Results of conditional model predicting groupitizing slope from Intercept, Grade, and Counting Slope.

Fixed Effects	Coefficient (Standard Error)	t
Intercept	513.46 (29.52)	17.39 **
Grade	-152.31 (12.41)	-12.274 **
Counting Slope	0.29 (0.08)	3.592 **
Random Effects	χ^2	
Intercept	305.64 **	
Grade	135.57 *	
Counting Slope	152.90 *	

Note. * p<0.05 ** p<0.001

Discussion

Experiment 1A verified that groupitizing slope decreases linearly, at rates that vary across individuals. This confirms that groupitizing fluency improves as children progress through elementary school. The best-fitting unconditional model specified random-effects intercept and grade terms, suggesting that there was significant between-child variance in initial groupitizing fluency in kindergarten and significant child-level individual differences in the rate of improvement in groupitizing fluency.

Analyses also demonstrated that changes in subitizing fluency were not significantly associated with developments in groupitizing slope, likely because there is so little change in subitizing fluency in elementary school. Subitizing slopes are already relatively flat when children enter elementary school, and may not provide enough variance to predict growth in other enumeration skills.

However, as predicted, developments in counting fluency did predict significant variance in individuals' change in groupitizing fluency across time (at Level 1), as well as individual differences in rate of change (at Level 2). The relationship between developments in counting fluency and developments in groupitizing fluency may be especially strong for younger children with low groupitizing fluency children. Low-groupitizing fluency children (i.e., those with steep positive groupitizing slopes) cannot directly access the numerosity of an array from the subgroup quantities, and therefore rely on less efficient processes that may include counting. The significance of random effects for Counting Slope indicated that the extent to which rate of change in groupitizing fluency is impacted by developments in counting fluency varies across individuals, which supports this claim. Future analyses should clarify this by examining counting fluency as a nonrandomly-varying predictor (Raudenbush & Bryk, 2002) that potentially accounts for Level 2 variance specifically for early grades, before children have achieved fluent groupitizing.

Experiment 1B: Predicting change in symbolic arithmetic skills from change in subitizing, counting, and groupitizing fluency

Hypotheses

Experiment 1B built on a finding from Starkey and McCandliss that groupitizing fluency predicts significant and unique variance in symbolic arithmetic skills, measured by raw scores on the MF subtest. Based on the hypothesis that groupitizing reflects developments in conceptual number knowledge, and may involve set combination, which is linked to children's arithmetic ability (Fuchs et al., 2006; Fuchs et al., 2010; Geary et al., 2009; Geary et al., 2007), I expected that this finding would be confirmed and extended in the longitudinal sample. Specifically, I predicted that a two-level model for predicting change in MF raw scores would demonstrate that developments in groupitizing fluency account for significant variance at Levels 1 and 2. This would demonstrate that developments in groupitizing fluency predict individuals' MF growth across grade, as well as individual differences in rate of growth, even after accounting for variance due to intercept, grade, developments in counting fluency, and developments in subitizing fluency.

Results

Math fluency raw scores across grades were visually inspected before building an unconditional model (Singer & Willett, 2003). This plot suggested that MF raw scores increase linearly with grade (Figure 5). MF scores are bounded by the total number of problems on the subtest, and the plot suggests that there may be a slight ceiling effect that would need to be addressed in the model. Therefore, to specify the best unconditional model, both grade and a logarithmic growth function were tested as predictors of change in MF scores.

The best-fitting unconditional model for describing growth in MF raw scores included intercept and grade with random effects. This model specifies that starting points for MF raw scores in kindergarten varied between participants, and that individuals' rate of growth also varied. Residuals were evenly distributed around zero, suggesting that this model was a good fit for the data. When the logarithmic growth function was added to the model, it was not a significant predictor of additional variance in growth; this may be due to the small number of fourth and fifth grade participants who had achieved MF scores approaching ceiling.

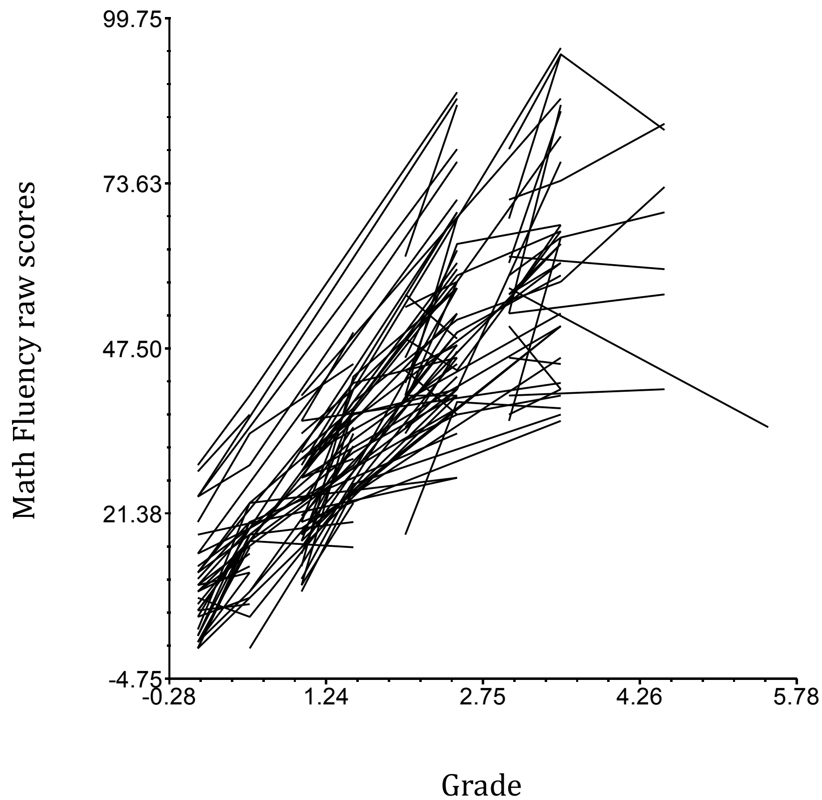

Figure 5. Change in MF raw scores over grades

Next, subitizing, counting, and groupitizing slopes were added to the model separately to test whether they predicted significant variance at Levels 1 and 2 after accounting for variance due to grade and intercept. As in Experiment 1A, these three time-varying predictors were individual-mean-centered in order to reduce potential bias that could be introduced by multicollinearity (Raudenbush & Bryk, 2002). Neither change in subitizing fluency nor change in counting fluency accounted for significant variance in developments in MF at either level. Groupitizing slopes with fixed effects, however, did predict significant variance at Level 1 (within-subjects change over time). Results showed a negative coefficient for groupitizing slope, indicating that decreases in groupitizing slope (i.e., increases in groupitizing fluency) predicted improvements in MF. Even after subitizing slope and counting slope were added back into the model, developments in groupitizing fluency predicted significant variance at Level 1. Table 5 displays Level 1 coefficients and results of significance tests for this model.

Table 5. Results of conditional model predicting MF raw scores from Intercept, Grade, Subitizing Slope, Counting Slope, and Groupitizing Slope.

Fixed Effects	Coefficient (Standard Error)	t
Intercept	13.64 (1.28)	17.39 **
Grade	13.18 (0.78)	-12.274 **
Subitizing slope	0.00 (0.00)	1.28
Counting slope	0.00 (0.00)	-0.68
Groupitizing slope	-0.13 (0.02)	-4.602 **
Random Effects	χ^2	
Intercept	178.12 **	
Grade	233.88 **	

Note. * p<0.05   ** p<0.001

Discussion

Experiment 1B demonstrated that changes in groupitizing fluency predicted variance in individuals' growth in symbolic arithmetic, even after accounting for changes due to intercept, grade, and changes in subitizing and counting fluency. The longitudinal aspect of the design of this experiment was critical, because it enabled predictions of individuals' change over time *and* individual differences in how rapidly this development occurs. Thus, Experiment 1 took regression findings from Starkey and McCandliss (2014) one step further, suggesting that improvements in groupitizing fluency, which reflect developments in conceptual number knowledge, contribute to improvements in symbolic math skills. It has been proposed that these developments in conceptual number knowledge relate to children's new insights into number composition, and that this enables direct access to the numerosity of an array once the quantities of subgroups have been determined. Experiment 2 will further examine these cognitive dynamics.

Integrating findings from Experiment 1

Experiment 1 used two-level hierarchical linear models to investigate relationships between developing enumeration skills and improvements in symbolic arithmetic. This experiment also examined whether developing enumeration skills predict within-subject change across time, and between-subject variance in rate of change. Results do not establish causality between predictors and outcomes, but do provide strong evidence for links between developments in counting and groupitizing skills, and developments in groupitizing and improving symbolic arithmetic.

Changes in counting fluency, but not subitizing fluency, predicted developments in groupitizing fluency. Although subitizing is part of the process of groupitizing, Starkey and McCandliss (2014) provided preliminary evidence that the developments that allow children to groupitize more fluently are not related to changes in subitizing fluency. Rather, developments in conceptual number knowledge - specifically, insights into number composition, and the ability to directly access the combinations of sets that make up numbers – appear to contribute to improvements in groupitizing fluency. This will be addressed further in Experiment 2.

Developments in groupitizing fluency overwhelmingly predicted changes in symbolic math fluency. Critically, changes in subitizing and counting fluency did not. Regression results from Starkey and McCandliss (2014) indicated that groupitizing slopes were strongly related to math fluency, but this analysis was limited by a cross-sectional sample. Experiment 1 extends this finding to longitudinal data, suggesting that developments in groupitizing fluency, which reflect insights into the composition of number, play a significant role in improvements in symbolic arithmetic and other higher-level math skills.

CHAPTER III

EXPERIMENT 2: INVESTIGATING THE COGNITIVE DYNAMICS OF GROUPITIZING ABILITY

The purpose of Experiment 2 was to apply Starkey and McCandliss's (2014) proposed mechanism of groupitizing to a new task with variations in grouping structure. A new version of the Feeding Nemo task from Experiment 1 was created and named "FN Grouped." During this task, children were presented with only grouped arrays to enumerate. These arrays varied on two parameters: the maximum quantity of dots within a subgroup (three or four), and total number of subgroups (two or three). This design enabled the following comparisons, which shed light on the mechanism underlying groupitizing.

In Experiment 2A, enumeration latencies and slopes were compared for three-group arrays that did and did not contain a subgroup of four dots. Importantly, the sample of participants was separated into two "subitizing span groups" (4-Counters and 4-Subitizers) based on enumeration data for unstructured arrays in the original Feeding Nemo task, which participants completed during the same session as the FN Grouped task. Experiment 2A addressed the influence of subitizing span on groupitizing speed and fluency under conditions in which participants can versus cannot subitize all of the subgroups. Experiment 2B compared enumeration latencies and slopes for arrays that were spatially grouped into two or three subgroups, but all contained at least one subgroup of four dots. By varying the number of subgroups that must be subitized, Experiment 2B addressed the question of whether subgroups are subitized individually or in parallel, and explored the interaction between subitizing range and groupitizing ability.

Children also participated in a symbolic version of FN Grouped, named FN Symbolic. In this task, digits were used in place of subgroups of dots. By eliminating the subitizing aspect of the task, this symbolic task isolated children's fluency in recruiting knowledge of number composition to figure out the total magnitude of an array. Experiment 2C examined the effects of manipulating the number of digits and digit magnitude on this component of the groupitizing process.

Experiment 2 Method

Participants

Experiment 2 participants were a subgroup of children who participated in Experiment 1 and returned to the Educational Cognitive Neuroscience Lab at Vanderbilt University in the spring of 2013 (TP4). A total of 43 children participated, and were enrolled in second (n=25) and third grade (n=18) when these sessions took place.

FN Grouped (Experiments 2A and 2B)

During this session, children participated in the FN Grouped task, which consisted of only grouped arrays in which two parameters were manipulated: maximum number of dots within subgroups, and total number of subgroups. The two levels of subgroup quantity were referred to as "SubMax 3" (subgroup maximum of three dots) and "SubMax 4" (subgroup maximum of four dots). The two levels of subgroup number were two and three subgroups. Overall set sizes of arrays ranged from 5 to 9 dots, with 5 and 9 as boundary set sizes, and 6 through 8 as target conditions. Target set sizes were presented in three grouping conditions (Table 6).

Table 6. Structure of subgroups in FN Grouped arrays.

Set size	Three Groups, SubMax 3	Three Groups, SubMax 4	Two Groups, SubMax 4
6	3, 2, 1	4, 1, 1	4, 2
7	3, 3, 1	4, 2, 1	4, 3
8	3, 3, 2	4, 3, 1	4, 4

It was not possible to maintain the same target condition range as the original Feeding Nemo (5 through 7), because 5 dots cannot be arranged in three subgroups with one containing 4 dots. However, the goal of Experiment 2 was to compare enumeration latencies for different grouping conditions, rather than to compare enumeration latencies between grouped and unstructured conditions, so it was not necessary to maintain the same target set size range as the original Feeding Nemo.

For FN Grouped, each target set size was presented four times in each grouping condition, resulting in 36 target trials. Boundary conditions were also presented four times in each of two grouping conditions, resulting in 16 boundary trials, or 52 total trials. This task was separated into two blocks of 26 trials.

Stimuli to fit the new array parameters were created using the same MATLAB 7 program and procedure as for Experiment 1 grouped stimuli. Two levels of dot size and two levels of density were employed again in order to avoid cuing of set size based on these perceptual factors (Gebuis & Reynvoet, 2011). A large number of stimuli were generated, and the mathematical approach from Experiment 1 was used again to select arrays that were distinctly structured into groups.

FN Symbolic (Experiment 2C)

Approximately half of the children who participated in FN Grouped also participated FN Symbolic (n=24). For this task, the MATLAB 7 script that generated subgroup locations for Feeding Nemo and FN Grouped was used. Instead of superimposing subgroups over these locations, symbolic numbers were used. The center of each digit was placed over each subgroup location, such that instead of arrays of dots, arrays of digits were created. Any stimuli that displayed digits in a horizontal line were excluded, because these stimuli could appear either as three-digit numbers, or as horizontal arithmetic problems without operational signs.

Digits for the FN Symbolic stimuli corresponded exactly to subgroup sizes in FN Grouped stimuli. Therefore, total set sizes were 6-8 for target trials and 5 and 9 for boundary trials. Target set sizes were presented in three grouping conditions (three-digit SubMax 3, three-digit SubMax 4, and two-digit SubMax 4), and boundary set sizes were presented in two conditions for a total of 52 trials. Again, these trials were separated into two blocks of 26 trials.

Procedure

Children participated in Feeding Nemo, followed by FN Grouped, followed by FN Symbolic. The procedure for the FN Grouped task differed from Feeding Nemo in that participants were informed of the grouped structure and upper and lower limits of the range of set sizes. They were told that the arrays were all arranged in groups, and would contain anywhere from 5 to 9 dots. For FN Symbolic, participants were told to "say the total number on the screen," to avoid giving instructions that might cue a particular strategy. If children were initially confused about the intentionally vague instructions, responding with the total number of digits instead of the total magnitude, the experimenter demonstrated the correct answer for one trial, rather than giving more specific instructions. This trial was then excluded from data analysis.

For both tasks, accuracy was near perfect, likely due to children having experience with similar enumeration tasks in previous sessions, and having unlimited time to respond (although they were encouraged to enumerate as quickly and accurately as possible). Accuracy was not significantly affected by set size or grouping condition for either task. Across all conditions, participants averaged 97.4% accuracy for FN Grouped and 96.8% accuracy for FN Symbolic.

Sound wave files from FN Grouped and FN Symbolic trials were coded, and median reaction times for each set size by grouping condition were calculated followed the same protocol as for Feeding Nemo (see Experiment 1 Method). Exclusion criteria were also the same. Perhaps because children were older and had participated in similar enumeration tasks previously, no children were completely excluded from analyses of enumeration latencies due to poor data quality or too many inaccurate responses.

As in Experiment 1, slopes for each condition were calculated across only three data points (set sizes 6, 7, and 8); therefore, a slope was excluded from analysis if one of these data points was missing. For FN Grouped, I excluded 5 participants' slopes for three-group SubMax 3 arrays, 3 participants' slopes for three-group SubMax 4 arrays, and 3 participants' slopes for two-group SubMax 4 arrays. For FN Symbolic, I excluded 3 participants' slopes for three-group SubMax 3 arrays, and no slopes for three-group SubMax 4 or two-group SubMax 4 arrays. Most of these exclusions were due to poor-quality voice recordings, rather than inaccurate responses.

Classification of participants into subitizing span groups

For this experiment, it was critical to accurately classify participants as 4-Counters and 4-Subitizers based on their behavioral data from enumerating unstructured arrays in the original Feeding Nemo task (see Experiment 1 Method). Calculations of subitizing capacity stemmed from well-substantiated evidence that reaction time dynamics for subitizing follow a bilinear function (Chi & Klahr, 1975; Klahr & Wallace, 1976; Trick & Pylyshyn, 1994). Previous studies have used a variety of methods to derive subitizing capacity, including fitting bilinear and even exponential models to reaction times and accuracy (Ester et al., 2012; Pagano, Lombardi, & Mazza, 2013; Pagano & Mazza, 2012). For the purposes of Experiment 2, however, it was unnecessary to identify the exact point of intersection between the two lines in a bilinear function fit to the data. Rather, the criteria for group classification was whether a participant subitized four.

I took a least-squares approach to classifying participants as 4-Counters or 4-Subitizers. For each participant, lines were fit to reaction times for unstructured arrays in the known subitizing range (1-3 dots) and known counting range (5-7 dots). Participants' reaction times for four dots were compared to their specific subitizing and counting functions, and the squared error was calculated between the predicted reaction time for four and observed reaction time for four. Participants were classified as 4-Counters if the squared error was greater for the subitizing function, indicating that the counting function was a better fit for the observed reaction time for four dots. Participants were classified as 4-Subitizers if the squared error was greater for the counting function, indicating that the subitizing function was a better fit for the observed reaction time for four dots.

As a precaution, group classifications were verified by visually inspecting line graphs of each participant's reaction time data. For one participant, the observed reaction time for four dots was greater than the observed reaction time for five dots, resulting in a larger squared error for the counting function. Consequently, this participant had been classified as a 4-Subitizers even though he or she clearly did not subitize four. This participant was re-classified as a 4-Counter.

For Experiments 2A and 2B, out of 43 participants, 29 were classified as 4-Counters and 14 were classified as 4-Subitizers. Experiment 2C had a smaller sample of 24 participants, with 15 4-Counters and 9 4-Subitizers.

Experiment 2A: How is groupitizing influenced by the maximum size of subgroups in an array?

Hypotheses

The groupitizing mechanism described by Starkey and McCandliss proposes that individuals subitize each subgroup, and then use their knowledge of number composition to access the overall set size from the combination of subgroup quantities, possibly via combining the sets. Experiment 2A examined enumeration latencies and enumeration slopes for three-group arrays that did (SubMax 4) or did not (SubMax 3) contain a subgroup of four dots. Enumeration latencies are informative about the process of subitizing in enumeration of grouped arrays, and the slopes across these latencies are informative about the fluency with which children can access the composition of a number after the quantities of subgroups have been determined.

Enumeration latencies were analyzed to examine the influence of subgroup magnitude on enumeration speed for children in the two subitizing span groups. Based on hypothesized role of subitizing in the groupitizing model, I predicted that 4-Counters would require serial counting to determine the quantity of the four-dot subgroup, and would therefore take significantly longer to enumerate SubMax 4 than SubMax 3 arrays. In contrast, I predicted that 4-Subitizers would not have significantly different enumeration latencies for arrays in these two conditions, because they should be able to subitize the four-dot subgroup.

Therefore, I anticipated a three-way interaction between set size, SubMax, and subitizing capacity, which would suggest that enumeration latencies for children in the two subitizing span groups are affected differently by the presence of a four-dot subgroup. Furthermore, I predicted that a series of follow-up paired sample t-tests would show that 4-Counters have greater enumeration latencies for SubMax 4 arrays than for SubMax 3 arrays. For the 4-Subitizers, enumeration latencies for the two grouping conditions should not significantly differ.

The second analysis used slope as a groupitizing fluency measure that controls for individual differences in speed of processing. As described in Experiment 1, fluent groupitizing is characterized by flat slopes, suggesting that once the quantity of each subgroup is obtained, children can access their knowledge of number composition with the same fluency even as set size increases. Individuals with flat slopes are not using serial counting or addition in the process of enumerating, as both of these processes would produce a positive slope (Butterworth et al., 2001; Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). Comparing 4-Counters' and 4-Subitizers' slopes for the two grouping conditions provided evidence for whether the presence of a four-dot subgroup affects how fluently children can access number based on the quantities of subgroups.

Previous studies have shown a link between subitizing range and math development (Aunio & Niemivirta, 2010; Clements, 1999; Fischer et al., 2008), although the basis of this relationship is somewhat unclear. One possibility is that being able to subitize a quantity either reflects, or enables a more developed concept of how that quantity operates in many contexts. This might include insight into how that quantity can serve as a subset in the composition of other, larger quantities.

Groupitizing provides a unique opportunity to investigate this. If it is the case that subitizing a quantity is related to an understanding of the characteristics of that quantity, then 4-Subitizers should have flat slopes for both grouping conditions because their greater subitizing span would suggest a more developed concept of four, as well as more advanced math skills. They may be able to use their knowledge of number composition more fluently, and may be able to more quickly connect the combination of subgroup quantities to the correct overall array size. Therefore, I anticipated an interaction of subitizing span group and SubMax condition, which would indicate that SubMax 3 and SubMax 4 arrays affect slopes differently for children with different subitizing capacities. I also predicted that follow-up t-tests would demonstrate that 4-Counters and 4-Subitizers have similar slopes for SubMax 3 arrays, but that 4-Counters would have a significantly steeper slope than 4-Subitizers for SubMax 4 arrays.

For within-subitizing span group comparisons, I expected that 4-Counters would have steeper slopes for SubMax 4 arrays than for SubMax 3 arrays, but that 4-Subitizers' slopes would not significantly differ for the two grouping conditions. This would demonstrate that 4-Subitizers enumerate arrays with four-dot subgroups more fluently than 4-Counters, suggesting that they can recruit and apply their knowledge of number composition more fluently.

Results

Two approaches were taken in investigating the impact of subgroup size on enumeration for children in the two subitizing span groups. First, enumeration latencies were compared for the two SubMax conditions at each set size. Second, for each participant, slopes were calculated across set sizes 6, 7, and 8 for each grouping condition, resulting in a SubMax 3 slope and a SubMax 4 slope. I examined these measures for 4-Counters and 4-Subitizers to determine whether grouping condition has different effects on their enumeration fluency.

Participants' median enumeration latencies for target set sizes in the SubMax 3 and SubMax 4 conditions were submitted to a repeated measures ANOVA with two within-subjects factors (Set Size and SubMax Condition) and Subitizing Span Group as a between-subjects factor. Greenhouse-Geisser corrections were applied as appropriate. Results revealed a main effect of Set Size, $F(2,70) = 16.594$, $p<0.001$, and a main effect of SubMax Condition, $F(1,35) = 55.816$, $p<0.001$, suggesting that the total quantity of dots in an array and whether it contained a four-dot subgroup both significantly impacted the time required to enumerate the array. An interaction between Set Size and SubMax Condition, $F(2,70) = 5.682$, $p<0.05$, suggested that this impact of overall quantity on enumeration latency was different for arrays in the two SubMax conditions. An interaction between Set Size and Subitizing Span Group, $F(2,70) = 10.768$, $p<0.001$, indicated that the overall set size of an array influenced enumeration latency differently for 4-Counters and 4-Subitizers. Finally, a three-way interaction between Set Size, SubMax Condition, and Subitizing Span Group was significant, $F(2,70) = 7.547$, $p<0.001$ (Figure 6). This result was in accordance with the original prediction, and it suggests that the presence of a four-dot subgroup within an array had different effects on enumeration latency across the three target set sizes for 4-Counters and 4-Subitizers.

To investigate this interaction further, a series of paired sample t-tests were conducted within groups of 4-Counters and 4-Subitizers, in order to compare reaction times at each set size between the two SubMax conditions. As expected, for all set sizes, the 4-Counters took significantly longer to enumerate SubMax 4 arrays than SubMax 3 arrays (Table 7; Figure 6). This supports the idea that participants who cannot subitize 4 must use a serial counting process to derive the quantity of the subgroup of 4 dots, accounting for the reaction time difference between the two grouping conditions.

Interestingly, participants who subitized 1-4 also took significantly longer to enumerate SubMax 4 arrays than SubMax 3 arrays (Table 7; Figure 7). This unexpected result contradicted the prediction of no difference between SubMax 4 and SubMax 3 enumeration latencies for 4-Subitizers. This suggests that even children who can subitize 4 may not be subitizing the four-dot subgroup when enumerating SubMax 4 arrays.

Table 7. T values for paired sample t-tests comparing enumeration latencies for SubMax 3 and SubMax 4 arrays within 4-Counters and 4-Subitizers.

	Set Size		
	6	7	8
4-Counters	3.329 *	4.939 **	10.542 **
4-Subitizers	3.604 *	2.657 *	5.397 **

* p < 0.05 ** p < 0.001

In the second analysis, each participant's slopes across set sizes 6, 7, and 8 in the SubMax 3 and SubMax 4 conditions were submitted to a repeated measures ANOVA with one within-subjects factor (SubMax Condition) and one between-subjects factor (Subitizing Span Group). Greenhouse-Geisser corrections were applied as necessary. This ANOVA revealed a main effect of SubMax Condition, $F(1,35) = 27.076$, $p<0.001$, indicating that slopes were impacted by the presence of a four-dot subgroup in an array. As predicted, the interaction between SubMax Condition and Subitizing Span Group was also significant, $F(1,35) = 35.977$, $p<0.001$, suggesting that enumeration fluency for SubMax 3 and SubMax 4 arrays may be different for 4-Counters and 4-Subitizers.

To follow up on this interaction effect, independent samples t-tests were calculated (not assuming equal variance). First, slopes were compared between 4-Counters and 4-Subitizers for both grouping conditions. For the SubMax 3 condition, 4-Counters and 4-Subitizers did not demonstrate differences in enumeration slope. However, as predicted, 4-Counters demonstrated a significantly steeper slope for the SubMax 4 condition than 4-Subitizers, $t(27.681) = 6.116$, $p<0.001$ (Figure 6). Regardless of enumeration *latency*, 4-Counters required more additional time for each incremental increase in set size than 4-Subitizers when there was a four-dot subgroup in the array. Even though 4-Subitizers did need more time to enumerate SubMax 4 arrays than SubMax 3 arrays, they were able to enumerate SubMax 3 arrays with greater fluency than the 4-Counters.

Finally, paired samples t-tests were calculated to compare slopes for the SubMax 3 and SubMax 4 arrays within each subitizing span group. The 4-Counters had significantly steeper slopes in the SubMax 4 condition than the SubMax 3 condition, $t(24) = 8.662$, $p<0.001$, suggesting that the presence of a four-dot subgroup decreased their enumeration fluency and resulted in substantially longer latencies for set size 8 than for set size 6. The 4-Subitizers, on the other hand, did not have significantly different slopes between the two SubMax conditions (Figure 6). Comparing slopes within subitizing span groups for SubMax 3 and SubMax 4 conditions demonstrates that the presence of a four-dot subgroup slowed enumeration fluency for 4-Counters, but did not slow enumeration fluency for 4-Subitizers. This suggests a relationship between subitizing skill and fluency with accessing and using knowledge of number composition.

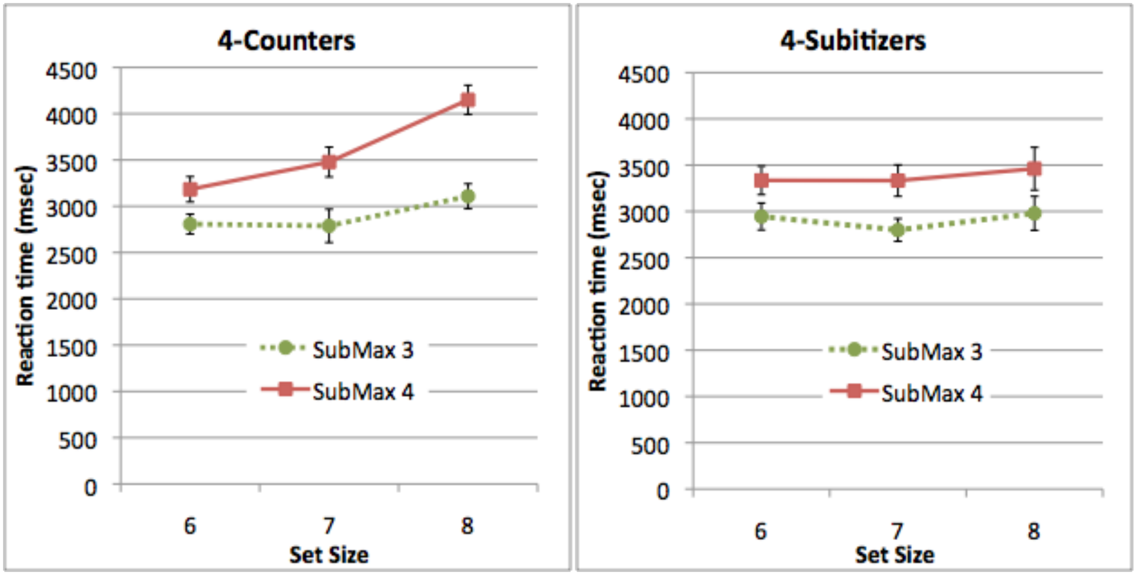

Figure 6. Comparison of enumeration latencies for SubMax 3 and SubMax 4 arrays, for 4-Counters and 4-Subitizers

Discussion

Experiment 2A focused on comparing enumeration latencies and slopes when the number of subgroups was held constant, and some arrays contained a four-dot subgroup. Results demonstrated two effects that are important in understanding the cognitive dynamics of groupitizing. First, regardless of subitizing capacity, participants took longer to enumerate arrays with a four-dot subgroup. This unexpected result suggests that even children who can subitize four dots do not do this in the context of a larger array. Previous studies have shown that elementary school-aged children whose subitizing skills are still developing are able to subitize four, but still count in some circumstances (Svenson & Sjöberg, 1983). Participants were classified into subitizing span groups based on their performance on unstructured arrays of 1-8 dots in the original Feeding Nemo task. It is possible that some children who were classified as 4-Subitizers had just begun to subitize four, and reverted back to a counting strategy when four dots were presented as a subgroup in a larger array.

Second, although 4-Subitizers did not appear to be subitizing the four-dot subgroup, they were able to access the overall array size just as fluently as they did when there was no four-dot subgroup, demonstrated by slopes that were not significantly different. (It is possible that this comparison was underpowered to detect a difference between SubMax 3 and SubMax 4 slopes for this group due to the small sample size; however, this is unlikely, as significant effects were detected in other within-group comparisons.) The 4-Subitizers appeared to have flat slopes for both SubMax 3 and SubMax 4 arrays, indicating that the four-dot subgroup has no impact on how fluently they can access their knowledge of number composition.

In contrast, 4-Counters had steeper slopes when a four-dot subgroup was present, meaning that it took more effort to access the total number of dots as set size increased. This positive slope suggests that the 4-Counters relied on a cognitive process that is influenced by set size. They may have resorted to serial counting, or they may have added the subgroup quantities; the use of either of these processes would have produced a positive slope (Butterworth et al., 2001; Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). In either case, it is clear that 4-Counters were unable to use a process that is not set size-sensitive to directly access overall array size from the quantities of the subgroups, as this would have produced a flat slope.

Moreover, when slopes were compared between the 4-Counters and 4-Subitizers for each grouping condition, results showed that 4-Subitizers had greater fluency for

SubMax 4 arrays than 4-Counters did. This result highlights a difference between 4-Counters and 4-Subitizers: regardless of the process used to derive the quantities of the subgroups, 4-Subitizers are better at making connections between combinations of subsets and overall array size. They may be more familiar with three-set combinations that include a set of four, while 4-Counters are still inexperienced with manipulating sets of four.

Experiment 2B: How is groupitizing influenced by the number of subgroups in an array?

Hypotheses

The hypothesized model of groupitizing proposes that subgroups are subitized in the process of groupitizing, and that incremental set combination may enable a direct connection between subgroup quantities and overall array size. Therefore, I predicted that all participants, regardless of subitizing capacity, would enumerate arrays with two subgroups faster than arrays with three subgroups. I anticipated that a comparison of enumeration latencies for two-group and three-group arrays would result in a main effect of grouping condition. This lead to the prediction that follow-up paired sample t-tests would reveal that, for each set size, arrays in the two-group condition are enumerated faster than arrays in the three-group condition. This finding would suggest that children are accessing the quantity of each subgroup one at a time, not in parallel. In addition, this would support the hypothesis that children recruit their set combination skills to combine the quantities of the subgroups one at a time. If reaction times for these grouping conditions were the same, however, this would indicate that children are able to subitize the subgroups in parallel, or that they are able to access the overall array size through a cognitive process other than knowledge of set combinations.

A second analysis comparing 4-Counters' and 4-Subitizers' slopes for two-group and three-group arrays explored differences in their fluency in enumerating arrays with different numbers of subgroups. Due to links between subitizing and math development, I predicted that children with a greater subitizing span can groupitize arrays more fluently regardless of the number of subgroups. The 4-Subitizers should have concepts of number that have developed to include the various combinations of sets that constitute numbers, and their ability to access this information should not be affected by the number of subgroups in the array. The 4-Counters may not have reached this stage, resulting in less fluent enumeration for two-group and three-group arrays. Therefore, I predicted that 4-Subitizers would demonstrate flatter slopes than 4-Counters for both grouping conditions. The 4-Counters should have substantial positive slopes in both grouping conditions. I anticipated finding a main effect of subitizing span group, and that follow-up t-tests would reveal steeper slopes for 4-Counters than for 4-Subitizers for both grouping conditions.

Results

All participants' median enumeration latencies for the Two Group, SubMax 4 and Three Group, SubMax 4 conditions were submitted to a repeated measures ANOVA with two within-subjects factors: Set Size (6, 7, and 8) and Grouping Condition (Two Groups and Three Groups). Results revealed a main effect of Set Size, $F(2,74) = 30.877$, $p<0.001$, suggesting that regardless of condition, as set size increases, the time required for successful enumeration also increases. ANOVA results also revealed a main effect of Grouping Condition, $F(1,37) = 36.816$, $p<0.001$. This main effect confirms that one grouping condition resulted in faster enumeration than the other grouping condition. Interestingly, the ANOVA also revealed an interaction of Set Size and Grouping Condition, $F(2,74) = 3.516$, $p<0.05$, indicating that the effect of set size on enumeration latency is different for arrays with two and three subgroups.

To follow up on the main effect of grouping condition, a series of paired sample t-tests were calculated to compare enumeration latencies for the two grouping conditions within each set size. T-test results for set size 6 showed that the two-group array was enumerated significantly faster than the three-group array, $t(39) = -5.408$, $p<0.001$. T-test results for set size 7 and 8 yield the same result ($t(41) = -5.523$, $p<0.001$ and $t(40) = -6.903$, $p<0.001$, respectively).

To investigate the interaction between Set Size and Grouping Condition, participants' slopes across set sizes 6-8 were calculated for the two grouping conditions, creating a Three-Group, SubMax 4 slope (also used previously in Experiment 2A) and a Two-Group, SubMax 4 slope. Participants' slopes were entered into a paired sample t-test. Slopes for arrays with three groups were significantly steeper than slopes for two groups, demonstrating greater fluency for two-group arrays. The next analysis will explore this further, but one possible explanation is that children gain more exposure to two-set combinations than three-set combinations in elementary school (for example, in the form of two-operand addition and subtraction). Therefore, two-set combinations may be more readily accessible than three-set combinations when groupitizing.

For the next analysis, participants' slopes were used instead of enumeration latencies. As described in Experiment 1, slope of enumeration latencies across set sizes controls for individual differences in speed, thereby providing a measure of fluency. This is especially important in this analysis because, as Experiment 2A demonstrated, subitizing capacity significantly impacts enumeration latency. An analysis of group differences using enumeration latency would potentially be confounded by constraints placed on the 4-Counters, in particular, by subitizing capacity. However, the impact of serial counting on reaction time will be the same across all set sizes, so including slope in this analysis eliminates this potential confounding factor.

To examine how enumeration fluency varies between the two subitizing span groups for two- and three-group arrays, participants' slopes were submitted a repeated-measures ANOVA with Grouping Condition as a within-subjects factor and Subitizing Span Group as a between-subjects factor. Results revealed a main effect of Grouping Condition, $F(1,36) = 5.619$, $p<0.05$, indicating that slopes were different for two- and three-group arrays (shown in the interaction effect from the previous ANOVA). Importantly, the interaction between Grouping Condition and Subitizing Span Group was significant, $F(1,36) = 6.294$, $p<0.05$.

To examine whether slopes differed for two- and three-group arrays within each subitizing capacity group, paired samples t-tests were calculated separately. The 4-Counters' slopes for three-group arrays were significantly steeper than slopes for two-group arrays, $t(23) = 5.022$, $p<0.001$. However, for 4-Subitizers there was no significant difference between slopes for two- and three-group arrays. As in Experiment 2A, it is possible that this analysis is simply underpowered to detect significant differences, given the small number of participants classified as 4-Subitizers. However, Figure 7 shows that reaction time slopes for 4-Subitizers appear to be parallel, which suggests that this result may hold true for a larger sample.

To determine how slopes varied between 4-Counters and 4-Subitizers, independent samples t-tests were calculated. For two-group arrays and three-group arrays, 4-Counters had substantially steeper slopes than 4-Subitizers ($t(27.681) = 6.116$, $p<0.001$, and $t(24.125) = 3.789$, $p<0.001$, respectively; equal variances not assumed). Figure 7 shows significant positive slopes for both types of arrays for 4-Counters, compared to relatively flat slopes for 4-Subitizers.

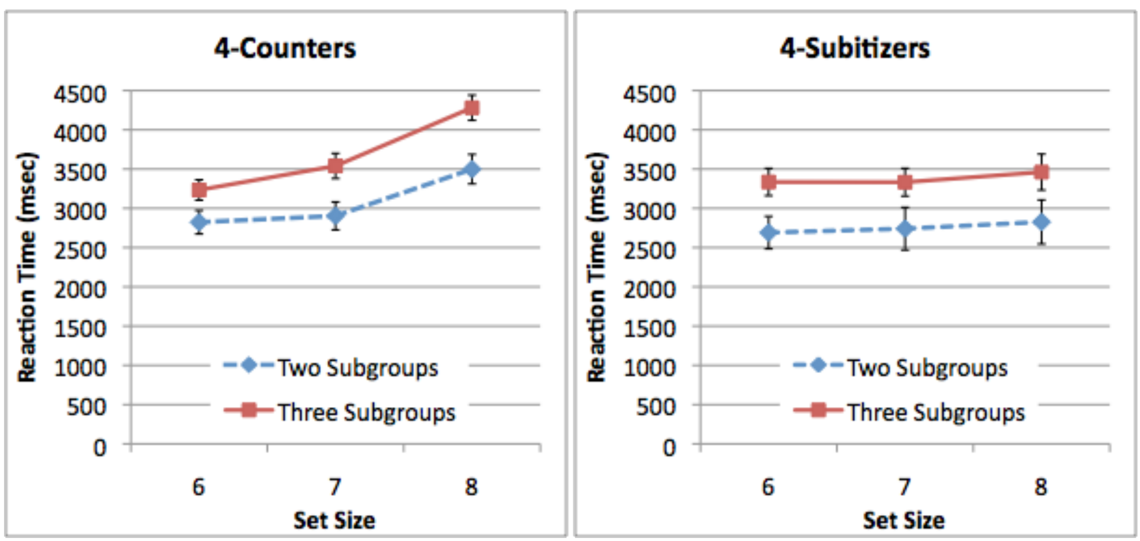

Figure 7. Comparison of enumeration latencies for two- and three-group arrays, for 4-Counters and 4-Subitizers

Discussion

These findings suggest several points. First, regardless of subitizing capacity, it took participants longer to enumerate arrays with more subgroups, even when the overall set size was held constant. This could indicate that subgroups were subitized one at a time during groupitizing, and not in parallel, resulting in longer enumeration latencies for arrays with more subgroups. It also suggests that participants are recruiting their knowledge of the numbers that result from specific combinations of sets. It makes sense that this would take longer for children when presented with three sets, compared to two sets. In school, children gain experience with combinations of two sets through learning traditional two-operand addition and subtraction. They are less often faced with combinations of three sets. Even if children's concepts of numbers are advanced enough to include the combinations of sets that a number can be decomposed into, combinations of three sets are likely to be less readily accessible than the two-set combinations, to which they are exposed more frequently. Experiment 2C will address this further in a symbolic task that eliminates the subitizing component of groupitizing.

Second, in a within-groups comparison of slopes, 4-Counters demonstrated an effect of grouping condition on groupitizing fluency, showing that they were less fluent at accessing numbers when given a combination of three sets than a combination of two sets. The 4-Counters' steeper slopes for three-group arrays suggest that they relied on a serial process of counting elements in the array, or incrementally adding the subgroups together. In contrast, 4-Subitizers did not demonstrate this effect; rather, they had relatively flat slopes for both two- and three-group arrays. This indicates that 4-Subitizers did not need to rely on any serial processes - counting or adding - but instead were able to directly access the total number from the quantities of subsets in both conditions. It is possible that this finding could be due to an underpowered comparison, given the small group size for 4-Subitizers; however, this is unlikely, because effects were successfully detected in other analyses, suggesting that there may be truth to this finding.

Third, in a between-groups comparison of slopes, 4-Counters had steeper slopes for both types of arrays, suggesting that the 4-Subitizers groupitize more fluently, regardless of the number of subgroups. This builds on findings from Experiment 2A, which suggest that 4-Subitizers also groupitize more fluently regardless of the maximum quantity of subgroups. Taken together, these results provide compelling evidence that 4-Subitizers can more readily access the overall size of an array from the quantities of subgroups, regardless of the number of sets involved. This reflects a greater awareness of number composition and ability to apply stored set combination information to the process of enumerating.

Experiment 2C: How do the maximum magnitude of digits and the number of digits influence enumeration of symbolic arrays?

The goal of Experiment 2C was to investigate reaction time dynamics when subgroups were presented as symbolic digits, eliminating any subitizing and counting components of groupitizing. By presenting combinations of digits without arithmetic symbols (which might inadvertently push children toward using addition), the FN Symbolic task aimed to isolate how children connect subgroup quantities to the total magnitude of a stimulus. Thus, this task allowed further investigation into set combination processes that may be recruited in groupitizing, and children's grasp of the number concepts that enable this.

Hypotheses

The following analyses built upon findings from Experiments 2A and 2B that pertained to children's fluency in accessing array quantities via number composition. Response latencies for symbolic stimuli addressed findings related to enumeration speed, and slopes across response latencies for set sizes 6-8 addressed findings related to groupitizing fluency.

Comparing stimuli with and without four

Experiment 1A examined enumeration latencies and slopes for three-group arrays that did (SubMax 4) and did not (SubMax 3) contain a subgroup of four dots. Surprisingly, both 4-Counters and 4-Subitizers required more time to enumerate SubMax 4 arrays than SubMax 3 arrays. In Experiment 2C, response latencies for symbolic SubMax 3 and SubMax 4 stimuli were compared within subitizing span groups to investigate whether this effect was replicated for symbolic digits. I hypothesized that this result was due to the time required to count four dots; if this is true, both 4-Counters and 4-Subitizers should not have significantly different response latencies for symbolic SubMax 3 and SubMax 4 arrays, because no subitizing or serial counting is needed.

The 4-Subitizers had the same slopes for SubMax 3 and SubMax 4 arrays, suggesting that even though the four-dot array impacted their enumeration speed, it did not affect fluency. The 4-Counters, in contrast, had steeper slopes for SubMax 4 arrays than SubMax 3 arrays, suggesting that their ability to directly access array magnitude from a combination of digits was affected by the presence of a four-dot subgroup. Slopes for symbolic SubMax 3 and SubMax 4 arrays were compared within each subitizing span group. I predicted that this effect would be replicated, suggesting that differences in slope were due to fluency in figuring out the total number that the subset quantities comprise (and not related to how fluency in determining subgroup quantities).

A between-groups comparison of slopes for non-symbolic SubMax 3 and SubMax 4 arrays showed that 4-Counters had steeper slopes for SubMax 4 arrays than 4-Subitizers. This group difference demonstrates that 4-Counters' fluency is affected by the four-dot subgroup. Experiment 2C followed up on this by comparing SubMax 4 slopes for 4-Counters and 4-Subitizers, addressing whether this difference in non-symbolic fluency can be attributed to accessing number composition.

Comparing two-digit and three-digit stimuli

In the classroom, arithmetic is typically taught in the form of two-operand problems. Children receive less exposure to three-addend arithmetic, which suggests that their concept of numbers above the subitizing range may have developed to include combinations of two smaller sets that make up a number, but not combinations of three smaller sets. In Experiment 2B, results showed that both 4-Counters and 4-Subitizers enumerated two-group arrays more quickly than three-group arrays. However, it was unclear whether this reflected incremental subitizing or faster access to overall array size for two sets than three sets, potentially due to a greater level of experience with two-set combinations.

To address this, response latencies from the FN Symbolic task were compared for two-digit and three-digit stimuli to examine the impact of the number of digits on response speed. If both 4-Counters and 4-Subitizers had faster response latencies for two-digit stimuli, this would suggest that at least part of what contributes to this effect in non-symbolic grouped arrays is how quickly children can determine the overall magnitude of an array given the sets that comprise the array. Another potential outcome was for only one subitizing span group to have significantly different response latencies, which would highlight a difference between the groups. This might support the link between subitizing capacity and higher-level math skills, which has been suggested in previous research (Aunio & Niemivirta, 2010; Clements, 1999; Fischer et al., 2008).

Slopes across response latencies for set sizes 6-8 in two-digit and three-digit were also compared within and between 4-Counters and 4-Subitizers. Whereas analyses of response latencies address the speed with which overall magnitudes are accessed, analyses of slope address fluency after controlling for individual differences in processing speed. Thus, this analysis of slopes accounts for any differences in how long it takes for children to process the digits (as this will be the same across set sizes). As discussed in previous sections, lower slopes indicate that set size has less impact on response speed, providing evidence for direct access to array magnitude.

Finally, Experiment 2B results suggested that 4-Counters had significantly steeper slopes for three-group arrays than for two-group arrays, indicating use of a cognitive process that is affected by set size (counting or adding). The 4-Subitizers had flat slopes, indicating direct access to overall quantity with no effect of set size. Slopes for the two-digit and three-digit symbolic stimuli will be compared within and between subitizing span groups. These analyses will show whether this is simply due to the cognitive processes used to derive subgroup quantity, or whether it reflects differences in children's ability to use insights into number composition.

Results

Analyses for Experiment 2C were driven by findings from Experiments 2A and 2B, and served as follow-up to these findings. In addition, the participant sample for Experiment 2C was relatively small, especially when separated into groups of 4-Counters and 4-Subitizers. Therefore, in lieu of repeated-measures ANOVAs, a series of independent and paired-sample t-tests were used to conduct analyses for this experiment.

Comparing stimuli with and without 4

For each subitizing span group, paired samples t-tests to compare SubMax 3 and SubMax 4 response latencies were calculated for each of the three target set size. Neither 4-Counters nor 4-Subitizers demonstrated significantly different response latencies for SubMax conditions at any set size. Although it cannot be concluded that response latencies are equal based on null effects, Figure 8 shows that response latencies for SubMax 3 and SubMax 4 were almost entirely overlapping for both 4-Counters and 4-Subitizers. Thus, this tentatively suggests that differences in enumeration latencies for non-symbolic arrays were due to participants needing more time to serially count the subgroup of four, rather than participants having more difficulty accessing the overall quantity.

Within-groups paired samples t-tests comparing slopes for SubMax 3 and SubMax 4 showed that neither 4-Counters not 4-Subitizers had significantly different slopes for the two SubMax conditions. This result is not surprising, given that their enumeration latencies did not significantly differ. This result tentatively suggests that perhaps the difference between 4-Counters' non-symbolic SubMax 3 and SubMax 4 slopes was not in fact related to their ability to access the overall magnitude from subgroup quantities. Instead, the process of deriving the quantity of the four-dot subgroup may have contributed to 4-Counters' steeper SubMax 4 slope.

Next, independent samples t-tests were calculated to compare 4-Counters' SubMax 3 and SubMax 4 slopes to 4-Subitizers' slopes. No significant difference was found between the groups' SubMax 3 slopes, but 4-Counters' SubMax 4 slopes were significantly steeper than 4-Subitizers' SubMax 4 slopes, $t(15.650) = 2.594$, $p<0.05$ (corrected for unequal variances). Figure 8 shows these between-group differences in SubMax 4 slope. This matches the effect found in Experiment 2A, which supports the idea that between-group differences in fluency are attributable to 4-Counters having more difficult with the number composition component of groupitizing when one of the subgroup quantities is four.

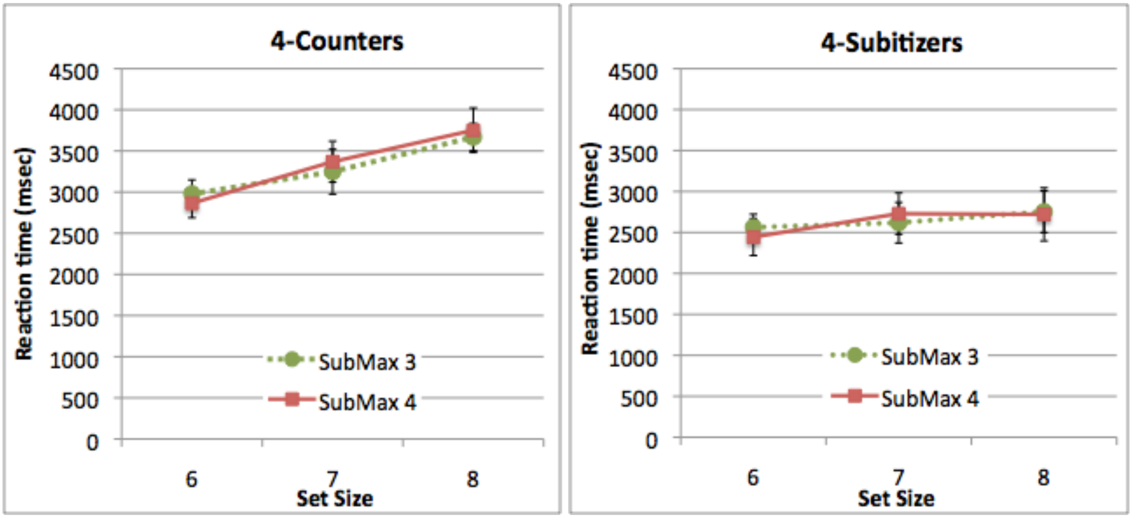

Figure 8. Comparison of enumeration latencies for symbolic SubMax 3 and SubMax 4 arrays, for 4-Counters and 4-Subitizers

Comparing two-digit and three-digit stimuli

Paired sample t-tests were calculated for each subitizing capacity group to compare response latencies for two-digit and three-digit stimuli. The 4-Counters and 4-Subitizers both demonstrated significantly faster response latencies for two-digit than three-digit stimuli. This means that conclusions still cannot be drawn about whether subitizing occurs incrementally or in parallel. However, this important result suggests that the time it takes to determine array quantity from a set of numbers plays a role in differences in enumeration latencies for non-symbolic arrays in the two-group and three-group conditions.

Note that in Figure 9, it appears that response latencies for two-digit stimuli in set size 8 are substantially faster than for set sizes 6 and 7; this is likely because 8 was presented as 4 and 4, which is considered a "tie," and is an exception to the set size effect in arithmetic. It is possible that participants responded faster to these stimuli because they can retrieve these ties from memory so quickly that they did not engage in the same cognitive process needed for responding to set sizes 6 and 7.

Independent samples t-tests compared slopes for two- and three-digit stimuli between 4-Counters and 4-Subitizers. Slopes were not different for two-digit stimuli, but 4-Counters had significantly steeper slopes than 4-Subitizers for three-digit stimuli, $t(15.560) = 2.594$, $p<0.05$ (Figure 9). This result supports conclusions from Experiment 2B, which proposed that 4-Counters relied on a serial cognitive process that is affected by set size (like counting or incrementally adding sets), rather than directly accessing the value of the array from the sizes of the subgroups. Moreover, this result suggests that this less efficient cognitive process was introduced when 4-Counters connected the subgroup quantities to the overall quantity of the array, because serial counting would not have been required in the symbolic task. Therefore, the 4-Counters likely had to add the digits together one at a time for three-digit stimuli, while 4-Subitizers could use the same process of direct access that they used for two-digit stimuli.

Finally, paired samples t-tests compared slopes for two-digit and three-digit stimuli for each subitizing capacity group. The 4-Counters had significantly steeper slopes for three-digit than two-digit stimuli (Figure 9); this indicates a set size effect on response latency, further suggesting that they may have been incrementally adding numbers for the three-digit stimuli. The 4-Subitizers, however, did not have significantly different slopes for two- and three-digit stimuli, demonstrating that they did not need to rely on adding to determine overall quantity.

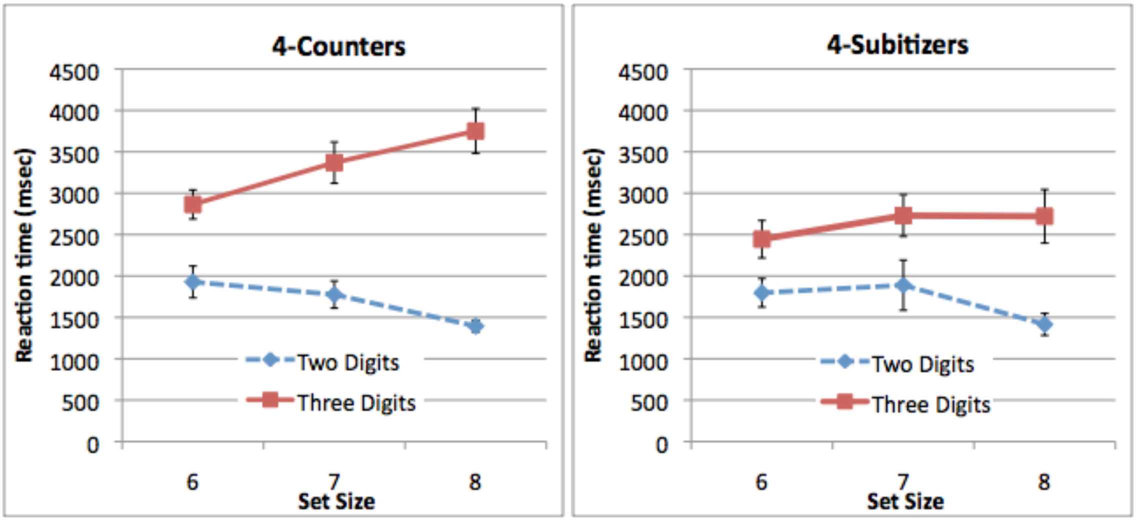

Figure 9. Comparison of enumeration latencies for symbolic two-digit and three-digit stimuli, for 4-Counters and 4-Subitizers

Discussion

The purpose of Experiment 2C was to test findings from Experiments 2A and 2C in a task with analogous grouping conditions, but that did not involve subitizing or counting. Many of the analyses in Experiment 2C yielded null effects, which must be interpreted cautiously. Nonetheless, results from this symbolic task made several important points.

Symbolic data supported findings that suggest that 4-Counters and 4-Subitizers both (a) counted four-dot subgroups in non-symbolic arrays, and (b) took longer to respond to three-group stimuli than two-group stimuli. However, compared to 4-Subitizers, 4-Counters were less fluent at enumerating non-symbolic and symbolic arrays when they contained a subgroup of four dots, or when a combination of digits contained a four. This was reflected in steeper slopes across enumeration latencies for non-symbolic and symbolic SubMax 4 stimuli. The 4-Counters also had more difficulty than 4-

Subitizers when an array contained three subgroups rather than two subgroups, demonstrated by steeper slopes across enumeration latencies for two-group arrays and two-digit stimuli. More specifically, results from within-group comparisons of symbolic slopes suggest that the 4-Counters actually add the subgroup quantities together, which accounts for the set size effect, rather than engaging in a process of directly accessing number combinations.

Meanwhile, 4-Subitizers' groupitizing fluency does not appear to be affect by the presence of a four-dot subgroup or a symbolic 4, or by the number of subgroups or digits in an array. Their relatively flat slopes across the various grouping manipulations demonstrate that regardless of the number or size of subgroups, they are able to directly (without a set size effect) access the overall quantity of an array after determining the subgroup quantities.

Overall, conclusions from Experiment 2C are somewhat limited by a small participant sample, but offer clarification for some non-symbolic findings and highlight important differences between participants who can and cannot subitize four items.

Integrating findings from Experiment 2

Experiment 2 successfully highlighted important similarities and differences between children who can subitize 4, and children who cannot. First, this experiment demonstrated that both 4-Counters and 4-Subitizers have to count the four-dot subgroups in non-symbolic arrays. This was an unexpected result, but is supported by previous literature suggesting that when children begin to be able to subitize four, they do not use this strategy under all circumstances (Svenson & Sjöberg, 1983). Given the age range of participants in this experiment, it is likely that most, if not all, of the 4-Subitizers were just beginning to subitize four. They may have subitized four in the original Feeding Nemo paradigm, upon which their group classification was based, but when a subgroup of four appeared surrounded by other subgroups of dots, these children reverted back to serial counting.

Second, both groups of participants also took longer to respond to three-group stimuli than two-group stimuli in the non-symbolic and symbolic tasks, which together suggest that it is easier to access combinations of two sets than combinations of three sets when trying to figure out the total magnitude of an array. One might argue that this reflects an incremental process of mentally adding the quantities together, but this process would produce a set size effect that is not present in the data. Instead, I propose that two-set combinations are easier to access than three-set combinations, due to more exposure to two-operand arithmetic in the classroom and more experience manipulating two sets of items than three sets of items. As children's concept of number develops, they likely associate these two-set combinations with a number before three-set combinations (e.g., concept of 7 also includes the combination of 4 and 3 before it includes 4, 2, and 1). Consequently, when faced with two subsets or two digits in an array, these combinations are more familiar and connections are made more quickly than when children are faced with three subsets or digits.

Experiment 2 also revealed some interesting group differences between 4-Counters and 4-Subitizers. Compared to 4-Subitizers, 4-Counters had more difficulty directly accessing and applying their knowledge of number composition when an array contained a subgroup of four dots, or when there were three instead of two subgroups. These results were echoed with symbolic data in Experiment 2C. Rather than directly accessing array quantities via number knowledge, non-symbolic and symbolic data together suggest that 4-Counters are adding the subgroups together, producing a positive slope that reflects a significant set size effect. In contrast, 4-Subitizers' enumeration fluency is not impacted by changes in the number or maximum quantity of subgroups, reflecting greater ability to handle sets of four in the context of enumeration. This supports the notion that the ability to subitize quantity may be linked to a more developed concept of that quantity and insight into its associated characteristics, accounting for the relationship between subitizing skill and math development that is noted in the literature (Aunio & Niemivirta, 2010; Clements, 1999; Fischer et al., 2008).

To summarize, Experiment 2 provides new support for the proposed mechanism of groupitizing by demonstrating how manipulations to grouping structure impact enumeration latency and fluency. In addition, non-symbolic and symbolic findings from Experiment 2 contribute to understanding how developments in subitizing capacity map onto developments in children's knowledge of number composition. However, because of the small sample size and inherent risks in calculating multiple t-tests, future studies should attempt to replicate these findings in a larger sample of children with an even distribution of 4-Counters and 4-Subitizers.

CHAPTER IV

EXPERIMENT 3: EVALUATING THE FEASIBILITY OF MEASURING N2pc AND VISUAL WORKING MEMORY IN CHILDREN

Electrophysiological research has examined the N2pc component as an index of selective attention in visual search paradigms (Drew & Vogel, 2008). The N2pc reflects the difference between amplitudes for posterior electrodes in a particular hemisphere when a target is presented in the ipsilateral and contralateral visual fields. Tasks that elicit an N2pc effect present target items among distractors, so that participants must individuate these items from the entire array. Some N2pc tasks require participants to detect change among target items (in color, shape, or orientation), but recently, multiple studies have employed small-number enumeration tasks to elicit the N2pc.

These studies have reported that the amplitude of the N2pc component is modulated by the number of target items presented (Ester et al., 2012; Pagano & Mazza, 2012). Furthermore, the N2pc reaches an asymptote at the number of items that can be subitized. Ester et al. demonstrated that individual differences in N2pc asymptote were strongly correlated with behavioral measures of subitizing capacity. These results support the theory that subitizing reflects a fixed-capacity model, in which a certain number of items can be attended to and held in working memory without requiring serial shifts of attention (Cowan et al., 2005; Fukuda et al., 2010; Pagano et al., 2013). However, these studies were conducted with adults; to my knowledge, at the time this study was conducted, there were no published electrophysiological studies of children's subitizing ability. Addressing this gap in the literature could help elucidate the neural underpinnings of children's developing enumeration skills.

Experiment 3 applied the notion of a fixed-capacity mechanism to a sample of third- and fourth-grade children. This is a stage of development in which subitizing range increases from 1-3 to 1-4 for some children. Therefore, this population offers significant individual differences in subitizing capacity that may modulate N2pc amplitude. Experiment 3 assessed the feasibility of conducting such an experiment in a sample of children. Because of the clear involvement of subitizing in groupitizing, this experiment could be a step toward understanding the neural substrates of groupitizing.

The subitizing tasks employed in experiments by Ester et al. (2012) and Pagano and Mazza (2012) were highly demanding, even for adults. In Experiment 3, steps were taken to make the task more suitable for children. However, some elements of the task that significantly increase cognitive demand were necessary to elicit an N2pc, and therefore had to be retained. For example, participants must maintain fixation in the center of the display in order for amplitudes to truly reflect whether a hemisphere is ipsilateral or contralateral to the target visual field.

Unfortunately, attrition due to poor data quality and participant dropout was a significant issue. Furthermore, results from remaining participants showed that although behavioral performance on the task demonstrated an ability to respond to stimuli as predicted, the task did not successfully elicit an N2pc component modulated by set size. N2pc amplitudes were calculated according the same procedure as previous studies, but did not follow any predictable pattern, and varied substantially across participants. Overall, results from this feasibility study indicate that there are substantial barriers to successfully measuring a reliable N2pc component in children, which future studies may be able to address.

However, the N2pc task did yield useful behavioral data. Participants' accuracy for each set size was calculated, collapsing across visual field. Subitizing capacity was calculated based on accuracy at each set size. For each participant, this measure of subitizing capacity was then compared to measures calculated from enumeration latency data for unstructured arrays in Feeding Nemo. Since these subitizing capacity indexes were calculated from two different measures in two different tasks, establishing significant overlap between them could strengthen claims about the validity of classifications of 4-Counters and 4-Subitizers.

Participants also completed a computer-based visual working memory (VWM) task, in which they detected changes in the color of one square out of an array of 1, 2, 4, or 8 squares. Consistent with the fixed-capacity model of subitizing, research has demonstrated that VWM capacity and subitizing capacity are strongly correlated (Piazza et al., 2011). However, to my knowledge, no published work has assessed children's VWM capacity in relation to subitizing capacity. Therefore, purpose of this task was to extract a behavioral measure of VWM capacity, which may strengthen claims that subitizing capacity is governed by the limitations of VWM.

Experiment 3 Method

Participants

For Experiment 3, participants include a subset of children from Experiments 1 and 2 who agreed to return to the lab for one two-hour session, conducted in the spring of 2014. A total of 27 children who were in third (n=16) and fourth grade (n=11) participated. Participants' parents were compensated for their time and mileage, and for potential participant discomfort due to EEG procedures. During this session, children participated in a shortened version of the Feeding Nemo enumeration task ("FN Unstructured"), a brief computer-based VWM task, and the N2pc subitizing task.

FN Unstructured

FN Unstructured consisted of only unstructured arrays of 1-8 dots. Stimuli from the original Feeding Nemo task were used. Each set size was presented four times, for a total of 32 trials, which were not separated into blocks. The "fish food store" context was not used for this task, as the task was very brief and children were old enough that they could complete the task with the additional game-like motivation. Sound files were coded and data were processed following the procedure specified in Experiment 1. No participants were excluded from analyses.

FN Unstructured allowed me to establish a concurrent behavioral measure of subitizing range for assessing whether N2pc asymptotes correspond to subitizing capacity. Another measure of subitizing range was derived from accuracy for the N2pc task, as other studies have done (Ester et al., 2012; Pagano & Mazza, 2012). However, data from FN Unstructured provided a way to calculate subitizing range from a separate, less cognitively demanding task than the N2pc task. In addition, FN Unstructured presented items in isolation, without distractors. The purpose of FN Unstructured, therefore, was to provide an additional measure of subitizing capacity, to which N2pc capacity measures and asymptotes could be compared.

N2pc task

This N2pc enumeration task was based on the task used in a previous study by Ester et al. (2012). Stimuli were created in MATLAB 7, and were designed to fit the stimulus parameters specified by Ester et al. Stimuli consisted of 12 or 14 squares on each of a fixation dot. For each trial, the same number of squares (0-6) on each side of the fixation were colored blue or green, and the remaining squares were black. Before beginning EEG data collection, children were instructed to attend to blue or green squares (counterbalanced across participants). For children who were assigned green as a target color, green squares were targets, blue squares were non-targets, and black squares served as distractors. Participants' target color remained the same throughout the session. No differences in performance were observed between target colors.

Two important changes were made to adapt this task for children. First, the presentation time for arrays was increased to better fit the speed at which children can process numerosity. Chi and Klahr (1975) and Svenson and Sjöberg (1983) demonstrated that 8-year-old children subitize at approximately 75 msec/item, and count at approximately 600 msec/item, while adults subitize at 50 msec/item and count at 300 msec/item. Since the children participating in this study were 9 and 10 years old, the length of stimulus presentation was scaled to fit the subitizing time for 8-year-olds.

Second, in the Ester et al. (2012) study, adult participants responded by pressing number buttons on a standard keyboard. This was deemed too risky for children, who would undoubtedly have to look down after each trial to find the right button, creating muscle artifacts that impact data quality. Instead, this N2pc task used verification responses in which children responded using only two buttons, indicating whether a symbolic number presented on the screen matched the quantity of target squares. To verify that this change did not impact the N2pc component, a pilot test was run with five adults. Results from this pilot test demonstrated the set size modulation described in Ester et al., so this verification task was used for Experiment 3. Digits were correct for half of the trials for each set size, and were incorrect (one away) for the other half.

The N2pc task was run using E-Prime 2.0 software (Psychological Software Tools, Pittsburgh, PA). During the N2pc task, children viewed a black fixation dot in the center of a gray screen for 400-600 ms (jittered), followed by an arrow cue for 300 ms, indicating the side of the fixation on which the target color squares (green or blue) would appear. After another 300 ms fixation, an array of squares appeared for 120 ms. Children then saw another 300 ms fixation, followed by a digit that remained on the screen until a response was made. Children held a button pad with two buttons labeled “Yes” and “No,” and used these buttons to indicate whether the digit matched the number of green or blue squares in the array. The order of response buttons (which was on the left/right) was also counterbalanced across subjects, resulting in four counterbalancing conditions. No differences were observed in response accuracy based on order of response buttons.

When a response was made, the digit disappeared. Trials were separated by an 800-1000 ms (jittered) inter-trial interval. Figure 2 in the Introduction shows a sample trial from Ester et al.; these stimuli were replicated as closely as possible in Experiment 3.

The number of target and non-target squares ranged from 0 to 6, with set sizes 0, 1, 5, and 6 as boundary trials. Set sizes 2, 3, and 4 were of particular interest, because subitizing capacity and N2pc asymptote should both be in this range. Therefore, the task consisted of 10 trials each of set sizes 0, 1, 5, and 6, and 200 trials each of set sizes 2, 3, and 4, for a total of 640 trials that were presented randomly. The task was separated into six blocks. The goal in focusing on set sizes 2-4 and having very few trials for boundary set sizes was to make the task shorter, and thus more suitable for children. Importantly, each block also contained 3 "catch trials." These trials were intended to identify when participants were not attending to the target side. Catch trials were implemented in set sizes 1-3, for which children's accuracy was expected to be very high. In these trials, the digit matched the non-target quantity, which was different from the target quantity. If participants responded with "Yes," this indicated that they may have been attending to an arbitrary side of the array, which would have implications for the laterality of neural responses. When catch trials indicated that children were not attending to the correct side, the experimenter paused the task and reminded the child to attend to the cued side.

Aside from catch trials, no specific trial-by-trial feedback was given regarding accuracy. Children were visited by the experimenter between blocks and given positive feedback on their performance. The experimenter also reminded participants to keep their eyes on the fixation, and not to look at the response pad when pushing a button, so as to minimize eye movements. Difficulty with fixation was an expected obstacle, so during each block, the experimenter used a video monitor to observe whether children were keeping their eyes on the fixation. If a child was looking around, shifting their eyes, or blinking excessively, the experimenter paused the task and went into the testing room to remind the child to keep their eyes as still as possible.

Of the 27 children who participated in the study, 6 ended early because of discomfort related to the EEG testing. Reasons included, but were not limited to, itchy skin (a common side effect of the solution used to help electrical conductance), headache (possibly from the electrode net putting pressure on the head), and nervousness about the EEG net. In addition, one participant did not begin EEG testing because her hair was extremely thick, and impedances would not go below threshold; this would not have yielded usable data. Thus, 20 children completed all six blocks of the task. This rate of attrition due to non-performance-related factors is not atypical of EEG studies with developmental populations.

EEG data were recorded using a 128-Channel Hydrocel Geodesic Sensor Net (Electrical Geodesics Inc., Eugene, Oregon) referenced to Cz. Sampling occurred at a rate of 250 samples/second. Impedance thresholds were set at 40 μV and no more than five non-consecutive impeded electrodes were allowed. Impedances were checked before data collection began, and after blocks 2 and 4. Raw data were filtered with a highpass filter of 0.3 Hz and a lowpass filter of 30 Hz. Trials contaminated by artifacts (blinks or large eye movements) were excluded from analyses, as were trials that had incorrect responses (Pagano et al., 2013; Pagano & Mazza, 2012). After exclusion of individual trials, participants who had less than 50% of trial remaining for set sizes 2-4 were excluded from N2pc analyses.

Unfortunately, 10 of the remaining participants had data that was too contaminated by artifacts to be usable. Thus, the final sample size of participants who had usable EEG data was 10 children (4 third graders and 6 fourth graders). Possible reasons for this high attrition rate are outlined in Experiment 3 Discussion.

Calculation of N2pc component

A different EEG net was used for this study than for previous studies. Consequently, electrode regions of interest were identified by overlaying a map of the 128-channel net over a map of the net used in previous studies. Channels were chosen that overlapped with the channels used by Ester et al. (2012) and Pagano and Mazza (2012). Figure 10 shows the channels that were selected for analysis (marked in red).

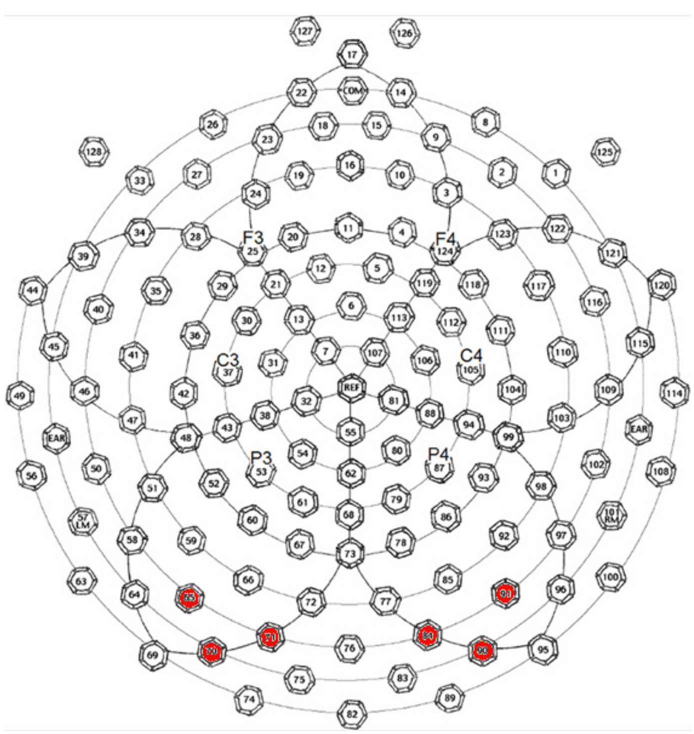

Figure 10. Channels selected for N2pc calculation (front of head is on top; selected channels are marked in red).

For each channel, I extracted the waveform across 200-300 ms after stimulus onset (the time period from which the N2pc should be calculated). Waveforms were averaged across the three electrodes in each region, resulting in a right hemisphere waveform and a left hemisphere waveform for each set size by target side condition (i.e., set size four with the target on the left). For each set size, contralateral waveforms were collapsed across the right hemisphere during left visual field target trials, and left hemisphere during right visual field target trials. Ipsilateral waveforms were collapsed across the left hemisphere during left visual field target trials, and the right hemisphere during right visual field trials. Mean amplitudes were calculated across contralateral and ipsilateral waveforms. Finally, the N2pc component for each set size was calculated by subtracting ipsilateral mean amplitude from contralateral mean amplitude. This method of calculating the N2pc component has been used in several previous studies (Ester et al., 2012; Mazza, Turatto, & Caramazza, 2009; Pagano & Mazza, 2012; Woodman, Kang, Rossi, & Schall, 2007).

Visual working memory (VWM) task

The VWM task was a computer-based change detection task. Children were presented with a fixation, followed by a 150 millisecond (ms) array of 1, 2, 4, or 8 colored squares around the fixation, followed by a brief (1000 ms) retention interval, and then an array with a single square in a location matching a square from the original array. Children had unlimited time to respond by pushing a “Same” or “Different” button on the keyboard to indicate whether the color of the square had changed. Participants were instructed to attend to the squares without moving their eyes away from the fixation.

Each set size was presented 24 times for a total of 96 trials, which took approximately 6 minutes for children to complete. Correct answers were evenly split between "Same" and "Different," with 12 "Same" trials and 12 "Different" trials per set size. Participants were not cued to which square from the original array would be tested. Square colors were randomly selected from eight highly discriminable color options, with replacement, such that two squares in an array could be the same color. Therefore, participants had to encode the color of each individual square, rather than simply remembering which colors were present in an array (Riggs, McTaggart, Simpson, & Freeman, 2006).

Hypotheses

First, I predicted that measures of subitizing capacity derived from behavioral performance on the N2pc task would correspond to FN Unstructured estimates of subitizing capacity. This result would support the validity of the method used to classify 4-Counters and 4-Subitizers in Experiment 2; alternatively, a different result could identify issues with these classifications.

Second, based on compelling evidence for the fixed-capacity model of groupitizing and its underlying neural components, there was no reason to hypothesize that the neural substrates of subitizing for children would be different from adults. Therefore, I predicted that, if the feasibility study succeeded, children's N2pc amplitudes would be modulated by set size, reaching an asymptote around a quantity that corresponds to subitizing capacity.

Third, I anticipated that behavioral data from the VWM task would demonstrate very high accuracy for set sizes 1 and 2, and that accuracy for set size 8 would be no greater than chance level. These predictions are based on previous studies suggesting that 1 and 2 items fall within visual working memory capacity for adults and children, but 8 is above capacity (Riggs et al., 2006; Vogel, Woodman, & Luck, 2001). Furthermore, I predicted that some participants would also perform very well for set size 4, while others would have a marked decrease in accuracy for set size 4. Children who have high accuracy for 4 squares may correspond to the 4-Subitizers on the FN Unstructured task, and children who have significantly lower accuracy may correspond to the 4-Counters. This finding would provide additional support for the fixed-capacity model of subitizing by demonstrating that estimates of subitizing range map onto visual working memory capacity.

Results

Behavioral measures of subitizing capacity

Children were re-classified as 4-Counters and 4-Subitizers based on their concurrent data from FN Unstructured, using the method described in Experiment 2. Of the 20 children who completed all six blocks of the N2pc task, 12 were 4-Counters, and 8 were 4-Subitizers. Subitizing capacity was also calculated from accuracy on the N2pc task. Participants performed very well on this task. Because participants had limited time to respond in the N2pc task, subitizing capacity could be estimated from accuracy (a method that could not be used for FN Unstructured because of unlimited response time, which enabled near-perfect accuracy for all set sizes).

Previous studies (Franconeri, Alvarez, & Enns, 2007; Pagano & Mazza, 2012) have used the criterion that accuracy below 90% for a given set size indicated counting. Again, the 20 children who completed all six blocks of the N2pc task were classified as 4-Counters and 4-Subitizers based on this criterion. Children who performed below 90% accuracy for set size 4 were considered 4-Counters, and children who performed at or above 90% were considered 4-Subitizers. Of these 20 children, 13 were classified as 4-Counters and 7 were classified as 4-Subitizers. Reaction time-based classifications from FN Unstructured and accuracy-based classifications from the N2pc task were compared in Table 8.

Table 8. Classifications of 4-Counters and 4-Subitizers based on FN Unstructured enumeration latencies and N2pc task accuracy.

Group Classification (Criterion)	4-Counter (N2pc)	4-Subitizer (N2pc)
4-Counter (FN Unstructured)	11	1
4-Subitizer (FN Unstructured)	2	6

As this table demonstrates, only 3 participants were classified differently by the two models, indicating that there is significant overlap between them. Although this sample is relatively small, this analysis supports the use of reaction time-based methods of classification in Experiment 2.

N2pc results

As a first step, averaged waveforms from stimulus onset for each hemisphere were visually inspected for each of the 10 participants. Sample waveforms are shown in Figure 11. The purpose of this visual inspection was twofold. First, it is important to check that pre-stimulus onset baseline amplitudes are close to zero. Second, I wanted to verify that early-stage components that are characteristic of visual tasks were present in the waveforms. Specifically, I looked for a P1 response (a positive amplitude around 100 ms) and an N1 response (a negative amplitude around 100 ms), both of which reflect that a visual stimulus is being processed. When a participant is doing a task with a visual component, these amplitudes should be present in the waveforms. Visual inspection verified that these components were present for each of the 10 participants in both hemispheres.

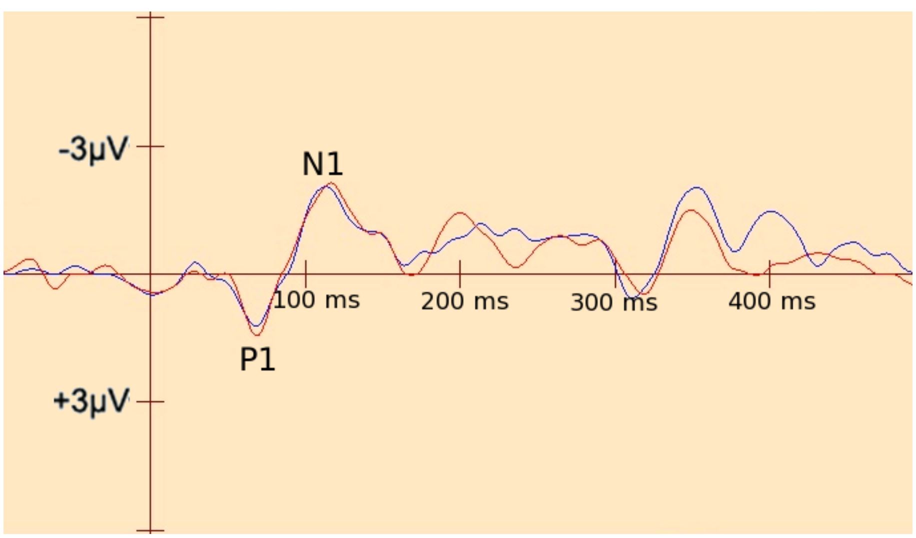

Figure 11. Sample averaged waveforms from the left hemisphere for set size 2. Blue lines indicate targets presented to the left visual field, and red lines indicate targets presented to the right visual field. P1 and N1 components are labeled. (*Note:* As is the standard for EEG waveforms, negative amplitudes are plotted upward, and positive amplitudes are plotted downward.

Next, for each participant, N2pc data were plotted at each set size for a preliminary visual assessment of whether amplitudes approximated the predicted result. Surprisingly, N2pc plots appeared very different across participants, and no single participant showed a pattern of N2pc amplitude increasing with set size and reaching a plateau around set size 3. For three participants, N2pc amplitude decreased, for five participants, amplitude crossed the x-axis from positive to negative or vice versa, and for the remaining two participants, N2pc showed no discernible change in amplitude. To check that N2pc did not follow any reliable pattern, a repeated-measures ANOVA was calculated for mean amplitudes, using two within-subjects factors of Set Size (2, 3, and 4) and Hemisphere (Left and Right). No main effects or interactions were significant, indicating that the N2pc measures derived from these 10 participants did not fit the predicted effect as seen in adults.

VWM results

Participants demonstrated very low accuracy for all set sizes in the VWM task. Chance-level accuracy was 50%, because there were two answer options for each trial. Mean accuracy for each set size across all participants was not far enough above chance to suggest that they were able to reliably represent arrays in working memory (50.1% for 1 square, 54.5% for 2 squares, 58.0% for 4 squares, and 63.9% for 8 squares).

Even more surprisingly, accuracy actually increased slightly as set size increased. However, reaction times also increased with set size, so this may have been due to a speed-accuracy tradeoff. Participants may have taken more time to respond during more difficult trials (i.e., trials with 4 or 8 squares, exceeding visual working memory capacity), enabling them to get a few more trials correct, while they responded quickly and without much thought for seemingly easier trials.

To explore whether different types of trials affected performance, “change trials” (in which the color of the test square was different from the initial array) were compared to “no-change trials” (in which the color of the test square was the same as in the initial array). Interestingly, participants demonstrated overall high accuracy for change trials (mean of 87.9% across set sizes) and low accuracy for no-change trials (mean of 25.4% across set sizes). Furthermore, for change trials, participants’ mean accuracy did show the expected decrease with increases in set size; however, for no-change trials, participants’ mean accuracy increased with set size.

“Same” and “Different” responses were also compared, revealing that all participants responded with “Different” far more often than “Same” despite having been instructed that some squares would not change color. The mean number of “Different” responses was 78 out of 96 total trials, whereas the mean number of “Same” responses was 18. This indicates a strong bias toward guessing “Different” when unsure of the correct answer. Thus, when participants had to guess on a change trial, they were very likely to get the answer correct, and when they had to guess on a no-change trial, they were very unlikely to guess correctly.

This response bias accounts for the substantial difference in accuracy between change and no-change trials. However, this does not account for the increase in accuracy at higher set sizes. Overall, it appears that the VWM task was unexpectedly difficult for participants for reasons that remain unclear. VWM has been successfully measured in children before, using a similar task (Riggs et al., 2006). One possibility is that temporal parameters for the VWM task need to be adjusted to better suit children's slower number processing speed. This could account for why children performed quite well on the N2pc task, but not on the VWM task.

Discussion

The first analysis in Experiment 3 focused on the relationship between subitizing capacity group classifications based on FN Unstructured enumeration latencies and N2pc task accuracy. Comparisons showed that classifications based on reaction times from a simple enumeration task, calculated using the method described in Experiment 2, map onto classifications based on performance on a task that is much more cognitively demanding. According to the fixed-capacity model of subitizing, accuracy on an N2pc enumeration task should reflect subitizing range. Therefore, this finding supports the validity of the reaction time-based method used to classify children into subitizing span groups in Experiment 2.

Although deriving subitizing capacity measures from the N2pc task was successful, calculating an N2pc at each set size for each subject did not yield the expected pattern of increases in amplitude with set size and an asymptote corresponding to subitizing capacity. Rather, change in N2pc across set sizes was unpredictable within and between subjects. High accuracy demonstrates that this was not due to participants' inability to process and represent the target squares in working memory.

One anticipated obstacle to the feasibility of the N2pc experiment was that children have difficulty keeping their eyes on a fixation point, because this requires inhibiting the urge to shift their eyes toward the target stimulus to which they are attending. However, for an N2pc component to be elicited, fixation in the center of the display must be maintained so that contralateral and ipsilateral measures are a true reflection of the side of the target visual field. Failure to derive a meaningful N2pc measure despite the presence of other components characteristic of visual attention may, therefore, partly reflect children's difficulty in maintaining fixation. Participants and individual trials were excluded from analysis on the basis of eye movements, but small shifts in eye fixation may not have exceeded the threshold implemented for rejecting trials. Trials with slight eye movements may have been retained in the data used in analysis, contaminating the N2pc.

It is also a distinct possibility that children do not elicit an N2pc response because of cognitive demands of the task that are unrelated to subitizing skill. In the N2pc task, target squares must be individuated, or perceptually separated, from distractors before they can be subitized (Ester et al., 2012). Previous research has shown that children have difficulty tracking three or four objects when they are presented among distractors (Trick, Jaspers-Fayer, & Sethi, 2005). This suggests that individuating target items from distractors is more difficult for children than for adults. Therefore, even though the stimulus presentation time was adjusted to reflect differences between adults' and children's numerosity processing speed, children may need more time to successfully individuate target items. Children were clearly able to determine the number of target items, as demonstrated by their high accuracy, but they may have used a strategy other than subitizing that did not elicit an N2pc (such as counting from a mental representation of the array, or estimating quantity from the area of color they perceived). This could account for the seemingly counterintuitive finding of high response accuracy but no predictable pattern for the N2pc component.

The VWM task, participants' unexpected low accuracy across all set sizes likely reflects a procedural difficulty with the task itself, rather than the cognitive limits of visual working memory. Again, high accuracy in the EEG task demonstrate that these children clearly are able to encode a quickly-presented array of items in the subitizing range. Previous studies have successfully shown that children can successfully and accurately detect changes in a display when the display contains 1, 2, 3, and sometimes 4 items. The present study did not replicate this result. Furthermore, comparisons of accuracy across different types of trials revealed that children were relying on unhelpful strategies (e.g., always pushing the "Different" button).

VWM task did vary slightly from traditional change detection tasks (e.g., Fukuda, Vogel, & Awh, 2010; Luck & Vogel, 1997; Riggs et al., 2006; Vogel, Woodman, & Luck, 2001, Experiments 1 and 2) in which participants are shown an array of colored squares, followed by a retention interval, then another array of the same quantity, and asked to indicate whether any of the squares in the array had changed color. In the VWM task used for the present experiment, participants responded to one test square, instead of matching a test array to their mental representation of the original array. It is possible that the VWM task in the present study was more cognitively demanding, as participants had to single out one particular square from their mental representation of an array and compare the color. Future studies should make use of traditional change detection tasks for measuring VWM. If accuracy is significantly greater, procedural aspects of the present task may account for why children had so much difficulty.

CHAPTER V

DISCUSSION AND INTEGRATION OF FINDINGS

Summary of findings

Although the effect of perceptual grouping on enumeration of quantities beyond the subitizing range has been established for a long time, the field of cognitive psychology has yet to identify a well-substantiated mechanism for this. This is an important task, given recent findings showing that groupitizing ability at an early age is a significant predictor of later math achievement. Starkey and McCandliss (2014) proposed a mechanism for groupitizing that involves two steps. First, subgroups in an array are subitized. Then, the combination of subgroup quantities enables direct access to array numerosity for individuals whose concept of number has developed to include the possible combinations of subsets that make up that number. It follows that fluent groupitizing, in which increases in set size do not impact enumeration latency, reflects an individual's advanced understanding of the composition of number, and ability to apply this concept to enumeration.

Situating groupitizing within the context of Dehaene's well-substantiated triple-code model of number processing (Dehaene, 1992; Dehaene & Cohen, 1995; 1997; Schmidthorst & Brown, 2004) may shed light on why some children are better at groupitizing than others. Dehaene (1992) proposed that every number is mentally represented in three formats, or *codes*. These include an *auditory verbal word frame* for spoken number word input and output, a *visual Arabic number form* for visually presented symbolic digits, and an *analog magnitude code* for representing quantity. When an array is subitized, the number of items is represented in an analog magnitude code; counting, however, is a verbally-mediated task, so when arrays are counted, the number of items is represented as a verbal word frame.

Importantly, when numbers are operated upon (e.g., in the context of addition or combination of sets), number representations can be transcoded into the preferred code for the operation (Dehaene, 1992). For example, verbal enumeration requires the pairing of a number word with a visually presented non-symbolic quantity. Therefore, the analog magnitude code for the presented quantity must be transcoded into a verbal word frame, and then produced through speech. The preferred code for arithmetic tasks is verbal word frames, so adding sets of items requires that each number representation be transcoded from an analog magnitude into a verbal word frame before this operation can be performed (Dehaene, 1992; Dehaene & Mehler, 1992).

In the first step of the proposed mechanism of groupitizing, subgroups are subitized, and would therefore be represented in the analog magnitude code. Starkey and McCandliss proposed that the second step of groupitizing involves connecting the subgroup quantities to the overall array size using a non-serial process that is not affected by increases in set size. One potential mechanism that would enable this access is set combination, in which the quantities of the subgroups are combined, producing an effect of number of subgroups, but not of overall set size. The triple-code model would suggest that part of this process involves transcoding subgroup quantities into the verbal word frame. In contrast, unstructured arrays must be counted, and are therefore initially represented in the verbal word frame. It is still unclear which domain is responsible for the process of connecting subgroup quantities to overall set size for grouped arrays. One possibility is that the process of set combination is verbally mediated, occurring in the verbal word frame code. However, it is also possible that the sets are combined in the analog magnitude code and then transcoded into a verbal word frame.

A critical part of what characterizes groupitizing is the reaction time benefit compared to counting, which does not involve transcoding. Thus, in the context of the triple-code model, fluent groupitizing must in part rely on the successful integration of, and fluent transcoding between, the three codes. The development of groupitizing fluency over elementary school, suggested by Starkey and McCandliss (2014), may reflect improvements in integration as children gain more experience with numbers, and thus become more adept at transcoding their mental representations of numbers. Additionally, individual differences in groupitizing ability could reflect variability in children's code integration and fluency in transcoding number representations.

The present study examined the development of groupitizing in a longitudinal sample of elementary school aged children, and further explored the relationship between change in groupitizing fluency and improvements in symbolic arithmetic skills. Results demonstrate that groupitizing fluency improves in a linear trajectory across elementary school, potentially reflecting greater integration of various mental representations of numbers. Changes in counting fluency predict some variance in individuals' rate of groupitizing development. Furthermore, developments in groupitizing fluency, but *not* developments in subitizing or counting fluency, significantly predict individual differences in how quickly symbolic math skills improve over elementary school. This important finding confirms the strong relationship between groupitizing ability and higher-level math skills like arithmetic. Moreover, this result enhances, previous findings, suggesting that developing fluent groupitizing skills reflects conceptual number knowledge that is fundamental to children's math achievement.

The idea that math skills improve as one's concept of number becomes increasingly enriched is not new in the field of number cognition. In particular, Dehaene (1992) and Dehaene and Cohen (1995; 1997) proposed that simple arithmetic facts (e.g., two-operand addition with small numbers) are stored as verbally-mediated associations. More complex arithmetic, on the other hand (e.g., three-operand addition, or two-operand addition with larger numbers) requires some level of "semantic elaboration," meaning that the mental representations of the numbers involved must carry semantic characteristics that aid in problem solving (Dehaene & Cohen, 1995).

For example, to solve 11+5, it is helpful for children to know that 11 is equal to 10+1; then, the statement can be rearranged to 10+6, which is easier to solve. This knowledge that numbers can be decomposed enables children to use alternative strategies in solving arithmetic problems (Siegler, 1988). Therefore, greater semantic elaboration (i.e., more semantically rich mental representations of numbers) facilitates complex arithmetic. For small enough numbers (such as the range of quantities employed in the enumeration tasks), adults have a "rich and precise internal semantic representation" (Dehaene & Cohen, 1995, p. 87), which contributes to their ability to fluently perform complex arithmetic. For elementary school children, developing these rich semantic representations may contribute to improvements in arithmetic ability.

Dehaene's notion of semantic elaboration captures the developments in concept of number that Starkey and McCandliss (2014) propose also influences groupitizing ability. Significant improvements in groupitizing fluency occur during the grades in which children are increasingly exposed to math activities that encourage insights into the semantics of number (such as combining sets of manipulatives, decomposing sets into subgroups, and later, simulating multiplication with manipulatives). Improvements in children's groupitizing fluency, as demonstrated in Experiment 1 of this study, may reflect the benefits of developments in their semantic number knowledge, particularly pertaining to number composition. Specifically, children whose representations of number carry more semantic information may be able to more successfully make associations between subset quantities and overall array size by accessing these number semantics, leading to more fluent groupitizing. Further, the contribution of counting improvements to groupitizing fluency suggests that counting may scaffold either developments in semantic elaboration, or the application of number semantics to enumeration of grouped arrays.

The second part of this study delved into the mechanism underlying groupitizing. By measuring enumeration latencies and slopes for grouped arrays with systematically manipulated parameters, this experiment examined the role of subitizing and insights into number composition in groupitizing speed and fluency. Results further suggest that subitizing is the mechanism by which children access the quantity of each subgroup, and revealed that the overall magnitudes of two-subgroup arrays are accessed more easily than three-subgroup arrays. Importantly, Experiment 2 also focused on comparisons between children who can and cannot subitize four, demonstrating that 4-Subitizers have greater groupitizing fluency that is not affected by manipulations to the number and magnitude of subgroups. Children who still have to count four, in contrast, become less fluent at enumerating grouped arrays when these parameters change. Fluency in groupitizing is clearly linked to developments in conceptual number knowledge that enable this access to the overall quantity of an array.

The triple-code model offers some potential explanations for the differences in performance between 4-Counters and 4-Subitizers. It is possible that, as kids become able to subitize four, the connections between the analog magnitude code and verbal word frame grow stronger, and that transcoding between these formats happens more quickly. This would increase the fluency with which 4-Subitizers can associate subset combinations that include a four with the overall array size.

Alternatively, the ability to subitize four may reflect, or produce, a richer semantic understanding of four (in addition to the perceptual capabilities that enable subitizing of four items). This could include information about the various number combinations that include four that comprise other, larger numbers. This enriched understanding of four could perhaps facilitate associations between subset quantities and overall array size for SubMax 4 grouped arrays. In contrast, 4-Counters, whose concept of four does not include this information, do not benefit from this facilitation.

Finally, examining children's behavioral performance in an N2pc enumeration task, in which target items to be enumerated are presented among distractors, corresponds to subitizing capacity as defined by response latencies in a simple enumeration task. This provides some preliminary evidence for the validity of the method used to identify 4-Counters and 4-Subitizers, strengthening conclusions from Experiment 2 regarding group differences in groupitizing speed and fluency. Furthermore, this result provided additional evidence for a fixed-capacity model of subitizing, in which subitizing range is governed by limitations on attention and working memory.

Taken together, the experiments in this study create a clearer picture of the cognitive processes that groupitizing reflects, and why developments in groupitizing are important to the larger picture of children's math achievement. The feasibility component of this study provides crucial feedback for conducting successful future investigations of the neural substrates of subitizing and, by extension, groupitizing.

Limitations and suggestions for future research

Experiment 1 demonstrated important relationships between developing variables. However, this study focused on three types of enumeration skills; there are many other early number skills that may contribute to the development of groupitizing. For example, results indicating that set combining skills may be involved in groupitizing suggest that this skill should be directly tested in future studies as a likely predictor of groupitizing development. In addition, causality cannot yet be established based on results from Experiment 1. Future studies should assess the effectiveness of an intervention for groupitizing skills, insights about number composition, and stored knowledge of set combinations that constitute larger numbers. This would justify claims about the direction of causality between conceptual number knowledge, developments in groupitizing ability, and changes in symbolic arithmetic ability.

Experiment 1 suffered from significant attrition over time, resulting in missing data. Attrition is common in longitudinal studies, and using hierarchical linear modeling helped to combat this limitation. Future longitudinal studies should recruit a larger subject pool from the same grade, and follow these participants at evenly-spaced intervals. This would enable flexibility in choosing the best approach for accurate model building. A study with a larger sample and more consistent follow-up would also have greater statistical power, and would allow for stronger conclusions about the potential implications of null effects. The goal of the present study was not to evaluate a variety of exploratory models for the best fit, but to test specific hypotheses about predictors of groupitizing fluency and symbolic arithmetic. In the future, because groupitizing is still not as well understood as other enumeration skills, it would be interesting and useful to build and compare models that include additional predictors of number skill (such as developments in approximate number skills).

Experiment 2 would also benefit from a larger sample size that would enable stronger conclusions about null effects (such as the absence of a significant difference between two slopes). Furthermore, future studies that build upon this experiment should aim to recruit equal numbers of 4-Counters and 4-Subitizers, and possibly even match these participants on other factors that influence academic achievement (such as SES and parental education).

Experiment 3 yielded some interesting results, but importantly, serves as an important lesson for future research. Clearly, additional adaptations need to be made in order to create a child-friendly N2pc task. First, this could be turned into an engaging video-game-like task to prevent children from quitting the task early, and to motivate them to follow rules like maintaining eye fixation. Second, more monitoring of eye movements would be helpful in ensuring high-quality data and consistent fixation. One possibility is to use an eye tracker in conjunction with the EEG system. This would allow the experimenter to give instant feedback or mark trials with eye movements, and could potentially build feedback into the video game task based on these eye movements (e.g., across the task, participants accumulate points for a later reward, and lose points every time they move their eyes toward a target).

In addition to these suggested changes, completing a training module before starting a task may be helpful. This would help to ensure that children are able to achieve an expected level of accuracy before proceeding with the task, and would also provide practice in maintaining eye fixation. Training modules could also allow an experimenter to preemptively assess the level and type of feedback that a child might need, which would help the session run more smoothly and possibly yield higher-quality data.

Concluding comments

In conclusion, this study provides new evidence that highlights the importance of groupitizing as a fundamental enumeration skill that reflects critical developments in conceptual number knowledge. Moreover, results of this study help to refine our understanding of what this conceptual knowledge may entail. Findings confirm that subitizing is recruited in groupitizing, and also further suggest that an understanding of numbers as being composed of subsets is critical to the development of children's groupitizing skills. The ability to quickly combine sets of quantities, and to fluently switch between representations of a number and the combinations of sets that compose it, may be essential in establishing a direct connection between subitizing subgroups, and enumerating a whole array.

The results of this study carry implications for math instruction in school, suggesting that early insights into number composition have a significant impact on enriching concept of number and building higher-level math skills. Emphasizing training in number composition and set combination may produce improvements in groupitizing, resulting in stronger skills in symbolic arithmetic and beyond. However, there are elements of groupitizing that are still not completely clear. Therefore, understanding groupitizing should be a focus of future studies working toward the goal of improving early math education.

REFERENCES

Anderson, D.E., Vogel, E.K., & Awh, E. (2011). Prevision in Visual Working Memory Reaches a Stable Plateau When Individual Item Limits Are Exceeded. *The Journal of Neuroscience 31*, 3, 1128-1138.

Ansari, D., Lyons, I.M., van Eimeren, L., & Xu, F. (2007). Linking Visual Attention and Number Processing in the Brain: The Role of the Temporo-Parietal Junction in Small and Large Symbolic and Nonsymbolic Number Comparison. *Journal of Cognitive Neuroscience, 19,* 11, 1845-1853.

Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences, 20*, 427-435.

Balakrishnan, J.D., & Ashby, F.G. (1992). Subitizing: Magical numbers or mere superstition? *Psychological Research 54*, 80-90.

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

Chi, M.T.H., & Klahr, D. (1975). Span and Rate of Apprehension in Children and Adults. *Journal of Experimental Child Psychology, 19*, 434-349.

Clements, D.H. (1999). Subitizing: What Is It? Why Teach It? *Teaching Children Mathematics, 5*, 7, 400-405.

Cowan, N., Elliott, E.M., Saults, J.S., Morey, C.C., Mattox, S., Hismjatullina, A., & Conway, A.R.A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51,* 42-100.

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

Dehaene, S., & Cohen, L. (1994). Dissociable Mechanisms of Subitizing and Counting: Neuropsychological Evidence from Simultanagnosic Patients. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 5, 958-975.

Dehaene, S., & Cohen, L. (1995). Toward an Anatomical and Functional Model of Number Processing. *Mathematical Cognition, 1*, 1*,* 83-120.

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

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition, 43,* 1-29.

Ding, C.S., & Davidson, M.L. (2005). A longitudinal study of math achievement gains for initially low achieving students. *Contemporary Educational Psychology, 20,* 81-95

Drew, T., & Vogel, E.K. (2008). Neural Measures of Individual Differences in Selecting and Tracking Multiple Moving Objects. *The Journal of Neuroscience 28*, 16, 4183-4191.

Duckworth, A., Tsukamaya, E., & May, H. (2010). Establishing Causality Using Longitudinal Hierarchical Linear Modeling: An Illustration Predicting Achievement From Self-Control. *Social Psychology and Personality Science, 1,* 4, 311-317.

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P, Pagani, L., Feinstein L., Engel, M., Brooks-Gunn, J., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology 43*, 6, 1428-1446.

Ester, E.F., Drew, T., Klee, D., Vogel, E.K., & Awh, E. (2012). Neural measures reveal a fixed item limit in subitizing. *Journal of Neuroscience, 32*, 1, 7169-7177.

Feigenson, L. (2008). Parallel non-verbal enumeration is constrained by a set-based limit. *Cognition, 107,* 1-18.

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

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

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences, 105*, 29, 9926-9930.

Fischer, B., Gebhardt, C., & Hartnegg, K. (2008). Subitizing and Visual Counting in Children with Problems in Acquiring Basic Arithmetic Skills. *Optometry & Vision Development, 39*, 1, 24-29.

Franconeri, S.L., Alvarez, G.A., & Enns, J.T. (2007). How Man y Locations Can Be Selected At Once? *Journal of Experiment Psychology: Human Perception and Performance, 33,* 5, 1003-1012.

Freeman, F.N. (1912). Grouped Objects as a Concrete Basis for the Number Idea. *The Elementary School Teacher 12*, 7, 306-314.

Fuchs, L.S., Fuchs, D., Compton, D.L., Powell, S.R., Seethaler, P.M., Capizzi, A.M., Schatschneider, C., & Fletcher, J.M. (2006). The cognitive correlates of third-grade skill in mathematics arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98,* 29-43.

Fuchs, L.S., Geary, D.C., Compton, D.L., Fuchs, D., Hamlett, C.L., & Bryant, J.D. (2010). The contributions of numerosity and domain-general abilities to school readiness. *Child Development, 81*, 1520-1533.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review 17*, 5, 673-679.

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

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

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

Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*, 1343-1359.

Gebuis, T., & Reynvoet, B. (2011). The Interplay Between Nonsymbolic Number and Its Continuous Visual Properties. *Journal of Experimental Psychology: General, 141,* 4, 642-648.

Gerhardstein, P., & Rovee-Collier, C. (2002). The development of visual search in infants and very young children. *Journal of Experimental Child Psychology, 81*, 2, 194-215.

Halberda, J., Sires, S.F., & Feigenson, L. (2006). Multiple Spatially Overlapping Sets Can Be Enumerated in Parallel. *Psychological Science 17*, 7, 572-576.

Hanushek, E.A., Peterson, P.E., & Woessman, L. (2012). Achievement Growth: International and U.S. State Trends in Student Performance. *PEPG Report No. 12-03.*

Hyde, D.C., & Spelke, E.S. (2011). Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. *Developmental Science 14*, 2, 360-371.

Jevons, W.S. (1871). The power of numerical discrimination. *Nature, 3*, 281-282.

Kaufman, E.L., Lord, M.W., Reese, T.W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology, 62*, 4, 498-525.

Klahr, D., & Wallace, J.G. (1973). The role of quantification operators in the development of conservation of quantity. *Cognitive Psychology, 4,* 3, 301-327.

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

Lassaline, M.E., & Logan, G.D. (1993). Memory-based automaticity in the discrimination of visual numerosity. *Journal of Experimental Psychology, 19,* 561-581.

Logan, G.D., & Zbrodoff, N.J. (2003). Subitizing and similarity: Toward a pattern-matching theory of enumeration. *Psychonomic Bulletin & Review 10*, 3, 676-682.

Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279-281.

Mandler, G., & Shebo, B.J. (1982). Subitizing: An Analysis of Its Component Processes. *Journal of Experimental Psychology: General 111*, 1, 1-22.

Mazza, V., Turatto, M., & Caramazza, A. (2009). Attention selection, distractor suppression and N2pc. *Cortex, 45*, 879-890.

Mazza, V., & Caramazza A. (2011). Temporal Brain Dynamics of Multiple Object Processing: The Flexibility of Individuation. *PloS One 6,* 2, e17453.

Mullis, I.V.S., Martin, M.O., Foy, P, & Arora, A. (2012). TIMSS 2011 International Results in Mathematics. Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement.

O'Connell, A.A., & McCoach, D.B. (Eds.). (2008). *Multilevel Modeling of Educational Data.* Charlotte, NC: Information Age Publishing, Inc.

Pagano, S., Lombardi, L., & Mazza, V. (2014). Brain dynamics of attention and working memory engagement in subitizing. *Brain Research, 1543,* 244-252.

Pagano, S., & Mazza, V. (2012). Individuation of multiple targets during visual enumeration: New insights from electrophysiology. *Neuropsychologia, 50,* 754-761.

Piazza, N., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition 121*, 147-153.

Piazza, M., Giacomini, E., Le Bihan, D., & Dehaene, S. (2003). Single-trial classification of pre-attentive and serial attentive processes using functional magnetic resonance imaging. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270,* 1237-1245.

Pylyshyn, Z.W., & Storm, R.W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision 3*, 3, 179-197.

Railo, H., Koivisto, M., Revonsuo, A., & Hannula, M.M. (2008). The role of attention in subitizing. *Cognition 107*, 82-104.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods (Second Edition).* Thousand Oaks, CA: Sage Publications, Inc.

Raudenbush, S.W., Bryk, A.S, & Congdon, R. (2011). HLM 7 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.

Raudenbush, S.W., Bryk, A.S, Cheong, Y.F., Congdon, R., & du Toit, M. (2011). *HLM 7: Hierarchical Linear & Nonlinear Modeling.* Lincolnwood, IL: Scientific Software International, Inc.

Riggs, K.J., McTaggart, J., Simpson A., & Freeman, R.P.J. (2006). Changes in the capacity of visual working memory in 5- to 10-year-olds. *Journal of Experimental Child Psychology, 95, 18-26.*

Schmithorst, V.J., & Brown, R.D. (2004). Empirical validation of the triple-code model of numerical processing for complex math operations using functional MRI and group Independent Component Analysis of the mental addition and subtraction of fractions. *NeuroImage, 22,* 1414-1420.

Schrank, F.A., McGrew, K.S., & Woodcock, R.W. (2001). *Technical Abstract* (Woodcock-Johnson III Assessment Service Bulletin No. 2). Itasca, IL: Riverside Publishing.

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

Siegler, R.S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 3,* 258-275.

Simon, T.J., & Vaishnavi, S. (1996). Subitizing and counting depend on different attentional mechanisms: Evidence from visual enumeration in afterimages. *Perception & Psychophysics, 58,* 6, 915-926.

Singer, J.D., & Willett, J.B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York, NY: Oxford University Press, Inc.

Starkey, G.S., Hubbard, E.M., & McCandliss, B.D. (2013). Linking subitizing fluency with neural systems for exact quantity. Poster presented at the Cognitive Neuroscience Society conference in San Francisco, CA.

Starkey, G.S., & McCandliss, B.D. (2014). The emergence of "groupitizing" in children's numerical cognition. *Journal of Experimental Child Psychology, 126*, 120-137.

Svenson, O., & Sjöberg, K. (1983). Speeds of subitizing and count processes in different age groups. *The Journal of Genetic Psychology, 142*, 203-211.

Trick, L.M., & Pylyshyn, Z.W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101,* 80-102.

Vetter, P., Butterworth, B., & Bahrami, B. (2008). Modulating Attentional Load Affects Numerosity Estimation: Evidence against a Pre-Attentive Subitizing Mechanism. *PLoS One, 3*, 9, e3269.

Vetter, P., Butterworth, B., & Bahrami, B. (2011). A Candidate for the Attentional Bottleneck: Set-size Modulation of Right TPJ during Attentive Enumeration. *Journal of Cognitive Neuroscience, 23*, 3, 728-736.

Vogel, E.K., & Machizawa, M.G. (2004). Neural activity predicts individual differences in working memory capacity. *Nature 428*, 748-751.

Vogel, E.K., Woodman, G.F., & Luck, S.J. (2001). Storage of Features, Conjunctions, and Objects in Visual Working Memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 1, 92-114.

Wender, K.F., & Rothkegel, R. (2000). Subitizing and its subprocesses. *Psychological Research, 64,* 81-92.

Woodman, G.F., Kang, M., Rossi, A.F., & Schall, J.D. (2007). Nonhuman primate event-related potentials indexing covert shifts of attention. *PNAS, 104,* 38, 15111-15116.

Xu, X., & Liu, C. (2008). Can subitizing survive the attentional blink? An ERP study. *Neuroscience Letters 440*, 140-144.