

The small-large divide: The development of infant abilities to discriminate small from large sets

Author: Tasha Irene Posid

Persistent link: <http://hdl.handle.net/2345/bc-ir:104371>

This work is posted on [eScholarship@BC](#),
Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2015

Copyright is held by the author, with all rights reserved, unless otherwise noted.

Boston College
The Graduate School of Arts and Sciences
Department of Psychology

THE SMALL-LARGE DIVIDE:
THE DEVELOPMENT OF INFANT ABILITIES TO DISCRIMINATE SMALL FROM
LARGE SETS

a dissertation

by

TASHA IRENE POSID

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

May 2015

© copyright by TASHA IRENE POSID

2015

The small-large divide: The development of infant abilities to discriminate small from large sets

Tasha Posid

Dissertation Advisor: Sara Cordes

Abstract

Evidence suggests that humans and non-human animals have access to two distinct numerical representation systems: a precise “object-file” system used to visually track small quantities (<4) and an approximate, ratio-dependent analog magnitude system used to represent all natural numbers. Although many studies to date indicate that infants can discriminate exclusively small sets (e.g., 1 vs. 2, 2 vs. 3) or exclusively large sets (4 vs. 8, 8 vs. 16), a robust phenomenon exists whereby they fail to compare sets crossing this small-large boundary (2 vs. 4, 3 vs. 6) despite a seemingly favorable ratio of difference between the two set sizes. Despite these robust failures in infancy (up to 14 months), studies suggest that 3-year old children no longer encounter difficulties comparing small from large sets, yet little work has explored the development of this phenomenon between 14 months and 3 years of age. The present study investigates (1) when in development infants naturally overcome this inability to compare small vs. large sets, as well as (2) what factors may facilitate this ability: namely, perceptual variability and/or numerical language. Results from three cross-sectional studies indicate that infants begin to discriminate between small and large sets as early as 17 months of age. Furthermore, infants seemed to benefit from perceptual variability of the items in the set when making these discriminations. Moreover, although preliminary evidence suggests

that a child's ability to verbally count may correlate with success on these discriminations, simply exposure to numerical language (in the form of adult modeling of labeling the cardinality and counting the set) does not affect performance.

Keywords: Manual Search Task, Object-File; Analog Magnitude; Numerical Cognition; Numerical Discrimination; Perceptual Variability

Table of Contents

Introduction	1
Experiment 1	16
Experiment 2	29
Cross-Experiment Analysis	33
Experiment 3	35
Cross-Experiment Analysis	40
General Discussion	43
References	53
Tables	66
Figures	69

Acknowledgements

I would like to thank my dissertation committee members: my advisor, Sara Cordes, my secondary advisor, Ellen Winner, as well as Hiram Brownell and Elida Laski for their help and continued support on this doctoral dissertation, as well as their invaluable feedback on this dissertation. I would also like to thank my friends and family, who have provided unconditional support during my time at Boston College and during this dissertation process. Specifically, I would like to thank the members of the Infant and Child Cognition Lab, especially Lizzy Bayoff, Danielle Brazel, Lauren Glenn, Alison Goldstein, Emily Kleinlein, Emily Kubota, Emma Lazaroff, Nick Leal, Rosemary O'Connor, Aarati Raghuvanshi, and Beth Sandham. Additionally, I would like to extend my gratitude to the Boston Children's Museum and the *Living Laboratory* at the Museum of Science, Boston, as well as to all of the families who participated in these studies. This work was supported by an NSF CAREER award (#1056726) to Sara Cordes.

The Small-Large Divide:

The Development of Infant Abilities to Discriminate Small from Large Sets

Converging evidence suggests that humans, throughout the lifespan, and non-human animals have access to two distinct systems for representing number: an exact object-file system (OF) that can precisely track small sets of items (<4) and an approximate number system (ANS) primarily responsible for representing large sets (>3 ; for reviews, see Cordes & Brannon, 2008; Feigenson, Dehaene, & Spelke, 2004). This dissertation first reviews evidence indicating that these two distinct systems operate from early in development, prior to and independent of linguistic abilities, and examines evidence demonstrating the incompatibility of these two systems in the context of comparing small and large sets in infancy. Finally, the present studies address the open question of (1) when, over the course of development, do we overcome these discrimination difficulties posed by representation incompatibilities, and (2) what circumstances facilitate the ability to overcome this observed failure to compare certain sets?

Background

Recent research suggests that our primitive abilities for tracking numerical information may serve as a preverbal foundation for formal mathematics across development. For example, data suggest that individual differences in the sensitivity of our preverbal numerical representation abilities may contribute to children's initial learning of formal mathematical symbols and their meaning, and may even contribute to variation in mathematical outcomes in adults (e.g., Bonny & Lourenco, 2013; Geary, 2011; Geary, Hoard, Nugent, & Bailey, 2013; Halberda & Feigenson, 2008; Libertus,

Odic, & Halberda, 2012). Although the relative importance of preverbal and verbal numerical processing is debated (e.g., De Smedt, Noel, Gilmore, & Ansari, 2013), importantly, this relationship appears to be causal, at least in some domains of formal mathematics, such that arithmetic processing has been shown to improve following training in approximate numerical tasks (Hyde, Khanum, & Spelke, 2014; Park & Brannon, 2013). In fact, the relationship between the ANS and symbolic mathematics is evident even prior to formal mathematics experience or education, with preverbal numerical abilities in infancy and early childhood predicting formal math achievement several years later (Libertus, Feigenson, & Halberda, 2011, 2013; Starr, Libertus, & Brannon, 2013). Given that math achievement upon entering school is strongly predictive of math achievement throughout later schooling (see Duncan et al., 2007; Geary, 2013), it is critical that we understand the origins of these numerical abilities.

Numerical Abilities in Infancy

Data from human infants, children, and adults, and non-human animals support the existence of two distinct systems for tracking set sizes. The first is a noisy analog magnitude system used to represent all natural numbers in an approximate manner (approximate number system or “ANS;” Barth, La Mont, Lipton, Dehaene, Kanwisher, & Spelke, 2006; Brannon & Terrace, 1998; Cantlon & Brannon, 2006; Cordes & Gelman, 2005; Cordes, Gelman, Gallistel, & Whalen, 2001; Dehaene, 1997; Gallistel & Gelman, 2000; Meck & Church, 1983; Whalen, Gallistel, & Gelman, 1999; Xu & Spelke, 2000; see Posid & Cordes, 2014, for a review). The signature characteristic of the ANS is its adherence to Weber’s Law, such that the ease of which two sets are discriminated depends upon their ratio, not their absolute difference (e.g., discriminating 6 from 8 items,

a 3:4 ratio, should result in slower and less accurate processing compared to discriminating 6 from 12 items, a 1:2 ratio; Barth et al., 2006; Halberda & Feigenson, 2008). The precision of the ANS increases with age, such that newborns require a 3-fold change in number to discriminate between sets (e.g., 4 vs. 12), 6-month-olds require a 2-fold change (e.g., 8 vs. 16), and 9-10-month-olds notice a 1.5-fold change (e.g., 8 vs. 12; Brannon, Abbot, & Lutz, 2004; Cordes & Brannon, 2008a; Izard, Sann, Spelke, & Streri, 2009; Lipton & Spelke, 2003, 2004; Wood & Spelke, 2005; Xu, 2003; Xu & Arriaga, 2007; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005).

The second system implicated in numerical processing is termed the parallel individuation system or object-file system. This exact, one-to-one representation system can be used to track a small number of items in the visual modality (<4; Carey & Xu, 2001; Dehaene, 1997; Feigenson, Carey, & Hauser, 2002; Feigenson et al., 2004; Hyde & Wood, 2011; Leslie, Xu, Tremoulet, & Scholl, 1998; Simon, 1997). Unlike the analog magnitude system, the object file system has an absolute set size limit, such that evidence suggests that human infants can hold exactly three or fewer items in visual working memory when making numerical discriminations, and human adults able to track up to 4 or 5 items before working memory becomes overly taxed (Alvarez & Cavanaugh, 2004; Alvarez & Franconeri, 2007; Awh, Vogel, & Oh, 2006; Carey & Xu, 2001; Feigenson, 2008; Feigenson & Yamaguchi, 2009; Feigenson et al., 2004; Hyde & Wood, 2011; Jordan & Brannon, 2006; Klahr, 1973; Luck & Vogel, 1997; Luria & Vogel, 2011; Piazza, Giacomini, Bihan, & Dehaene, 2003; Scholl & Pylyshyn, 1999; Trick & Pylyshyn, 1994; Uller, Carey, Huntley-Fenner, & Klatt, 1999; Vogel, Woodman, & Luck, 2001; Zosh & Feigenson, 2012; Xu, 2003; Zosh, Halberda, & Feigenson, 2011).

Critically, the object file system seems to be a function of the visual attention system and thus is employed only when tracking visual objects, and not sounds (Luck & Vogel, 1997; vanMarle & Wynn, 2009; see Mou & vanMarle, 2013).

Importantly, although the object file system is a one-to-one tracking system, which does not inherently represent set size (and thus is non-numerical in nature, unlike the ANS), the system has been implicated in numerical tasks across the lifespan. For example, data indicate that adults employ the object file system during enumeration, such that they generally enumerate small sets (4 or fewer) effortlessly, accurately, and quickly (termed “subitizing”) whereas enumeration of larger sets (>4 items) involves effortful, slower, and error-prone processing (verbal counting; Balakrishnan & Ashby, 1982; Piazza, Mechelli, Butterworth, & Price, 2002; Trick, Enns, & Brodeur, 1996; Trick & Pylyshyn, 1993, 1994). Similarly, a numerical advantage for small sets (attributed to the object file system) has been found when human infants are presented with numerical discrimination tasks, in which greater numerical discrimination precision has been demonstrated for comparisons involving exclusively small sets of items (3 or fewer). For example, evidence suggests 6-month-olds successfully discriminate 2 from 3 items, despite being unable to discriminate a 2:3 ratio when comparing larger sets (presumably via the ANS system; e.g., 4 vs. 6, 8 vs. 12, 16 vs. 24; Antell & Keating, 1983; Bijeljac-Babic, Bertoncini, & Mehler, 1993; Cordes & Brannon, 2009b; Jordan, Suanda, & Brannon, 2008; Kobayashi, Hiraki, & Hasegawa, 2005; Lipton & Spelke, 2003, 2004; Wood & Spelke, 2005; Xu, 2003; Xu & Arriaga, 2007; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005; see Cordes & Brannon, 2008b, for a review).

Consistent with behavioral data, neuroscientific evidence also points to clear differences in how humans process small and large sets. Differences in the location and timing of brain activation have been demonstrated as a function of set size, such that adults process small, non-symbolic sets (arrays of 1-3 dots, but not symbolic number) in the area of the brain associated with visual attention (right temporo-parietal junction; Ansari et al., 2007; Hyde & Spelke, 2012). Similarly, ERP studies reveal that adults and infants who view small sets evoke earlier brain activity (with the magnitude of the activity depending on the absolute set size of the set), whereas viewing large sets results in later brain activation (with the magnitude of the activity depending on the relative magnitude of the set; Hyde & Spelke, 2009, 2011). Thus, both behavioral and neuroscientific data indicate that small and large sets are processed in a very different manner across the lifespan.

Arguably, the strongest evidence to date of the existence of these two systems comes from work with infants, where discrimination of small sets from large sets yields robust failures, despite a seemingly favorable ratio (e.g., Cordes & Brannon, 2009a; Feigenson & Carey, 2003, 2005; Feigenson et al., 2002; Lipton & Spelke, 2004; vanMarle, 2013; Wood & Spelke, 2005; Xu, 2003; see also Mou & vanMarle, 2013). For example, 6-month-old infants can reliably discriminate a 1:2 change in ratio for large sets (e.g., 4 vs. 8, 16 vs. 32; Xu & Spelke, 2000); however, when presented with a comparable 1:2 ratio for sets spanning small and large sets (e.g., 2 vs. 4 or 3 vs. 6), infants repeatedly fail on these discriminations (Cordes & Brannon, 2009a; Lipton & Spelke, 2004; Wood & Spelke, 2005; Xu, 2003). Similarly, paradigms requiring a behavioral response from infants (such as searching for toys placed within a box or crawling to a container with a

greater number of food items) reveal successful discrimination of sets containing exclusively small (1 vs. 2, 2 vs. 3) or exclusively large sets (4 vs. 8), but failures when sets cross the small-large divide (1 vs. 4, 2 vs. 4 or 3 vs. 6; Feigenson & Carey, 2003, 2005; Feigenson & Halberda, 2004; Feigenson et al., 2002; vanMarle, 2013). For example, when shown crackers being placed into two different containers, 10-12-month olds reliably crawl to the container containing the larger of two sets of crackers when one container contains 1 cracker and the other contains 2, or similarly for 2 and 3 crackers. But when one of the containers contains a small number of crackers and the other container holds a large number of crackers (2 vs. 4, 1 vs. 4, 3 vs. 6), infants choose between the containers at random (Feigenson et al., 2002). Similarly, studies employing a manual search task paradigm (MST) ask children to reach inside a box to retrieve items hidden by the experimenter. On key trials, the experimenter surreptitiously removes some of the items from the back of the box, and then the duration of the infant's searching inside the box for the hidden items is recorded. In these tasks, infants search significantly longer (compared to when the box should be empty) when a small number of items is hidden in the box and some of the items are retrieved (e.g., when 3 items are initially placed in the box and 1 item is removed), but do not search longer when a large number of items are initially placed inside the box and a small number of items are removed (4 items in and 2 removed; Feigenson & Carey, 2003, 2005; Van de Walle, Carey, & Prevor, 2000).

Despite these robust failures to compare small and large sets, there has been little research examining *when* in development infants overcome this small-large incompatibility. Although infants robustly fail to discriminate small from large sets up

until 14 months of age, there are few studies exploring this phenomenon in infants during toddlerhood. Some evidence suggests that infants continue to demonstrate this failure as late as 22 months (Barner, Thalwitz, Wood, Yang, & Carey, 2007), however once children reach 3 years of age, these discriminations do not appear to pose any difficulties (Cantlon, Safford, & Brannon, 2010). Given the robustness of these failures in early infancy, it is remarkable that by 3 years of age, young children appear to show no signs of difficulty in discriminating small from large sets. Why might this be? In particular, when, during the course of development, do children overcome these discrimination difficulties, and what factors may contribute to their ability to succeed?

Increasing ANS Precision with Age

Evidence suggests that the ANS can be used to represent values across the numerical continuum (Cordes et al., 2001) presumably making it a much more stable and reliable means for representing number across the continuum. Yet, robust small-large discrimination failures instead suggest that infants have a greater reliance upon object files for representing small sets. One account that has been provided for this over-reliance upon object files early in human development posits that relative representational precision may be the key. Specifically, object files are thought to be exact, one-to-one representations of items in the world – they are therefore reliable and precise. In contrast, ANS representations are subject to scalar noise (i.e., noise in the representation that scales in proportion to the value being represented) and as such are somewhat unreliable, particularly early in development (e.g., Cordes & Brannon, 2009; Posid & Cordes, 2014, for a review). Evidence suggests that the ANS undergoes dramatic changes in precision over the first 2 years of life, such that newborns start-out only discriminating a 3-fold

change in number, whereas by the time children reach 3 years of age, they have been shown to discriminate as fine a ratio as 1.3-fold (Izard, Sann, Spelke, & Streri, 2009; Halberda & Feigenson, 2008). It has been posited that the demonstrated reliance on object file representations early in development may be accounted for an early imprecision in ANS representations (making this system unreliable). If so, then increases in precision in the ANS with age (attributed to either maturity and/or experience; see Hyde et al., 2014; Park & Brannon, 2013) may correspond to an increasing reliance upon this system, particularly for representing small sets.

Numerical Language: A third, integrated representation system?

In addition to increases in representational precision, another observable cognitive change between infancy and early childhood is the acquisition of language, and in this case, the acquisition of *numerical* language seems particularly relevant. Several lines of research suggest that sometime in the second two years of life (ages 1-3), children begin to count and, more importantly, learn the cardinal meanings associated with those words (e.g., Condry & Spelke, 2008; Gallistel & Gelman, 1992; Wynn, 1990, 1992). Thus, it is quite possible that this newly acquired ability to talk about set sizes using a common integrated system – numerical language (i.e., number words) – may provide the first means for children to fluidly represent small and large sets.

The idea that numerical concepts may hinge upon linguistic abilities is not a new one (e.g., Carey, 2001). Researchers have posited that language provides a third and precise means of representing number across all set sizes. Dehaene and colleagues have proposed a neuro-functional model (termed the “Triple Code Model;” Dehaene, 1992; Dehaene & Cohen, 1995) of linguistic and semantic numerical integration, which

assumes that there are essentially three categories of mental representations in which number can be manipulated in the human brain: (1) a visual identification process that allows Arabic numerals to be mapped rapidly, (2) a visual or auditory process which allows numerosities to be extracted directly along a mental number line (subitizing and estimation), and (3) a verbal code that is linked to input and output routines for parsing this multi-modal numerical information (Dehaene, 1992; Dehaene & Cohen, 1995). This language faculty gives humans the ability to develop number notations especially tailored to their calculation and communication needs. This account suggests that a non-verbal, magnitude-related network and a verbally mediated fact-retrieval network operate in a closely integrated fashion (Dehaene, 1992; Dehaene & Cohen, 1995; Klein et al., 2014). Thus, numerical information, including both semantic information and rote arithmetic facts, are posited to be verbally-mediated (Dehaene, 1992; Dehaene & Cohen, 1995).

Some data point to the efficacy of the Triple Code Model and further suggest the role of verbal language as a third and integrated system for representing quantity across a small and large number range. Patients who suffered damage to their left ventral occipito-temporal area (a region involved in high-level visual identification) were unable to read words while they also demonstrated calculation errors that stemmed from a miscalculation of the digits shown. For example, when visually reading a problem such as $2+3$, they may verbally answer “seven” (tapping into their approximate number sense); however, when they hear the problem read aloud (targeting their exact verbal representation of arithmetic facts), they instead provide the correct answer (i.e., 5). Because the memory retrieval of rote arithmetic facts is assumed to depend on the verbal format, this reading deficit entailed a calculation deficit as well (Dehaene, 1998). Work

with bilinguals also demonstrates the importance of language in certain types of numerical representations. Specifically, Russian and English bilingual college students retrieved numerical information about exact numbers more effectively in the language of their training (i.e., trained in English, tested in Russian, or vice-versa). In contrast, participants retrieved numerical information about approximate numbers – as they did with non-numerical facts – with equal efficiency in their two languages (Spelke & Tsivkin, 2001). Thus, language appears to play a role in learning about exact numbers in a variety of contexts.

The idea of a linguistically-mediated representation of number does not only hold in adult or clinical populations. Support for a language-dependent account of numerical concepts early in development comes from work by Barner et al. (2007) who found support for a bootstrapping account of numerical discrimination. Their study, the only one to investigate small-large discriminations in children between the ages of 14-36 months, found infants continued to search for items in the MST task when 4 items were initially placed inside the box and 1 item was retrieved (4 versus 1 comparison) around 22-24 months of age -- an age which was found to coincide with the initial onset of the production of plural nouns in spoken language in a second population of children. The authors theorized that children's developing singular-plural morphology aided children's abilities to discriminate 1 item (singular) from 4 items (plural)¹. Importantly, however, infants of this age continued to fail to discriminate 2 from 4 items in their MST task (plural from plural; Barner et al., 2007; Li, Ogura, Barner, Yang, & Carey, 2009). Thus,

¹ It should be noted that a between-subject design was employed, such that plural noun usage was not assessed in the children who succeeded in their MST 4 vs 1, leaving open the question whether a direct relationship between the two exists.

although preliminary evidence supports a language-dependent account of 1 vs. 4 discriminations, it is an open question whether linguistic abilities may also play a role in small-large discriminations involving sets larger than 1 item (2 vs. 4).

Perceptual Variability:

Real world experiences may also play a role in infant abilities to succeed in discriminating small from large sets. In particular, sets in the real-world are rarely homogeneous (i.e., all the same shape, color, size), making the use of exclusively homogeneous dot arrays or sets of toys in laboratory research less consistent with numerical experiences in children's environments. As such, greater experience (and success) discriminating heterogeneous sets in the world may support and further enhance infant abilities to determine when to rely upon object files versus ANS.

Previous research finds that perceptual variability aids verbal counting in adults, facilitating the distinction (accuracy and speed of processing) between counted and to-be-counted items (Frick, 1987; Trick, 2008). Similar findings have been demonstrated in young children (3-10 years), such that children are better at counting perceptually rich displays (particularly when they have low knowledge of the objects; Peterson & McNeil, 2012) and are better at rapidly discriminating sets based on number when they are perceptually variable (made up of all different shapes and colors; Posid, Huguenel, & Cordes, in preparation²). Moreover, heterogeneity has been shown to facilitate discrimination in infancy as well (Feigenson, 2005; Tremoulet, Leslie, & Hall, 2001;

² Notably, the opposite pattern has been found when children engage in exact numerical matching or cardinality tasks, such that performance is better on homogeneous trials compared to heterogeneous (Mix, 1999, 2008; Posid & Cordes, 2014).

Wilcox, 1990). For example, seven-month olds are more likely to discriminate small sets and large sets based upon number when presented with perceptually variable items (Cordes & Posid, in progress; Feigenson, 2005). Data from these studies spanning a developmental range suggest that, at least in some instances, perceptual variability may aid infants', children's, and adults' abilities to attend to numerical attributes of an array, helping to highlight the salience of the individual items to be enumerated.

Taken together, evidence suggests that perceptual variability may aid numerical discriminations across the lifespan. But why? One possibility is that distinguishing characteristics of each item may make it easier to track individual objects in a display, thereby reducing the demands of working memory, and potentially increasing the capacity of the number of items that may be tracked in working memory (via object files; e.g., Zosh & Feigenson, 2015). Specifically, work across the lifespan indicates that infants can hold approximately three items in working memory, with adult working memory ability tapering out at three or four items (e.g., Kaldy & Leslie, 2005; Ross-Sheehy, Oakes, & Luck, 2003). In fact, adults' estimates of arrays containing a certain number of items (specifically in the case of this number exceeding typical working memory storage capacity, >4) decreases as array size increases (Luck & Vogel, 1997). In contrast, evidence to date suggests that infants fail to remember any of the items presented to them when their working memory capacity is exceeded; however, these previous studies have exclusively relied on homogenous sets to make-up their stimuli (e.g., Barner et al., 2007; Feigenson et al., 2002; Feigenson et al., 2004; Feigenson & Carey, 2003, 2005). What if infants could distinguish between these elements and individuate the specific elements? Research with adults has employed arrays containing

perceptually variable items when testing working memory, thus testing purely homogeneous arrays with infants does not yet speak to how this adult ability to hold items in working memory may develop and mature across the lifespan (except see Feigenson, 2005). Therefore, perceptual variability may aid infants' memory for items to be stored in working memory. Thus, in contrast to numerical language – which may facilitate small/large discriminations via providing a third, integrated representation system – experiences with perceptually variable arrays may help to increase the capacity of the object file system in infancy (and presumably later in development), similarly resulting in successful discriminations of small from large. In the proposed studies, I test this hypothesis.

The Present Study

The present study examines when it is, over the course of early human development, that young children overcome the numerical discrimination difficulties posed by the interaction of the object file system and the analog magnitude system, and whether numerical language or perceptual variability may impact these abilities. Almost no research to date has examined the development of this understanding in toddler-hood (between 15-35 months of age; but see Barner et al., 2007). Although infants have demonstrated robust small/large discrimination failures up until 14 months (e.g., Feigenson & Carey, 2003, 2005; Feigenson et al., 2002; Xu, 2003), evidence suggests that they overcome these difficulties by at least 3 years of age (Cantlon et al., 2010). Yet, it is unclear when in development this occurs and what the mechanism is behind these changes in numerical abilities. In a series of experiments, we employ a manual search task (MST) paradigm to explore infant abilities to compare across exclusively small sets

(2 vs. 3) and across sets that span the small-large divide (2 vs. 4). Unlike any previous study to date, we look across a wide developmental age range (14-36 months) to map out the developmental trajectory of these numerical discrimination failures.

Additionally, we explored two distinct factors that may facilitate these discriminations in young children. The first of these factors is numerical language. In one of three conditions (see Methods below), items to be hidden were explicitly counted for the child, and the set cardinality was labeled, prior to being placed inside the box. It was hypothesized that the exposure to numerical language just prior to the task may prompt children to track items hidden within the box using a linguistically-dependent system, number words. Additionally, children's counting proficiency was assessed after the discrimination task to determine whether, similar to as has been posited for plural morphology, successful search patterns for small/large discriminations may be dependent upon a child's ability to produce numerical language.

In a third condition, we will also investigate the contribution of perceptual variability. Recent research suggests that, under certain circumstances, perceptual variability aids exact numerical judgments, such as counting items (adults: Frick, 1987), increasing speed of enumeration (adults: Trick, 2008), improving discrimination abilities (children: Posid et al., in preparation), and facilitating discrimination during infancy (Feigenson, 2005). Results from these studies suggest that perceptual variability may aid in infants', children's, and adults' ability to discriminate between sets because heterogeneity highlights the salience of the individual items to be enumerated, lightening the cognitive load imposed on working memory (Feigenson, 2005; Posid et al., in preparation). In the present study, we explore whether a heterogeneous array will increase

success on a manual search task by improving infants' individuation of objects by infants, thus potentially expanding their working memory in this specific context.

If, on the other hand, we find no impact on small-large discrimination abilities following exposure to numerical language and/or perceptually variable sets, this may alternatively support claims that increasing representational precision, with development, may prompt eventual successful discriminations. Specifically, equal performance between either of these experimental conditions and the baseline, or “standard”, condition (see Methods below) would imply that the increasing precision of the ANS predicts this transition in natural development and that neither numerical language (or, at least, exposure to numerical language) nor perceptual variability provides enough reliable and precise information to overcome this initial bias to represent numerical sets via two distinct systems.

Thus, the present set of studies explore three non-mutually exclusive predictions:

- (1) Sometime over the course of 14-36 months, infants begin to successfully compare small and large sets. Based on one previous study reporting failure on a small-large discrimination at 22-24 months of age and another study reporting success at 36 months of age, it is anticipated that this natural change in development may occur sometime during the third year of life (24-36 months).
- (2) Numerical language may provide a third, integrated system on which infants can rely when discriminating across sets. This could be due to several mechanisms; however, in the present study we explore two of these. First, we examine whether providing exposure to numerical language (specifically, verbal counting and cardinality) may facilitate infant abilities to discriminate small vs. large sets by priming infants with a third, reliable

system to think about the number of items within a set. Second, we examine whether infant abilities to produce the count words may significantly predict success when comparing sets.

(3) Perceptual variable sets may facilitate small-large discriminations as they may make it easier for infants to track items in the set, thereby decreasing demands on working memory, and momentarily expanding the limits of their visual working memory. If so, infants may rely on the object file system exclusively to represent individual items within a set at a given time (e.g., 4 items instead of 3 or fewer items), thereby taxing visual working memory to a lesser degree than a strictly homogenous set of items.

Experiment 1

Methods

Participants

One hundred and sixty nine infants between 18- and 36-months participated in this study. Participants were divided into three age groups: 18- to 23-month-olds ($N=43$; $M=21.2$ mos, $SD=1.51$ mos; 21 females), 24- to 29-month-olds ($N=57$ $M=26.5$ mos, $SD=1.66$ mos; 29 females), and 30- to 36-month-olds ($N=70$; $M=33.1$ mos, $SD=2.0$ mos, 41 females). Infants were randomly assigned to one of three conditions: “Standard” ($N=58$), “Heterogeneity” ($N=63$), and “Language” ($N=49$). See Table 1 for a breakdown of the number of participants per age in each Condition. Additionally, children were randomly assigned to one of two counterbalanced orders: 2v4 first ($N=84$) or 2v3 ($N=86$) first (counterbalanced within each age group and condition). An additional 14 subjects

completed half of the test trials (one of the two comparisons)³. Lastly, 53 additional participants were excluded from the study due to parental interference ($N=10$), failure to complete a sufficient number of test trials (at least one comparison, $N=32$), equipment error (video camera did not record; $N=3$), experimenter error ($N=4$), fussiness ($N=2$), or never searching over the course of the experiment ($N=2$). Children were recruited via phone or email for a single visit to the Infant and Child Cognition Lab or from the Boston Children's Museum and Living Laboratory at the Museum of Science, Boston.

Stimuli

Stimuli and procedure were adapted from Feigenson & Carey's (2003, 2005) manual search task paradigm. Infants watched the researcher hide small rubber ducks in a black cardboard box measuring 31.5 x 25 x 12.5 cm. The box was painted black with a 14 x 7.5 cm opening on both the front and back sides of the box. Both openings were covered with the same blue spandex material to prevent subjects from seeing the contents of the box (and to prevent subjects from seeing light enter the back of the box). For the Standard and Language conditions, yellow 2-inch rubber ducks homogeneous in appearance were used as stimuli. For the Heterogeneous condition, 2-inch rubber ducks of the same size were used as stimuli; however, each duck looked like a different animal (e.g., white cow, pink pig, brown monkey, grey elephant), making the stimuli heterogeneous in appearance.

Procedure

Standard Condition:

³ These participants were excluded from all primary analyses, but were included in secondary analyses (see Results below) examining infants' performance on the first test comparison only.

Infants sat at a table facing the box with the experimenter on the opposite side. A video camera recorded the trials for coding purposes. The experimenter showed the participant that he or she could reach into the box and encouraged the subject to do so (“Look! I can reach into my box! Can you reach into my box?”). During familiarization, the experimenter first showed the infant one rubber duck and said, “Look what I have!” and then placed it inside the box. The researcher pushed the box towards the infant and asked, “What’s in the box?” The infant reached inside the box and retrieved the hidden duck. The familiarization trial was repeated a second time to ensure the infant understood the rules of the game. Each child participated in two comparisons (3 in the box, 2 taken out (2v3) and 4 in the box, 2 taken out (2v4)) and the order in which each infant saw the comparisons was counterbalanced. In all conditions, each comparison (2v3 and 2v4) as well as baseline search trials (2 in the box, 2 taken out (2v2)) were repeated twice. Half of the children were run in the order: 2v3, 2v3, 2v2, 2v2, 2v4, 2v4 while the other half were run in the reverse order: 2v4, 2v4, 2v2, 2v2, 2v3, 2v3. The 2 vs. 2 condition was run to determine baseline search times (as per Feigenson & Carey, 2003, 2005); in other words, this baseline measure was indicative of each individual child’s propensity to search in the box (e.g., some children may just like to search inside the box more than others).

2 vs. 3 Comparison

For the 2v3 comparison, the experimenter sequentially placed three ducks on top of the box and said, “Look at my toys!” To control for longer exposure to the full set of stimuli in the Language condition (see below), the experimenter pointed to each duck one time in sequential order (as per Feigenson & Carey, 2003, 2005). The toys were then

placed sequentially inside the box through the opening facing the infant. Unaware to the participant, the experimenter secretly removed one duck from the back of the box. The experimenter then pushed the box towards the infant and said, “What’s in the box?” The infant retrieved two ducks from the box and was allowed to handle them for a few seconds before the experimenter asked for them back, placing them on top of the box in view of the infant. After the infant had released the first two ducks, a 10-second measurement period (“Expected Full” trial) was given in which the infant was allowed to search inside the box while the experimenter watched the time. After the 10-second measurement period, the experimenter said, “Can I try?” and reached into the box, retrieving the “missing” duck. Then, a second 10-second measurement period (3v3 “Expected Empty” trial) occurred in which the infant was allowed to search in the box while the experimenter watched the time.

2 vs. 4 Comparison

The 2v4 comparison mirrored the procedure used in the 2v3 comparison, with the following differences: The experimenter placed four ducks on top of the box before initially hiding them within the box. After hiding four ducks in the box, the experimenter secretly removed two of the ducks from the back opening of the box. After the child removed the first two ducks from inside of the box and the 10-second “Expected Full” measurement period occurred, the experimenter “found” the two remaining ducks. This was again followed by a 10-second measurement period (4v4 “Expected Empty”).

2 vs. 2 Baseline Search

To determine baseline search times, the experimenter sequentially placed two ducks on top of the box and said, “Look at my toys!” The toys were then placed

sequentially inside the box through the opening facing the infant. The experimenter then pushed the box towards the infant and said, “What’s in the box?” The infant retrieved both ducks from the box and was allowed to handle them for a few seconds before the experimenter asked for them back, placing them on top of the box in view of the infant. After the infant had released the two ducks, a 10-second measurement period (“Expected Empty” trial) was given in which the infant was allowed to search inside the box while the experimenter watched the time.

Language Condition:

The stimuli and procedure of the Language condition was identical to that of the Standard condition with the following difference: After the items to-be-hidden were first placed on top of the box, the experimenter tapped each item individually while counting and labeling the set (e.g., “Look at my toys! I have 1, 2, 3, 4! That’s 4!”). The experimenter then proceeded to place the items in the box and the procedure continued as in the standard condition.

Heterogeneity Condition:

The procedure of the Heterogeneity condition was identical in all regards to that of the Standard condition with one exception: Instead of a homogeneous array of toys to-be-hidden (i.e., all yellow ducks), the items were heterogeneous in appearance (i.e., different color animals); however, the size and dimensions of the toys were identical to the yellow ducks used in the Standard and Language conditions (see Figure 1).

Data Coding and Analysis

1. All data coding was performed offline using Preferential Looking Coder (Version 1.0; Libertus, 2008) in 100 ms frames. Inter-coder reliability was calculated for

each subject on a trial-by-trial basis (as reliability = (# of frame agreements – # of frame disagreements) / total time). Two independent researchers -- blind to the Experimental Condition -- coded each video (thus, each participant's video was coded twice by two different experimenters). Inter-rater reliability for all trials across subjects was 87.0%. Any disagreements were settled by a third (independent) coder for any participant where agreement was less than 80% .).

2. As per Feigenson & Carey (2003, 2005), data analysis was done using mean search times, in which an individual difference score was calculated for each participant (one per comparison).

Expected Full Trials: For the 2v4 comparison, an “Expected Full” score was created by taking the average search time (out of 10 seconds) on both “Four Full” trials; that is, after four ducks were placed in the box, the infant retrieved two of them, and was then given 10 seconds to search for any remaining items. Because each comparison was done twice, this created two individual “Four Full” trials that were averaged to create an “Expected Full” score. Identical data preparation was performed for the 2v3 comparison.

Expected Empty Trials: Similarly, after the experimenter retrieved the final two ducks from the box, infants were again given a 10-second search period (“Four Empty” trials). These two measurement periods were averaged to create an “Expected Empty” score. The 2v2 baseline search time was conducted in order to measure infants' individual propensities to search in the box (presumably some infants are more inclined to search than others). This resulted in two “Two Empty” measurement scores, one which was conducted directly adjacent to the 2v4 comparison and one which was conducted directly adjacent to the 2v3 comparison. Per Feigenson and Carey (2003, 2005), in order

to account for this individual propensity to search in general during the task, the associated “Two Empty” trial was averaged with the infant’s “Expected Empty” score for that comparison, thus creating an “Expected Empty” score for each comparison that accounted for baseline searching. For example, an infant’s search time from both “Four Empty” Trial would be averaged with an infant’s search time for the adjacent “Two Empty” trial. The same was done for the 2v3 comparison and its associated 2v2 trial.

Difference Scores: Finally, a difference score was calculated for each comparison by subtracting the “Expected Empty” score from each infant’s “Expected Full” score (for each comparison). Therefore, each participant had two Difference Scores, one for the 2v3 comparison and one for the 2v4 comparison, which were entered in subsequent analyses as our dependent variable. Difference scores that are positive and significantly greater than zero indicate success on that numerical comparison.

Results

Per Feigenson and Carey (2003, 2005), the effects of Trial Type (Expected Full vs. Expected Empty) and Numerical Comparison (2v3 and 2v4) on infants’ numerical discriminations were examined by computing difference scores between Expected Full and Expected Empty trials across the two comparisons. These difference scores were subjected to a Comparison (2: 2v3, 2v4) X Age Group (3: 18-23 months, 24-29 months, 30-36 months) X Condition (3: Standard, Heterogeneity, Language) X Order (2v3 first, 2v4 first) mixed measures ANOVA. Results revealed a main effect of Age Group ($F(2, 152)=10.9, p<.001, \eta_p^2=.126$; Figures 2 and 3); however, there was no main effect of Condition ($F(2, 152)=2.2, p>.1, \eta_p^2=.028$ nor Comparison ($F(1, 152)=1.31, p>.2, \eta_p^2=.009$), indicative of general and robust success across all age groups and conditions.

There was a significant Comparison X Order interaction ($F(1, 152)=12.4, p=.001, \eta_p^2=.075$). There were no other main effects or interactions (all p 's $>.1$).

The main effect of Age Group indicated that children searched significantly longer on Expected Full trials (compared to Expected Empty trials) with age ($M_{18-23\text{ mos}}=1.7; M_{24-29\text{ mos}}=2.91\text{s}; M_{30-36\text{ mos}}=3.53\text{s}$; all p 's $<.01$; Figure 4). Although there was no significant main effect of condition, one interesting trend was longer searching in the Heterogeneous condition relative to the Standard condition (Standard vs. Heterogeneity: $t(122)=2.26, p=.026, \text{Cohen's } d=.406$), with the Language condition falling in between the two ($M_{\text{heterogeneity}}=3.43\text{s}; M_{\text{language}}=2.46\text{s}; M_{\text{standard}}=2.50\text{s}$; Heterogeneity vs. Language: $p=.106$; Standard vs. Language: $p>.8$). Lastly, the Comparison X Order interaction indicated that participants performed better in the first comparison they were presented (2v4 first: $M_{2v4}=3.42\text{s}, M_{2v3}=2.16\text{s}$; 2v3 first: $M_{2v4}=2.33\text{s}, M_{2v3}=2.97\text{s}$). No other significant main effects or interactions were obtained⁴.

Despite these differences, children searched significantly longer in the Expected Full trials compared to the Expected Empty trials across all Age Groups X Comparisons X Conditions, with the exception of one (Language: 18-23 months, $t(10)=1.21, p>.2, \text{Cohen's } d=.77$). All other difference scores were found to be significantly greater than 0 (positive scores indicate success; Tables 2 and 3, Figure 5). Additionally, age (in days) was significantly correlated with an increased difference score for both comparisons (2v3: $r=.254, p=.001$; 2v4: $r=.214, p=.005$). Thus, these results confirm that even children in the youngest age groups successfully discriminated sets of 2 vs. 4, even in the

⁴ To ensure that experience with the task did not influence searching behavior, similar analyses were performed on the first comparison presented to the participant. An identical pattern of results was obtained.

Standard Condition. Together, these preliminary analyses indicate that infants younger than 3 years of age can successfully discriminate between a comparison spanning the small-large divide in a manual search paradigm, with this spontaneous ability apparent as early as 18-23 months of age.

Counting Behavior:

A chi-square analysis of children who verbally counted the items (“Counters”: coded as “1” if they demonstrated successful counting of 2 and/or 4 items), compared to those who did not count the items (“Non-Counters”: coded as “0”), revealed that a similar number of children counted items following each of the three conditions (Standard: Non-Counters: 62.5% vs. Counters: 37.5%; Heterogeneity: Non-Counters: 62.7% vs. Counters: 37.3%; Language: Non-Counters: 62.5% vs. Counters: 37.5%; $\chi^2(2, N=187)=.001, p>.9, Cramer's V=.002$; see Table 4). However, not surprisingly, the proportion of Counters increased as a function of age (18-23 months: Non-Counters: 76.6% vs. Counters: 23.4%; 24-29 months: Non-Counters: 71.2% vs. Counters: 28.8%; 30-36 months: Non-Counters: 45.9% vs. Counters: 54.1%; $\chi^2(2, N=187)=14.8, p<.001, Cramer's V=.281$; see Table 5).

Although modeling counting behavior to infants in our task (Language condition) did not appear to improve discrimination abilities, we also explored whether a child’s ability to verbally count items in the set (assessed after the discrimination task) may have predicted performance in our task. Linear regression analyses were conducted entering Counting Behavior, Age (Group), Language Condition, and Heterogeneity Condition as predictors of difference scores in the two comparison conditions. In the 2v3 Comparison, although the final model was an overall significant fit ($R^2=.100, p=.001$), only Age Group

emerged as a significant predictor (Age: *Standardized Beta* = .308, $p < .001$; Counting Behavior: *Standardized Beta* = -.047, $p > .5$, Language Condition: *Standardized Beta* = .022, $p > .7$, Heterogeneity Condition: *Standardized Beta* = .105, $p > .3$). In the 2v4 Comparison, the final model was also an overall significant fit ($R^2 = .075$, $p = .008$). Age (group) was again a significant predictor of success (*Standardized Beta* = .187, $t(174) = 2.47$, $p = .015$) but additionally, Counting Behavior was also a marginally significant predictor (*Standardized Beta* = .129, $t(174) = 1.71$, $p = .089$). Neither Heterogeneity nor Language Conditions predicted success on the 2v4 Comparison (Heterogeneity: *Standardized Beta* = .09, $p > .2$; Language: *Standardized Beta* = -.023, $p > .7$). A secondary linear regression analysis of entering Age, Counting Behavior, and an Age X Counting Behavior interaction term as predictors indicate that the interaction term was not a predictor of difference scores on 2v4 (*Standardized Beta* = .266, $p > .2$); thus, Age and Counting Behavior were not simply intertwined, but seemed to contribute unique variance to searching behavior. Together, these analyses suggest that children learn to compare sets with maturation. Although their ability count the number of items in the set may *also* be a helpful component in their developing ability to compare small and large sets along the same continuum, it is not necessary for this ability.

Discussion

The main finding from this experiment was that children succeeded in comparing both small sets (2 vs. 3) *and* sets spanning the small-large divide (2 vs. 4). Although previous work indicates that infants can compare small sets or large sets (e.g., 1 vs. 2 or 5 vs. 10; see Posid & Cordes, 2014, for a review) with ease (Xu, 2003; Xu & Spelke, 2000), a phenomenon exists whereby infants fail to compare small and large sets (e.g., 2 vs. 4 or

3 vs. 6; Cordes & Brannon, 2009; Feigenson, Carey, & Hauser, 2002; Xu, 2003; Xu & Spelke, 2000). This failure is posited to stem from infants' use of two incompatible systems for representing number (Cordes & Brannon, 2009; Feigenson et al., 2002).

Barner and colleagues found that 22-24 month old infants, who may have begun to produce their own plural nouns, succeed at a 1 vs. 4 comparison (spanning the small-large divide), yet continue to fail to compare plural sets (i.e., where both sets contain more than 1 item; 2 vs. 4) across the range (Barner et al., 2007). However, notably, that is the only study to explore small/large discriminations in children between the ages of 14-36 mos. Thus, the earliest documented evidence of successful discrimination appears in 3-year olds (Cantlon et al., 2010), suggesting this is a very late-developing ability. Findings of the current study, however, provide the first evidence to suggest that infants younger than 36 months may be able to compare sets spanning the small/large boundary.

Our results revealed that infants searched more with age (longer relative searching on Expected Full trials relative to Expected Empty across development, as indicated by higher and more positive difference scores). It appears that older infants, more confident in their numerical representations, were more inclined to believe that objects remained in the box and thus continued to search longer for them compared to younger infants. Thus, consistent with a developmental trend from a documented early failure at 14 months of age to discriminate small from large sets (Feigenson & Carey, 2003, 2005) to a late success at 3 years of age (Cantlon et al., 2007), our data reveal increased searching with age consistent with increases in confidence in numerical abilities across toddlerhood.

Secondary analyses reveal that, unsurprisingly, infants were able to engage in verbal counting with age. Of interest, however, was the fact that these infants who engaged in verbal counting were the ones more likely to succeed in discriminating 2 vs. 4, even when controlling for age, but this pattern did not emerge for the 2 vs. 3 condition, consistent with two distinct patterns across the comparison types. That is, as has been previously posited for exclusively small set comparison (e.g., Feigenson & Carey, 2003), infants may have primarily relied upon object file representations, and thus an ability to use numerical language may not have been an important contributor to success. In contrast, in the 2 vs. 4 comparison, being able to talk about number in a meaningful way may have provided infants a third, precise means for tracking both large and small sets together.

Lastly, we also examined the impact of numerical language and heterogeneity on infants' ability to compare small and large sets across conditions. Due to our high level of success in this age range, the present study does not truly determine whether the use of numerical language or item heterogeneity aided performance on the task relative to the Standard condition. Analyses suggests that infants in the heterogeneous condition did out-perform those in the Standard condition, suggesting that the inclusion of perceptually variable items may have improved infant abilities to track items placed within the box. However, given the pattern of success in both the small-small and small-large comparisons in the standard condition, it is unclear whether perceptual variability actually increased the number of available object files used for tracking, or instead provided one more cue to increase infant confidence in their abilities in the task cannot be determined.

In additional experiments, we explored the impact of these variables in younger infants (14-17 month olds), where it was expected that infants would show the documented pattern of successful discrimination for exclusively small sets (2 vs. 3) and failure for those crossing the small/large boundary (2 vs. 4). To begin, in Experiment 2 we tested 14-17 month old infants in the Standard condition to ensure that the documented pattern is observed.

Additionally, we identified one minor methodological difference between our manual search task and that used by Feigenson and Carey (2003, 2005) that may have contributed to the overall pattern of success in Experiment 1. In Experiment 1, items retrieved from the were placed on top of the box, in plain view of the infant, possibly allowing the child to compare the number of items on top of the box with a mental representation of the number of items that were originally placed inside the box. In contrast, Feigenson and Carey (2003, 2005) did not leave retrieved items in visual sight of the child, but had children deposit the toy into a receptacle that was out of view. It is possible that the visual working memory of infants who participated in Feigenson and Carey's manual search paradigm was taxed more than that of infants in our version of the task, who could visually track the items that had already been retrieved. If this was the case, perhaps the youngest children to demonstrate success in our first Experiment (18-23 month olds) would fail to compare small and large sets if they could not visually track the items hidden and retrieved during the task. Thus, in Experiment 2, in addition to testing 14-17 month olds, we also tested infants throughout the age range again (18-36 months) following the specific parameters used in Feigenson and Carey (2003, 2005), such that retrieved items were immediately deposited in an opaque container so that infants could

not visually track the numerical pairs to be enumerated. In this way, Experiment 2 assessed the developmental trajectory of the natural emergence of small-large discriminations in toddlerhood.

Experiment 2

Participants

Sixty-four infants between 14- and 36-months-old participated in this study. The participants were divided into four age groups (consistent with Experiment 1): 14- to 17-month-olds ($N=14$; $M=15.4$ mos, $SD=1.02$ mos, 6 female), 18- to 23-month-olds ($N=12$; $M=20.8$ mos, $SD=1.71$ mos, 7 female), 24- to 29-month-olds ($N=16$; $M=25.9$ mos, $SD=1.78$ mos, 11 female), and 30- to 36-month-olds ($N=22$; $M=33.1$ mos, $SD=2.14$ mos, 13 female). Again, to control for the order in which infants saw each numerical comparison (per Feigenson & Carey, 2003, 2005), children were randomly assigned to one of two counterbalanced orders: 2 vs. 4 first ($N=36$) or 2 vs. 3 ($N=28$) first. Additionally, 26 subjects completed half of the test trials (one of the two comparisons)⁵. These participants were excluded from all primary analyses, but were included in secondary analyses (see Results below) examining infants' performance on the first test comparison only. Twenty-one additional participants were excluded from the study due to parental interference ($N=2$), not completing a sufficient number of test trials ($N=6$), fussiness ($N=4$), equipment error (video camera did not record; $N=2$), experimenter error ($N=5$), and for never searching during the experiment ($N=2$). Children were recruited via

⁵ These participants were excluded from all primary analyses, but were included in secondary analyses (see Results below) examining infants' performance on the first test comparison only.

phone or email for a single visit to the Infant and Child Cognition Lab or from the Boston Children's Museum and Living Laboratory at the Museum of Science, Boston.

Stimuli, Procedure, and Data Coding

The stimuli and procedure were identical to that of the Standard Condition in Experiment 1 with the following difference: After the hidden items were retrieved from the box (either by the child or the experimenter), they were immediately placed in an opaque bucket per Feigenson and Carey (2003, 2005). This container was tall and wide enough to conceal the toys that had been retrieved from the box.

As in Experiments 1, per Feigenson and Carey (2003, 2005), difference scores were calculated for each numerical comparison based on an average of each participant's search time in the Expected Full trials and the Expected Empty trials (collapsed across search time for the Four Empty trials and the associated Two Empty baseline trial, e.g.). Again, a significant positive difference score is indicative of success on that numerical comparison.

Two independent researchers -- blind to the Experimental Condition -- coded each video (thus, each participant's video was coded twice by two different experimenters). Inter-rater reliability for all trials across subjects was 96.3%. Any disagreements were settled by a third (independent) coder for any participant where agreement was less than 80%.

Results

As in Experiment 1, we examined the impact of Numerical Comparison (2v3 and 2v4) on infants' numerical discriminations by computing difference scores between Expected Full and Expected Empty trials across the two comparisons. These difference

scores were subjected to a Comparison (2: 2v3, 2v4) X Age Group (4: 14-17 months, 18-23 months, 24-29 months, 30-36 months) X Order (2v3 first, 2v4 first) mixed measures ANOVA. Results revealed a main effect of Age Group ($F(3, 56)=6.12, p=.001, \eta_p^2=.303$), such that difference scores increased positively with age (14-17 mos: $M=-.483s$, 18-23 mos: $M=1.62s$, 24-26 mos: $M=2.91s$, 30-36 mos: $M=2.89s$). Although infants in the youngest age range differed significantly from all other age groups ($p<.05$ for all), older children (18-36 mos) did not differ significantly from each other (all $p's>.1$). There was no main effect of Comparison ($p>.7, \eta_p^2=.006$), nor did Comparison interact with age ($p>.5, \eta_p^2=.04$), indicating that infants across the board did not differ in their performance on the 2v3 and 2v4 trials (all $p's>.1$; see Figure 7).

Post-hoc analyses of difference scores for each Age Group X Comparison were calculated (versus 0). For the 14-17 month olds, their difference scores did not significantly differ from zero (2v3: $M=-.78s, t(13)=1.45, p=.172, Cohen's d=.80$; 2v4: $M=-.661s, t(13)=.784, p>.4, Cohen's d=.43$). In contrast, difference scores for nearly all older age groups across comparisons were positive, indicating they searched significantly longer in Expected Full trials compared to Expected Empty trials: 18-23 mos: 2v3: $M=1.08s, t(11)=1.76, p=.106, Cohen's d=1.06^6, approaching$; 2v4: $M=2.6s, t(11)=5.45, Cohen's d=3.29, p<.001$; 24-29 mos: 2v3: $M=3.03s, t(15)=3.33, p=.005, Cohen's d=1.72$; 2v4: $M=2.71s, t(15)=4.09, p=.001, Cohen's d=2.11$; 30-36 mos: 2v3: $M=3.2s, t(21)=4.71, p<.001, Cohen's d=2.06$; 2v4: $M=2.85s, t(21)=3.59, p=.002, Cohen's d=1.57$).

⁶ Although this p-value is currently not significant at the $\alpha=.05$ level, we believe that this is due to a small sample size, as it generally follows the same trend as the 2v4 comparison, as well as both comparisons across the older age range. More data are needed to further probe this potentially spurious result.

Additionally there was a three-way significant interaction between Comparison, Age Group, and Order ($F(3, 56)=2.99, p=.039, \eta_p^2 =.128$). Generally speaking, this interaction revealed that, unsurprisingly, infants generally searched longer (indicated by higher positive difference scores) on the comparison with which they were first presented⁷.

Thus, these results confirm that, although infants in the 14-17 month age range fail at both comparisons, again, children between 18 and 36 months of age successfully discriminated sets of 2 from 4, even in this Standard Control Condition (replicating the methods used in Feigenson & Carey, 2003, 2005). Together with findings from Experiment 1, these preliminary analyses indicate that infants younger than 3 years of age can successfully discriminate between a comparison spanning the small-large divide in a manual search paradigm, with this spontaneous ability apparent as early as 18-23 months of age.

Counting Behavior:

We also analyzed children's ability to verbally count items at the end of the manual search task. A chi-square analysis of children who verbally counted the items ("Counters": coded as "1" if they demonstrated successful counting of 2 and/or 4 items⁸), compared to those who did not count the items ("Non-Counters": coded as "0"), children

⁷ When the first comparison was singularly analyzed as a dependent variable in a Repeated Measures ANOVA, with Comparison and Age entered as between- and within-subject variables, the same trend was observed. Therefore, these analyses are not provided, as they confirm the results presented above in the omnibus ANOVA.

⁸ Note, preliminary analyses indicate that entering Counting Behavior as 0 or 1 yielded the same results as entering Counting Behavior as 0, 2 and 4. Thus, we collapsed "Counting" into those participants who could count either 2 and/or 4 items.

were more likely to count the items with age ($X^2(3, N=76)=25.5, p<.001$, *Cramer's V*=.395; see Table 6).

To explore whether counting abilities may have played a role in performance in the MST task, linear regression analyses were again conducted entering Counting Behavior and Age (Group) as predictors of difference scores in the two comparison conditions. Overall, both models significantly predicted search times ($R^2_{2v4}=.168, p=.004$; $R^2_{2v3}=.190, p=.001$), with Age being the only significant predictor in both comparison types (2v3 comparison: Age *Standardized Beta* = .492, $t(74)=4.06, p<.001$; Counting Behavior: *Standardized Beta* = -.170, $p>.1$, 2v4 comparison: Age: *Standardized Beta* = .363, $t(61)=2.68, p=.01$; Counting Behavior: *Standardized Beta* = .08, $p>.5$). Together, these analyses suggest that children learn to compare sets with maturation, in contrast to Experiment 1, where their ability to count (marginally) predicted success on the 2v4 comparison. Thus, although children's ability to count clearly increases with age, at least in this version of the MST paradigm, it does not appear to aid their ability to compare sets.

Cross-Experiment Analyses:

Findings from both Experiments 1 and 2 indicate that 18-23 month olds can successfully discriminate small-small (2 vs. 3) and small-large (2 vs. 4) sets, regardless of whether the items retrieved from the box were visible to the infant or not. This is the first evidence to date to demonstrate that infants younger than 3 years of age can successfully cross this small-large divide when making spontaneous numerical comparisons. To assess the impact of our different methods (whether items were visible or not), we ran a

cross-experiment analysis (Experiments 1 (Standard Condition only) vs. Experiment 2 to examine whether infants' searching differed across the two paradigms.

We first ran a Mixed Measures ANOVA using difference scores on Comparison (2: 2 vs. 3, 2 vs. 4), Age Group (3: 18-23 months ($N=29$), 24-29 months ($N=32$), 30-36 months ($N=41$)), and Experiment (2). Again, there was a main effect of Age Group ($F(2, 96)=16.5, p<.001, \eta_p^2=.256$), indicating that infants searched more with age. Importantly, there was no main effect of Experiment ($F(1, 96)=2.75, p>.1, \eta_p^2=.028$), nor any interactions involving this variable ($p's>.1$), suggesting that whether or not the items were visible after being retrieved from the box did not impact search times. Additionally, there was no main effect of Comparison ($F(1, 96)=1.1, p>.2, \eta_p^2=.011$), nor any other interactions ($p's>.1$).

Discussion

Data from Experiment 2 replicate findings from Experiment 1, indicating that infants as young as 18-23 months of age can discriminate both small-small (as previously documented; e.g., Feigenson & Carey, 2003, 2005; Feigenson et al., 2002; Xu, 2003; Xu & Spelke, 2001) and small-large sets. This is the *first* study (except see Barner et al., 2007) to track infants' developing understanding of small and large sets using the same task across this wide age range, revealing an early failure (in 14-17 months; as in Feigenson & Carey, 2003, 2005) to later success (by 18-24 months). These data indicate that infants as young as 18-months of age can compare small vs. large sets, and that they can do so when they have visual access to the small set (Experiment 1) and without visual access to the set (Experiment 2).

Thus, results of Experiments 1 and 2 provide an answer to the first research question by revealing that an ability to discriminate small from large sets emerges sometime by the age of 18-24 months. However, robust successes by infants in Experiment 1 three conditions prohibited our ability to address whether numerical language and/or perceptual variability may facilitate small-large discriminations in infancy.

The results of Experiments 1 and 2 also indicate no main effect of Experiment, such that infants exposed to visual cues (Experiment 1) did not benefit, compared to those who did not (Experiment 2). Thus, as infants' performance did not differ based on the visual information received in Experiment 1, in order to directly compare performance between Experiments 3 and 1, the procedure used in Experiment 1 was used in Experiment 3 as well.

Experiment 3

Experiment 3 (14-17 months) was conducted to (a) replicate our finding that infants younger than 18 months fail to discriminate small from large sets (in our standard condition), and (b) assess the impact of numerical language and perceptual variability on these discriminations in infants in this age group.

Methods

Participants

Forty six infants between 14- and 17-months participated in this study ($M=15.6$ mos, $SD=1.22$ mos, 26 female). Infants were run in one of three conditions, per

Experiment 1⁹: “Standard” ($N=20$), “Heterogeneity” ($N=13$), and “Language” ($N=13$).

Additionally, infants were randomly assigned to one of two counterbalanced orders: 2v4 first ($N=26$) or 2v3 first ($N=20$; counterbalanced within each condition). An additional 18 subjects completed half of the test trials (one of the two comparisons; Standard: $N=7$, Heterogeneity: $N=3$, Language: $N=8$; Secondary analysis participants: Standard: $N=27$, Heterogeneity: $N=16$, Language: $N=19$). These participants were excluded from primary analyses, but were included in secondary analyses (see Results below) examining infants’ performance on the first test comparison only. Lastly, 17 additional participants were excluded from the study due to failure to complete a sufficient number of test trials (at least one comparison, $N=3$), equipment error (video camera did not record; $N=4$), experimenter error ($N=3$), fussiness ($N=3$), or never searching over the course of the experiment ($N=4$). Children were recruited via phone or email for a single visit to the Infant and Child Cognition Lab or from the Boston Children’s Museum and Living Laboratory at the Museum of Science, Boston.

Stimuli, Procedure, and Data Coding

Because cross-experiment analyses following Experiment 2 (Standard Control, Feigenson & Carey, 2003, 2005) confirmed that infants’ performance did not differ based on the visual information received in Experiment 1, Experiment 3 was run using the same procedure as Experiment 1, allowing a direct comparison between performance on

⁹ Cross-experiment analyses following Experiment 2 (Standard Control, Feigenson & Carey, 2003, 2005) confirm that infants’ performance did not differ based on the visual information received in Experiment 1; thus, in order to directly compare performance between Experiments 3 and 1, the procedure used in Experiment 1 was used in the present Experiment.

Experiments 3 and 1. Thus, all procedures were identical to that of Experiment 1 with the exception that participants were between the ages of 14-17 months in Experiment 3.

After the MST task, infants were asked to count 2 and 4 items, as in Experiments 1 and 2; however, none of the infants in this age range did so; thus, this variable was not entered into any of the analyses in this experiment.

Reliability was calculated for each subject on a trial-by-trial basis as in Experiments 1 and 2 using the Preferential Looking Coder (Libertus, 2008). Two independent researchers -- blind to the Experimental Condition -- coded each video (thus, each participant's video was coded twice by two different experimenters). Inter-rater reliability for all trials across subjects was 91.9%. Any disagreements were settled by a third (independent) party coder for any participant where agreement was less than 80%.

Results

Difference scores were subjected to a Comparison (2: 2v3, 2v4) X Order (2v3 first, 2v4 first) X Condition (3: Standard, Heterogeneity, Language) mixed measures ANOVA. Results revealed a marginal main effect of Comparison ($F(1, 40)=3.02, p=.09, \eta_p^2=.07$), such that infants were more accurate in their searching for the 2v4 comparison ($M=1.48s$) than the 2v3 comparison ($M=.491s$). However, this was qualified by a significant Comparison X Order X Condition interaction ($F(1, 40)=3.6, p=.037, \eta_p^2=.152$).

To examine this interaction further, a preliminary post-hoc analysis of difference scores for each age group X comparison were calculated (versus 0; Figure 8). For the Standard condition, infants' difference scores did not significantly differ from zero in the 2v3 comparison ($M=-.095s, t(19)=-.263, p>.7, Cohen's d=.12$) or the 2v4 comparison

($M=.875s$, $t(19)=1.47$, $p>.1$, *Cohen's* $d=.40$). In the Heterogeneity condition, infants demonstrated marginal success in the 2 vs. 3 Comparison, but were successful when comparing 2v4 ($M_{2v3}=1.77s$, $t(12)=1.91$, $p=.081$, *Cohen's* $d=1.1$; $M_{2v4}=1.31s$, $t(12)=2.64$, $p=.022$, *Cohen's* $d=1.52$). In the Language condition, infants failed to compare either 2v3 or 2v4 ($M_{2v3}=.617s$, $t(12)=.858$, $p>.4$, *Cohen's* $d=.50$; $M_{2v4}=1.34$, $t(12)=1.48$, $p>.1$, *Cohen's* $d=.85$). However, our initial ANOVA suggested that Order was also important for infants' ability to discriminate sets successfully. Unsurprisingly, infants generally searched less on the second comparison that they saw, compared to the first comparison (as seen in previous manual search paradigms; e.g., Feigenson & Carey, 2003, 2005; Zosh & Feigenson, 2015; e.g., infants demonstrated the most success on the first comparison that they viewed). A secondary analysis including those infants who were initially excluded for failing to complete both test comparisons was run using the same statistics reported above. Results confirm these initial trends: Standard (2v3: $M=.153s$, $t(24)=.467$, $p>.6$, *Cohen's* $d=.19$; 2v4: $M=1.07s$, $t(22)=1.89$, $p=.072$, *approaching*, *Cohen's* $d=.81$); Heterogeneity (2v3: $M=1.32s$, $t(14)=1.5$, $p>.1$, *Cohen's* $d=.80$; 2v4: $M=1.44s$, $t(15)=3.29$, $p=.005$, *Cohen's* $d=1.7$), Language (2v3: $M=.97s$, $t(15)=1.5$, $p=.143$, *Cohen's* $d=.77$; 2v4: $M=.876s$, $t(17)=1.23$, $p>.2$, *Cohen's* $d=.60$).

We also assessed whether infants in this Experiment improved at this task with age. To examine this, we ran a correlational analysis between age (in days) and each difference score. Age (in days) was not significantly correlated with the 2v3 comparison ($r=.081$, $p>.5$), but it was correlated with the 2v4 comparison ($r=.326$, $p=.013$). This held for the Standard condition alone ($r=-.018$, $p>.9$; $r=.385$, $p=.069$); however, there were no significant correlations for either the Heterogeneity or Language conditions (all $p's>.2$).

This was confirmed by running a linear regression on the difference score for each comparison, with Age (months), Heterogeneity Condition, and Language Condition as predictors. The 2v3 Model was not a significant fit ($R^2=.043, p>.5$), with none of our predictors emerging as significant (Age: *Standardized Beta*=.02, $p>.8$; Heterogeneity: *Standardized Beta*=.204, $p>.1$; Language: *Standardized Beta*=.147, $p>.3$). However, the 2v4 Model was a marginally significant fit for the data ($R^2=.120, p=.077$). Age emerged as a significant predictor (*Standardized Beta* = .347, $t(53)=2.61, p=.012$); however, neither Heterogeneity (*Standardized Beta* = -.030, $p>.8$) nor Language were significant predictors (*Standardized Beta* = -.056, $p>.6$). See Table 7 and Figure 8 for difference scores vs. 0 for Comparison X Age (months).

Discussion

Experiment 3 asked two non-mutually exclusive questions: (a) whether infants younger than 18 months can discriminate small vs. large sets (i.e., success in the Standard condition), and (b) if additional factors may facilitate this understanding in infants younger than 18 months (i.e., success in either the Heterogeneity or Language conditions). Results indicate that infants younger than 18 months were able to compare small vs. large sets, but only when items were perceptually variable. Perceptual variability seemed to aid infants' ability to compare sets, with significant success in the 2v4 comparison. However, exposure to numerical language (counting by the experimenter in the Language condition) did not seem to impact discrimination abilities, and infants in this age range did not verbally count the number of items (i.e., two or four items).

In order to explore these remaining questions further, a cross-experiment analysis was conducted in order to analyze the contribution of Age Group (14-36 months) and

Condition (Standard, Heterogeneity, and Language) across development (Experiments 1, 2, and 3).

Cross-Experiment Analysis:

Cross-Experiment Analysis 1: Standard Condition (14-17 months): Experiments 2 and 3

In order to confirm when it is over the course of natural infant development that they can compare small and large sets, a Repeated Measures ANOVA was conducted examining the following variables: Comparison (2: 2v3, 2v4) X Age in Months (4: 14, 15, 16, 17 months) X Experiment (2: Experiment 2: Standard Control ($N=27$), Experiment 3: Standard Non-Control ($N=26$)). Results revealed no main effect of Experiment ($p>.9$, $\eta_p^2<.001$). Although there was a main effect of Comparison ($F(1, 33)=10.9$, $p=.002$, $\eta_p^2=.248$) and a main effect of Age Group ($F(3, 33)=8.02$, $p<.001$, $\eta_p^2=.422$), this was qualified by a significant Comparison X Age Group interaction ($F(1, 33)=5.7$, $p=.003$, $\eta_p^2=.342$; Figure 9). Follow-up analyses that this was driven by significant differences in searching behavior across the two comparisons in the oldest age group (14 mos: $M_{2v3}=-.269$ s, $M_{2v4}=.985$ s; $t(11)=1.64$, $p=.130$, *Cohen's d*=.99; 15 months: $M_{2v3}=.075$ s, $M_{2v4}=-.875$ s; $t(13)=1.84$, $p=.089$, *Cohen's d*=1.02; 16 months: $M_{2v3}=1.04$ s, $M_{2v4}=3.8$ s; $t(3)=2.21$, $p=.114$, *Cohen's d*=2.6; 17 months: $M_{2v3}=-.924$ s, $M_{2v4}=3.03$ s; $t(10)=3.22$, $p=.009$, *Cohen's d*=2.04). Moreover, the two oldest age groups successfully discriminated small from large sets, yet all four age groups failed to discriminate exclusively small sets (14 months: 2v3 vs. 0: $t(11)=.660$, $p>.5$, *Cohen's d*=.40; 2v4 vs. 0: $t(11)=1.66$, $p=.125$, *Cohen's d*=1.0; 15 months: 2v3 $t(17)=1.32$, $p=.206$, *Cohen's d*=.64; 2v4 vs. 0: $t(13)=2.25$, $p=.042$, *Cohen's d*=1.3; 16 months: 2v3 vs. 0: $t(5)=1.33$, $p=.241$, *Cohen's d*=1.2; 2v4 vs.

0: $t(3)=4.54, p=.02, \text{Cohen's } d=5.2$; 17 months: 2v3 vs. 0: $t(12)=.359, p>.7, \text{Cohen's } d=.21$; 2v4 vs. 0: $t(14)=3.46, p=.004, \text{Cohen's } d=1.8$).

Thus, the present cross-experiment analysis suggests that although infants in Experiment 1 and 2 demonstrated robust success in the Standard condition as young as 18 months of age, initial success comparing small and large sets can be seen in 16- and 17-month old infants, whether or not visual cues are present (i.e., no main effect of Experiment). Thus, our data provide strong evidence to suggest that infants may be able to successfully compare small and large sets successfully before 18 months of age.

Cross Experiment Analysis 2: Effects of Condition and Counting Behavior across Development: Experiments 1 and 3

To confirm and explore interim conclusions from Experiments 1 and 3, an omnibus Mixed Measures ANOVA was conducted to examine Comparison (2: 2v3, 2v4) X Age Group (4: 14-17mos ($N=46$), 18-23 mos ($N=43$), 24-29 mos ($N=57$), 30-36 mos ($N=70$)) X Condition (3: Standard ($N=78$), Heterogeneity ($N=76$), Language($N=62$)). Analyses revealed no main effect of Comparison ($p>.1, \eta_p^2 =.01$), such that infants in this version of the manual search task did not search longer in the 2v4 comparison ($M=2.45s$) than the 2v3 comparison ($M=2.1s$). Both of these means differed significantly from zero, indicating success on both of these comparisons, generally (2v3: $t(215)=12.2, p<.001, \text{Cohen's } d=1.7$; 2v4: $t(215)=13.7, p<.001, \text{Cohen's } d=1.9$).

There was a main effect of Age Group ($F(3,204)=19.7, p<.001, \eta_p^2 =.225$), with searching for each age group significantly differing from each other age group ($M_{14-17mos}=.970s, M_{18-23mos}=1.68s, M_{24-29mos}=2.92s, M_{30-36mos}=3.52s$; all $p's<.1$). See Figures 10 and 11 for correlations between age (in days) with each set of difference scores (2v3:

$r=.377, p<.001$; 2v4: $r=.344, p<.001$). Finally, there was a main effect of Condition ($F(2, 204)=3.88, p=.022, \eta_p^2=.037$), such that infants searched longest in the Heterogeneity condition ($M=2.77s$), followed by the Language condition ($M=2.12s$) and Standard condition ($M=1.93s$). Searching in the Heterogeneity condition differed significantly from the Standard condition ($p=.008$); however, Heterogeneity and Language only differed marginally ($p=.054$). Standard and Language did not differ from each other ($p>.5$). Furthermore, Condition did not interact with Comparison ($p>.6, \eta_p^2=.004$).

A secondary omnibus Mixed Measures ANOVA was run with the inclusion of Counting Behavior as a between-subjects variable. With the inclusion of Counting Behavior, there was no longer a main effect of Condition ($p>.1, \eta_p^2=.023$). However, there was still a main effect of Age Group ($F(3, 194)=9.79, p<.001, \eta_p^2=.132$). Of note, in our omnibus ANOVA, there was no main effect of Counting Behavior ($p>.6, \eta_p^2=.001$). Additionally, none of the variables of interest (Condition, Comparison, Counting Behavior, or Age) produced any significant interactions ($p's>.4$). Thus, to examine the relative impact of these variables, they were subjected to a linear regression. The 2v3 Model was a significant fit for the data ($R^2=.173, p<.001$), with Age (Group) significantly predicting difference scores (*Standardized Beta* = .408, $t(229)=6.2, p<.001$) and Heterogeneity marginally predicting search times (*Standardized Beta* = .125, $t(229)=1.84, p=.067$; Counting Behavior: *Standardized Beta* = -.045, $p>.5$; Language: *Standardized Beta* = .052, $p>.4$). The 2v4 Model was also a significant fit for the data ($R^2=.141, p<.001$), with Age (Group) singularly predicting difference scores (*Standardized Beta* = .300, $t(231)=4.44, p<.001$; Heterogeneity: *Standardized Beta* = .082, $p>.2$; Language: *Standardized Beta* = -.027, $p>.6$; Counting Behavior: *Standardized Beta* = .105, $p>.1$).

See Table 8 for a breakdown of Age Group by Comparison (difference scores vs. 0) for the Standard Condition, which indicates that infants successfully discriminated small from large sets as early as 14-17 months of age, although this was not the case for small vs. small sets until 18-23 months of age. .

General Discussion

These studies addressed two research questions: (a) when it is over the course of natural development that infants overcome the documented inability to compare small vs. large sets, and (b) what factors may facilitate this understanding. These questions yielded several specific predictions: (1) Infants will overcome this inability to compare small and large sets over the course of natural development between 18 and 36 months of age; namely, at approximately two years of age, when numerical language also appears (e.g., Wynn, 1990, 1992). (2) Numerical language may provide a third, integrated system on which infants can rely when discriminating across sets. Numerical language could impact discrimination in two distinct ways: via receptive language, in which case simply exposure to numerical language (and, specifically, verbal counting and cardinality), may facilitate these discriminations, and/or via expressive abilities, such that infant abilities to produce count words may predict discrimination abilities. (3) Perceptual variability may provide infants with the ability to represent items individually within a set, thus momentarily expanding the limits of their visual working memory. In this vein, infants could rely on the object file system *exclusively* to represent individual items within a set at a given time (e.g., 4 items instead of 3 or fewer items), thereby taxing visual working memory to a lesser degree than a strictly homogenous set of items.

The main finding from these experiments indicates that children older than 18 months succeeded in comparing both small sets (2 vs. 3, as previously documented: Feigenson & Carey, 2003, 2005; Xu, 2003; Xu & Spelke, 2000) *and* small vs. large sets (2 vs. 4). This is the first study to show this ability, spanning a developmental age range and using the same experimental paradigm. Previous work indicates that infants fail to compare small and large sets (Cordes & Brannon, 2009; Feigenson et al., 2002; Xu, 2003; Xu & Spelke, 2000) and the dominant theory suggests that this failure is due to infants' use of two incompatible systems for representing number (Cordes & Brannon, 2009; Feigenson et al., 2002): an object file system used to precisely represent visual sets in working memory and a noisy, approximate analog magnitude system used to represent all natural numbers (see Posid & Cordes, 2014, for a review). The earliest documented evidence of successful discrimination across this small-large divide appears at 3-years-old (Cantlon et al., 2010); however, findings from the current experiments provide the first evidence to suggest that infants younger than 36 months can compare sets spanning the small-large boundary. In fact, data from Experiments 1 and 2 indicate that this ability can be seen as young as 18-23 months of age, with this young age group successfully discriminating 2v3 and 2v4 in both standard conditions (Experiment 1 and Experiment 2: Standard Control). In support of this conclusion is the finding that infants searched more with age (longer relative searching on Expected Full trials relative to similar searching seen across Expected Empty trials across development, as indicated by higher and more positive difference scores) in Experiments 1 and 2 (Standard conditions). Thus, it appears that older infants, more confident in their numerical representations, were more inclined to believe that objects remained in the box and thus continued to search longer for them

compared to younger infants. In fact, this main effect of Age (Group) and its prevalence throughout each Experiment in the present study as a significant predictor of difference scores in both of our comparisons, indicates that infants continue to gain confidence in their numerical representations with maturation and experience. Specifically, if the present data indicated that 18-23 month olds could successfully compare small and large sets, but did not reveal Age effects in the remaining older age groups, this would be indicative of an “all or nothing” understanding, where infants either understood and represented the number of items which they were asked to enumerate, or did not. In contrast, infants continued to search longer with age on Expected Full trials in particular – but not Expected Empty trials – revealing that this representation of the number of items in the hidden becomes, presumably, more concrete with development. Thus, demonstrating a novel developmental trend, the present findings first extend the relevant literature and bridge the gap between documented early failure to compare small-large sets at 14 months of age (Feigenson & Carey, 2003, 2005) to a late success at 3 years of age (Cantlon et al., 2007).

Due to the unexpected and robust success of infants in the youngest age range of Experiment 1, Experiments 2 and 3 tested infants from 14-17 months of age, since Feigenson and Carey (2003, 2005) indicate a failure to compare small-large sets at 14-months. Initial results indicated that this age group, generally, failed to compare small vs. large sets; however, a closer examination of data collapsed across Experiments 2 and 3 for this youngest age range reveals that 16-17 month old infants have begun to succeed at this difficult discrimination. Thus, although some clear individual variability exists,

results indicate that infants between 16-17 months old may begin to cross this small-large divide, with robust success evident in the 18-23 month old sample.

Why might 18-24 month old infants succeeded in our 2v4 condition when Barner et al (2007) reported a failure in this comparison as late as 22-24 months of age? One potential explanation in support of this theory is that slight methodological differences between the present task and previous versions may have contributed to these differences (e.g., Barner et al., 2007). Although the task was modeled after previous studies, one minor difference between the current study and that of previous studies is that after presenting the items to the infant on top of the box, previous studies have used a single motion to place all items in the box simultaneously (Barner et al., 2007). In contrast, items were placed individually into the box in the current task, thus perhaps providing additional confounding information regarding the numerosity of the set. Thus, perhaps this sequential placing of items on and into the box allowed infants to encode the information as separate entities rather than a set or single group of items.

Relatedly, another question that stems from the results of the present study is why 14-17 month old infants show some success in our task, but fail to compare small-large sets in manual search paradigms used previously (Barner et al., 2007; Feigenson & Carey, 2003, 2005). One possibility is that infants could use visual cues in our version of the manual search task, whereas retrieved items were hidden from sight in the Standard Control version of the task. However, this does not seem to be the case, as there was no main effect of Experiment, nor any interactions with Experiment, following participation in our Standard Control experiment. In this case, although redundant visual cues were not important for infants over 18 months of age, these cues may have provided redundant

perceptual information for which 14-17 month olds benefited (as seen in the Experiment X Comparison interaction in the cross-experiment analysis of the standard conditions following Experiment 3).

Results across all three of our present Experiments lend themselves to the theory that the ANS increases in precision with development. Specifically, our results indicate that even infants in our youngest age range (14-17 months) successfully compared small and large sets (2v4); however, of interest, they failed to compare small and small sets (2v3). Why might this be the case? Evidence suggests that perhaps they were able to represent the items in the sets to-be-enumerated via the ANS. The consistent finding that Age Group predicts difference scores across experiments suggests that true success on this task comes with the natural and increasing precision of the ANS. Support for this notion comes in two forms. First, much evidence indicates that the ANS becomes more precise over natural development, such that 6-month-olds can discriminate a 1:2 ratio, 9-month-olds can discriminate a 2:3 ratio, and so on, with adults discriminating a 7:8 ratio accurately (Halberda & Feigenson, 2008; Lipton & Spelke, 2004). The largest gains in this increasing precision can be seen in these infant years. Thus, although infants early in development demonstrate an incompatibility when representing sets that span the small-large divide, this ANS precision increases rapidly during this time frame, and is also subject to individual variability across the lifespan (e.g., Halberda & Feigenson, 2008). It stands to reason, then, that perhaps infants in our task successfully compared small and large sets exclusively through reliance on the ANS, as adults and children can do with ease (for a review: Posid & Cordes, 2014). This theory gains further support by the fact that 14-17 month olds continue to fail in their ability to compare sets of 2 and 3 items in

our version of the manual search task. Presumably, if they were either (a) relying exclusively on the object file system or (b) attempting to represent smaller numbers via the object file system and larger numbers via the ANS, they would have demonstrated the opposite pattern of results that we find; that is, they should succeed at a 2v3 comparison earlier in development than 2v4. However, we find the opposite pattern: success at 2v4 earlier than 2v3. Although seemingly irreconcilable with previous reports (e.g., with Feigenson & Carey, 2003, 2005), this pattern of findings instead provides *novel* insight into the developmental trajectory of infants' ability to track sets with age. Specifically, although early literature from infant numerical tasks suggests that they fail to compare sets when invoking two incompatible systems, the present study suggests that the nature of the task (e.g., the context) prompts the use of this single system (ANS) even earlier in development than previously thought. The specific mechanism by which this type of task may engage the ANS (and not object files) is unclear; however, one reason for this proclivity of infants to use the ANS may stem from the non-visual nature of this task. Although infants see a certain quantity of items hidden in the box, the Standard Control experiment indicates that infants need not rely on these visual cues (no main effect of Experiment, nor any interactions). Thus, the present study gains support from a select number of studies in which infants make difficult small-small or small-large discriminations in the non-visual domain (i.e., sounds; vanMarle & Wynn, 2009). Importantly, in these non-visual tasks, infants also demonstrate a ratio-dependence, such that they can compare sounds of a 1:2 ratio, but not a 2:3 ratio. Therefore, in line with these findings, our data suggest that, in contrast to other small-large discrimination failures with visual sets, ratio-dependent numerical discriminations involving non-visual

sets strongly suggest that infants, like older children, adults, and non-human animals have access to ANS representations all the way down the number line (see Posid & Cordes, 2014).

Second, several recent lines of research suggest that training the ANS can lead to increased ANS precision, even in adulthood where a plateau in discrimination ability is observed (DeWind & Brannon). Specifically, DeWind & Brannon addressed the malleability of the ANS by providing trial-by-trial feedback to adults making numerical discriminations, with participants showing rapid improvements (measured by a greater precision in their ANS) following this feedback. In fact although this was initially observed for accuracy (which later plateaued), reaction time for numerical discriminations continued to decrease over time (DeWind & Brannon, 2012). In line with results from the present study, the authors suggest that children who have not yet reached asymptotic performance in ANS acuity might benefit more from this type of training, compared to adults who should have a pretty precise ANS already. Although more work is needed to explore this theory, it provides a possible explanation for the current pattern of findings, thus highlighting one potential mechanism for success on this small-large discrimination in toddlerhood.

To address our secondary research question, we asked what factors could facilitate infants' understanding of small vs. large sets across this same age range, namely, numerical language and perceptual variability. Of note, neither of these conditions produced *worse* performance, compared to the Standard condition for either numerical comparison across any observed age range. Therefore, our discussion will focus on the potential benefits of these variables of interest.

We first investigated the impact of numerical language via two mechanisms: production and exposure. Work from Barner and colleagues (Barner et al., 2007; Li et al., 2009) suggest that infants who successfully compare small vs. large sets (1 vs. 4) may be those infants who are at the same developmental stage as those who have started to produce quantity-related words in their own vocabulary. Although the acquisition of the singular/plural distinction may account for success in 1 vs 4 comparisons, this is clearly not the whole story. Thus, we examined whether or not infants could count a set of items (2 and 4) and whether this ability impacted their understanding of small vs. large sets. Results from Experiment 1 indicate that those infants who could produce count words were slightly more advanced than those who could not; however, this finding did not hold in any other analyses from Experiments 2 and 3, nor in our Cross-Experiment analyses. This could be due to the fact that 18-36 month olds were successful regardless of number word production or because 14-17 month olds rarely (if ever) engaged in overt counting of sets in our task. Thus, we cannot speculate about whether younger infants, who might potentially benefit from this ability, would show improved performance in our test comparison. Therefore, these preliminary findings do not support an account of linguistically-mediated small-large discriminations, but instead suggest that infants succeed at these difficult discriminations with development, such that numerical language does not play a specific or necessary role in this ability.

Our secondary variable of interest was perceptual variability. Recent research suggests that, under certain circumstances, perceptual variability aids exact numerical judgments, such as counting items (adults: Frick, 1987), increasing speed of enumeration (adults: Trick, 2008), improving discrimination abilities (children: Posid et al., in

preparation), and facilitating discrimination during infancy (Feigenson, 2005). Results from these studies suggest that perceptual variability may aid in infants', children's, and adults' ability to discriminate between sets because heterogeneity highlights the salience of the individual items to be enumerated (Feigenson, 2005; Posid et al., in preparation). In the present study, we explored whether a heterogeneous array would increase success in a manual search task by improving infants' individuation of objects, thus potentially expanding their visual working memory to hold four *distinct* items in working memory. Data from our experiments indicated that heterogeneity increased infants' confidence, as evident by higher search times in the Heterogeneity condition, compared to the Standard condition, in Experiment 1 and (marginally) in Experiment 3. Although infants successfully discriminated both exclusively small and small-large sets in Experiment 1, the increased search time for the Heterogeneity condition suggests that infants searched longer on Expected Full trials compared to Expected Empty trials specifically in this condition. Support for this theory also emerges when examining the 14-17 month olds' success in Experiment 3. Specifically, although infants were nearing successful comparisons across this small-large boundary in the youngest age group, the only robust discrimination was in the Heterogeneity condition. This pattern indicates that, although perceptual variability was not necessary for success, it promoted infants' remembering of the items hidden in the box. Support for this theory comes from recent work by Zosh and Feigenson (2015), who found that arrays that contrast in perceptual features increased the storage capacity of visual working memory in infants, as seen in adulthood. Results from Experiment 3 support this claim, such that 14-17 month olds in the Heterogeneity condition (Experiment 3) also successfully compare sets of two and four items. In line

with conclusions from Zosh and Feigenson (2015), this suggests that perceptual variability in the make up of sets to-be-enumerated increases the storage capacity of visual working memory, such that these young infants can rely on object files to compare sets successfully.

In conclusion, the present study addressed the open questions of (1) when, over the course of development, do we overcome the observed inability to compare small vs. large sets posed by representation incompatibilities, and (2) what circumstances or factors facilitate this ability. Results from three cross-sectional studies indicate that infants begin to discriminate between small and large sets as early as 16-17 months old, with robust ability to overcome this failure by 18 months of age. Furthermore, infants seemed to benefit from perceptual variability in the form of item individuation in their ability to make this discrimination. In contrast, neither exposure to numerical language (in the form of adult modeling of cardinality and counting) nor the ability to produce numerical language (i.e., counting behavior) led to greater confidence or proficiency when comparing sets. Future research should continue to examine the proposed potential mechanisms for overcoming small-large discrimination failures, including infants' reliance on exclusively object files or exclusively analog magnitudes. Some research to date has examined circumstances under which each of these can be implicated, but more work is needed, as well as a more thorough investigation of the longitudinal or broader effects of implicating one of these systems over the other across the lifespan when making numerical judgments.

References

- Alvarez, G.A. & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106-111.
- Alvarez, G.A. & Franconeri, S.L. (2007). How many objects can you track? Evidence for a resource-limited tracking mechanism. *Journal of Vision*, *7*, 1-10.
- Ansari, D., Lyons, I., van Eimeren, L., & Xu, F. (2007). The role of the temporo-parietal junction in small and large number processing: an fMRI study. *Journal of Cognitive Neuroscience*, *19*, 1845-1853.
- Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, *54*(3), 695-701.
- Awh, E., Vogel, E.K., & Oh, S.H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*, 201-208.
- Balakrishnan, J.D., & Ashby, F.G. (1992). Subitizing: Magical numbers or mere superstition? *Psychological Research*, *54*(2), 80-90.
- Barner, D., Thalwitz, D., Wood, J., Yang, S., & Carey, S. (2007). On the relation between the acquisition of singular-plural morpho-syntax and the conceptual distinction between one and more than one. *Developmental Science*, *10*(3), 365-373.
- Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*, 199-222.
- Bisazza, A., Piffer, L., Serena, G., & Agrillo, C. (2010). Ontogeny of numerical abilities in fish. *Plos One*, *5*, 1-9.

- Bijeljacabic, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances. *Developmental Psychology*, 29(4), 711-721.
- Bonny, J.W., & Lourenco, S. F. (2013). The approximate number system and its relation to early Math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology*, 114, 375-388.
- Brannon, E.M., Abbot, S., & Lutz, D.J. (2004). Number bias for the discrimination of large visual sets in infancy. *Cognition*, 93, B59-B68.
- Brannon, E.M., & Terrace, H.S. (1998). Ordering of the numerosities 1-9 by monkeys. *Science*, 282, 746-749.
- Cantlon, J.F., & Brannon, E.M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, 17, 401-406.
- Cantlon, J., Fink, R., Safford, K., & Brannon, E. M. (2007). Heterogeneity impairs numerical matching but not numerical ordering in preschool children. *Developmental Science*, 10(4), 431-440.
- Cantlon, J.F., Safford, K.E., & Brannon, E.M. (2010). Spontaneous analog number representations in 3-year-old children. *Developmental Science*, 13(2), 289-297.
- Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition. Special Issue: Objects and attention*, 80, 179-213.
- Condry, K.F. & Spelke, E.S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, 137, 22-38.

- Cordes, S., & Brannon, E. M. (2008). The difficulties of representing continuous extent in infancy: Using number is just easier. *Child Development, 79*(2), 476-489.
- Cordes, S. & Brannon, E.M. (2008). Quantitative competencies in infancy. *Developmental Science, 11*, 803-808.
- Cordes, S. & Brannon, E. M. (2009). Crossing the divide: infants discriminate small from large numerosities. *Developmental Psychology, 45*(6), 1583-1594.
- Cordes, S., & Brannon, E. M. (2009). The relative salience of discrete and continuous quantity in young infants. *Developmental Science, 12*, 453-463.
- Cordes, S., & Gelman, R. (2005). The young numerical mind: When does it count? In J. Campell (Ed.), *Handbook of mathematical cognition* (pp. 127-142). London: Psychology Press.
- Cordes, S., Gelman, R., Gallistel, C.R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review, 8*, 698-707.
- Cordes, S., & Posid, T. (in progress). 7-month-olds' discrimination of difficult sets under heterogeneous, but not homogenous, circumstances.
- Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition, 123*(1), 162-173.
- De Smedt, B., & Boets, B. (2010). Phonological processing and arithmetic fact retrieval: Evidence from developmental dyslexia. *Neuropsychologia, 48*, 3973-3981.
- De Smedt, B., Noel, M, Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior.

- Trends in Neuroscience and Education*,
<http://dx.doi.org/10.1016/j.tine.2013.06.001>
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44, 1-42.
- Dehaene, S., & Cohen, L. (1992). Levels of representation in number processing. In B. Stemmar And H.A. Whitaker, eds., *The handbook of neurolinguistics*, pp. 331-341.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122, 371-396.
- Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, 1, 83-120.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.
- DeWind, N.K., & Brannon, E.M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience*, 6(68), 1-10.
- Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428-1446.
- Feigenson, L. (2005). A double-dissociation in infants' representation of object arrays. *Cognition*, 95, B37-B48.
- Feigenson, L. (2008). Conceptual knowledge increases infants' memory capacity.
- Feigenson, L. & Carey, S. (2003). Tracking individuals via object-files: Evidence from

- infants' manual search. *Developmental Science*, 6, 568–584.
- Feigenson, L. & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, 97, 295–313.
- Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files vs. analog magnitudes. *Psychological Science*, 13(2), 150–156.
- Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314.
- Feigenson, L. & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition*, 91, 173-190.
- Feigenson, L., & Yamaguchi, M. (2009). Beyond “What” and “How many”: Capacity, complexity, and resolution of infants' object representations. In *The Origins of Object Knowledge*. Laurie Santos and Bruce Hoods (Eds.), Oxford University Press.
- Frick, R. W. (1987). The homogeneity effect in counting. *Perception & Psychophysics*, 41(1), 8-16.
- Gallistel, C.R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44, 43-74.
- Geary, D.C. (2011). Cognitive predictors of individual differences in achievement growth in mathematics: A five year longitudinal study. *Developmental Psychology*, 47, 1539-1552.
- Geary, D.C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science*, 22(1), 23-27.

- Geary, D.C., Bow-Thomas, C.C., Liu, F., & Siegler, R.S. (1996). Development of arithmetic competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development, 67*, 2022-2044
- Geary, D.C., Hoard, M.K., Nugent, L., & Bailey, D.H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS One, 8*(1), e54651.
- Gelman, R., & Gallistel, C.R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Halberda, J. & Feigenson, L. (2008). Developmental change in the acuity of the "Number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*, 1457-1465.
- Holmes, V.M., & McGregor, J. (2007). Rote memory and arithmetic fact processing. *Memory & Cognition, 35*(8), 2041-2051.
- Hyde, D.C. & Spelke, E.S. (2009). All numbers are not equal: An electrophysiological investigation of small and large number representations. *Journal of Cognitive Neuroscience, 21*, 1039-1053.
- 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.
- Hyde, D.C. & Spelke, E.S. (2012). Spatiotemporal dynamics of processing nonsymbolic number: An event-related potential source localization study. *Human Brain Mapping, 33*, 2189-2203.
- Hyde, D.C., & Wood, J.N. (2011). Spatial attention determines the nature of nonverbal

- number representation. *Journal of Cognitive Neuroscience*, 23, 2336-2351.
- Izard, V., Sann, C., Spelke, E.S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, 106, 10382-10385.
- Jordan, K.E., Suanda, S.H., & Brannon, E.M. (2008). Intersensory redundancy accelerates preverbal numerical competence. *Cognition*, 108, 210-221.
- Kaldy, Z., & Leslie, A.M. (2003). Identification of objects in 9-month-old infants: Integrating 'what' and 'where' information. *Developmental Science*, 6(3), 360-373.
- Klein, E., Suchan, J., Moeller, K., Karnath, H., Knops, A., Wood, G., Nuerk, H., & Willmes, K. (2014). Considering structural connectivity in the triple code model of numerical cognition: Differential connectivity for magnitude processing and arithmetic facts. *Brain Structure and Function*, doi:10.1007/s00429-014-0951-1.
- Kobayashi, T., Hiraki, K., & Hasegawa, T. (2005). Auditory-visual intermodal matching of small numerosities in 6-month-old infants. *Developmental Science*, 8, 409-419.
- Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105, 395-438.
- Leslie, A.M., Xu, F., Tremoulet, P.D., & Scholl, B.J. (1998). Indexing and the object concept: Developing 'what' and 'where' systems. *Trends in Cognitive Sciences*, 2, 10-18.
- Leybaert, J., & Van Cuyck, M. (2002). Counting in sign language. *Journal of Experimental Child Psychology*, 81(4), 482-501.

- Li, P., Ogura, T., Barner, D., Yang, S., & Carey, S. (2009). Does the conceptual distinction between singular and plural sets depend on language? *Developmental Psychology, 45*, 1644-1653.
- Libertus, K. (2008). Preferential Looking Coder (Version 1.0) [Computer software]. Durham, NC: Duke University. Retrieved August 5, 2008. Available from <http://www.duke.edu/~kl41>.
- Libertus, M.E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*, 1292-1300.
- Libertus, M.E., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences, 25*, 126-133.
- Libertus, M.E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica, 141*(3), 373-379.
- Lipton, J.S. & Spelke, E.S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science, 14*, 396-401.
- Lipton, J. S. & Spelke, E. S. (2004). Discrimination of large and small numerosities by human infants. *Infancy, 5*(3), 271-290.
- Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.
- Meck, W.H., & Church, R.M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9*, 320-334.

- Miller, K.F., & Stigler, J.W. (1987). Counting in Chinese: Cultural variation in a basic cognitive skill. *Cognitive Development*, 2, 279-305.
- Mix, K.S., Sandhofer, C.M., Moore, J.A., & Russell, C. (2012). Acquisition of the cardinal word principle: The role of input. *Early Childhood Research Quarterly*, 27(2), 274-283.
- Mou, Y., & vanMarle, K. (2013). Two core systems of numerical representation in infants. *Developmental Review*, <http://dx.doi.org/10.1016/j.dr.2013.11.001>.
- Piazza, M., Giacomini, E., Le Bihan, D., & Dehaene, S. (2003). Single-trial classification of parallel pre-attentive and serial attentive processes using functional magnetic resonance imaging. *Proceedings of the Royal Society of London. Serial B: Biological Sciences*, 270(1521), 1237-1245.
- Piazza, M., Mechelli, A., Butterworth, B., & Price, C.J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes? *Neuroimage*, 15(2), 435-446.
- Piffer, L., Miletto, Petrazzini, M.E., & Agrillo, C. (2013). Large number discrimination in newborn fish. *PLoS ONE*, 8(4), e62466.
- Posid, T., & Cordes, S. (under review). How high can you count? Probing the limits of children's counting. *Cognition*.
- Posid, T. & Cordes, S. (2014). The small-large divide: A case of incompatible numerical representations in infancy. In D. Geary, D. Berch, & K. Mann-Koepke (Eds.), *Evolutionary Origins and Early Development of Basic Number Processing*.

- Posid, T., Huguenel, B., & Cordes, S. (in preparation). Heterogeneity facilitates numerical abstraction in preschoolers.
- Ross-Sheehy, S., Oakes, L.M., & Luck, S.J. (2001). Visual short-term memory in the first year of life: Capacity and recency effects. *Developmental Psychology, 37*, 539-549.
- Schmithorst, V.J., & Douglas Brown, R. (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.
- Scholl, B.J., & Pylyshyn, Z.W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38*, 259-290.
- Simon, T.J. (1997). Reconceptualizing the origins of number knowledge: A non-numerical account. *Cognitive Development, 12*, 349-372.
- Spelke, E.S., & Tsivkin, S. (2001). Language and number: a bilingual training study. *Cognition, 78*(1), 45-88.
- Starr, A., Libertus, M.E., & Brannon, E.M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *PNAS, 110*(45), 18116-18120.
- Thevenot, C., & Barrouillet, P. (2010). Children's number processing is context dependent. *European Journal of Cognitive Psychology, 22*:3, 348-359.
- Tremoulet, P.D., Leslie, A.M., & Hall, D.G. (2000). Infant individuation and identification of objects. *Cognitive Development, 15*, 499-522.
- Trick, L. M. (2008). More than superstition: Differential effects of featural heterogeneity and change on subitizing and counting. *Perception &*

- Psychophysics*, 70(5), 743 – 760.
- Trick, L.M., Enns, J.T., & Brodeur, D.A. (1996). Life-span changes in visual enumeration: The number discrimination task. *Developmental Psychology*, 32(5), 925-932.
- Trick, L.M., & Pylyshyn, Z.W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2), 331-351.
- Trick, L.M., & Pylyshyn, Z.W. (1994). Why are small and large numbers enumerated differently: A limited capacity preattentive stage in vision. *Psychological Review*, 101(1), 80-102.
- Uller, C., Carey, S., Huntley-Fenner, G., & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development*, 1(3), 249-280.
- Van de Walle, G., Carey, S., & Prevor, M. (2000). Bases for object individuation in infancy: Evidence from manual search. *Journal of Cognition and Development*, 1, 249–280.
- vanMarle, K. (2013). Infants use different mechanisms to make small and large number ordinal judgments. *Journal of Experimental Child Psychology*, 114, 102-110.
- vanMarle, K. & Wynn, K. (2009). Infants' auditory enumeration: Evidence for analog magnitudes in the small number range. *Cognition*, 111, 302-316.
- vanMarle, K. & Wynn, K. (2011). Tracking and quantifying objects and non-cohesive substances. *Developmental Science*, 14(3), 502-515.
- 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, 92-114.
- Wang, Y., Lin, L., Kuhl, P., & Hirsch, J. (2007). Mathematical and linguistic processing differences between native and second languages: An fMRI study. *Brain Imaging and Behavior*, 1, 68-82.
- Whalen, J., Gallistel, C.R., & Gelman, R. (1999). Non-verbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10, 130-137.
- Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition*, 72, 125-166.
- Wilson, A.J., Dehaene, S., Pinel, P., Revkin, S.K., Cohen, L., & Cohen, D. (2006). Principles Underlying the design of "The Number Race," an adaptive computer game for Remediation of dyscalculia. *Behavioral and Brain Functions*, 2(19), doi:10.1186/1744-9081-2-1.
- Wood, J. N. & Spelke, E. S. (2005). Infants' enumeration of actions: Numerical discrimination and its signature limits. *Developmental Science*, 8, 173–181.
- Wylie, J., Jordan, J., & Mulhern, G. (2012). Strategic development in exact calculation: Group and individual differences in four achievement subtypes. *Journal of Experimental Child Psychology*, 113, 112-130.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36, 155-193.
- Wynn, K. (1992). Children's acquisition of number words and the counting system. *Cognitive Psychology*, 24, 220-251.
- Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, 89, B15-B25.
- Xu, F. & Arriaga, R.I. (2007). Number discrimination in 10-month-old infants. *British*

- Journal of Developmental Psychology, 25, 103-108.
- Xu, F., Carey, S., & Welch, J. (1999). Infants' ability to use object kind information for object individuation. *Cognition*, 70, 137-166.
- Xu, F. & Spelke, E. S. (2000). Large number discrimination in 6-month old infants. *Cognition*, 74, B1–B11.
- Xu, F., Spelke, E.S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, 8, 88-101.
- Zosh, J.M., Halberda, J., & Feigenson, L. (2011). Memory for multiple visual ensembles in infancy. *Journal of Experimental Psychology: General*, 140, 141-158.
- Zosh, J.M., & Feigenson, L. (2015). Array heterogeneity prevents catastrophic forgetting in infants. *Cognition*, 136, 365-380.
- Zosh, J.M., & Feigenson, L. (2012). Memory load affects object individuation in 18-month old infants. *Journal of Experimental Child Psychology*, 113, 322-336.
- Zosh, J.M., & Feigenson, L. (2015). Array heterogeneity prevents catastrophic forgetting in infants. *Cognition*, 136, 365-380.
-

Table 1.

Age Group Condition	Standard Condition	Heterogeneity Condition	Language
18-23 mos	N=14	N=18	N=11
24-29 mos	N=20	N=19	N=18
30-36 mos	N=24	N=26	N=20

Experiment 1: The number of children per Condition by Age Group.

Table 2.

Age Group	Standard	Heterogeneity	Language
18-23 months	$M=1.11s$ $t(14)=2.85, p=.013$ $Cohen's d=1.45$	$M=2.22s$ $t(17)=3.86, p<.001$ $Cohen's d=1.87$	$M=1.47s$ $t(10)=2.85, p=.017$ $Cohen's d=1.8$
24-29 months	$M=1.99s$ $t(20)=3.28, p=.004$ $Cohen's d=1.47$	$M=2.96s$ $t(19)=4.38, p<.001$ $Cohen's d=2.01$	$M=2.77s$ $t(17)=5.4, p<.001$ $Cohen's d=2.62$
30-36 months	$M=3.58s$ $t(23)=5.91, p<.001$ $Cohen's d=2.46$	$M=3.67s$ $t(26)=7.52, p<.001$ $Cohen's d=3.25$	$M=3.23s$ $t(19)=5.36, p<.001$ $Cohen's d=2.46$

Experiment 1: Difference scores (2v3 Comparison) vs. 0.

Table 3.

Age Group	Standard	Heterogeneity	Language
18-23 months	$M=1.5s$ $t(14)=1.94, p=.073$ $Cohen's d=1.0$	$M=3.31s$ $t(17)=6.3, p<.001$ $Cohen's d=3.06$	$M=.796s$ $t(10)=1.21, p>.2$ $Cohen's d=.77$
24-29 months	$M=3.17s$ $t(20)=5.37, p<.001$ $Cohen's d=2.4$	$M=2.97s$ $t(19)=5.65, p<.001$ $Cohen's d=2.59$	$M=3.64s$ $t(17)=6.75, p<.001$ $Cohen's d=3.27$
30-36 months	$M=3.62s$ $t(23)=5.36, p<.001$	$M=3.9s$ $t(26)=8.29, p<.001$	$M=3.09s$ $t(19)=4.24, p<.001$

*Cohen's d=2.23**Cohen's d=3.25**Cohen's d=1.9*

Experiment 1: Difference scores (2v4 Comparison) vs. 0.

Table 4.

Condition	Non-Counters	Counters
Standard	62.5%	37.5%
Heterogeneity	62.7%	37.3%
Language	62.5%	37.5%

Experiment 1: Counters X Condition.

Table 5.

Age Group	Non-Counters	Counters
18-23 months	76.6%	23.4%
24-29 months	71.2%	28.8%
30-36 months	45.9%	54.1%

Experiment 1: Counters X Age Group.

Table 6.

Age Group	Non-Counters	Counters
14-17 months	88.9%	11.1%
18-23 months	94.7%	5.3%
24-29 months	66.7%	33.3%
30-36 months	33.3%	66.7%

Experiment 2: Standard Control. Counters X Age Group.

Table 7.

Age Group	2v3	2v4
14 months	<i>M</i> =.843s	<i>M</i> =.267s

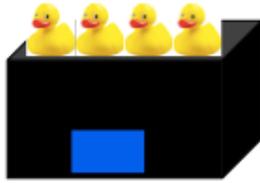
	$t(10)=1.52, p>.1$ <i>Cohen's d</i> =.96	$t(11)=.538, p>.6$ <i>Cohen's d</i> =.33
15 months	$M=.233s$ $t(12)=.968, p>.3$ <i>Cohen's d</i> =1.47	$M=-.06s$ $t(11)=-.102, p>.9$ <i>Cohen's d</i> =.062
16 months	$M=.263s$ $t(9)=.982, p>.3$ <i>Cohen's d</i> =.65	$M=1.11s$ $t(10)=1.14, p>.2$ <i>Cohen's d</i> =.72
17 months	$M=1.06s$ $t(19)=1.3, p>.2$ <i>Cohen's d</i> =.60	$M=2.24s$ $t(20)=3.97, p=.001$ <i>Cohen's d</i> =1.78

Experiment 3: Difference Scores for each Comparison X Age (months)

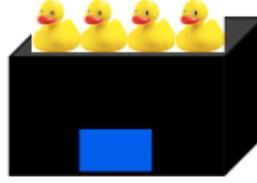
Table 8.

Age Group	2v3	2v4
14-17 months	$M=.153s$ $t(24)=.467, p>.6$ <i>Cohen's d</i> =.2	$M=1.07s$ $t(22)=1.89, p=.072$ <i>Cohen's d</i> =.81
18-23 months	$M=1.11s$ $t(14)=2.85, p=.013$ <i>Cohen's d</i> =.59	$M=1.5s$ $t(14)=1.94, p=.073$ <i>Cohen's d</i> =1.04
24-29 months	$M=1.84s$ $t(20)=1.84, p=.006$ <i>Cohen's d</i> =.82	$M=3.12s$ $t(21)=5.82, p<.001$ <i>Cohen's d</i> =.72
30-36 months	$M=3.66s$ $t(24)=6.24, p<.001$ <i>Cohen's d</i> =2.5	$M=3.62s$ $t(23)=5.36, p<.001$ <i>Cohen's d</i> =2.24

Cross-Experiment Analysis: Difference Scores for each Comparison X Age (months) in the Standard Condition.

Conditions**Standard Condition**

(Replicating Feigenson & Carey, 2003)

**Language Condition**

"Look at my toys. 1, 2, 3, 4. See, 4!"

**Heterogeneous Condition**

Figure 1.

Experiment 1: There were three conditions in which infants participated: Standard, Language, and Heterogeneity.

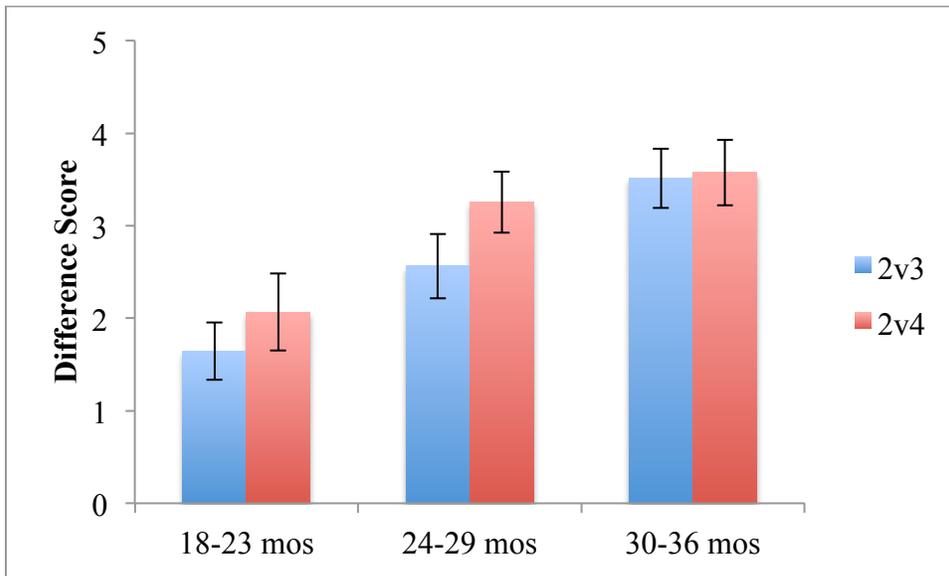


Figure 2.

Experiment 1. Age Group X Comparison (compared to average difference score vs. 0)

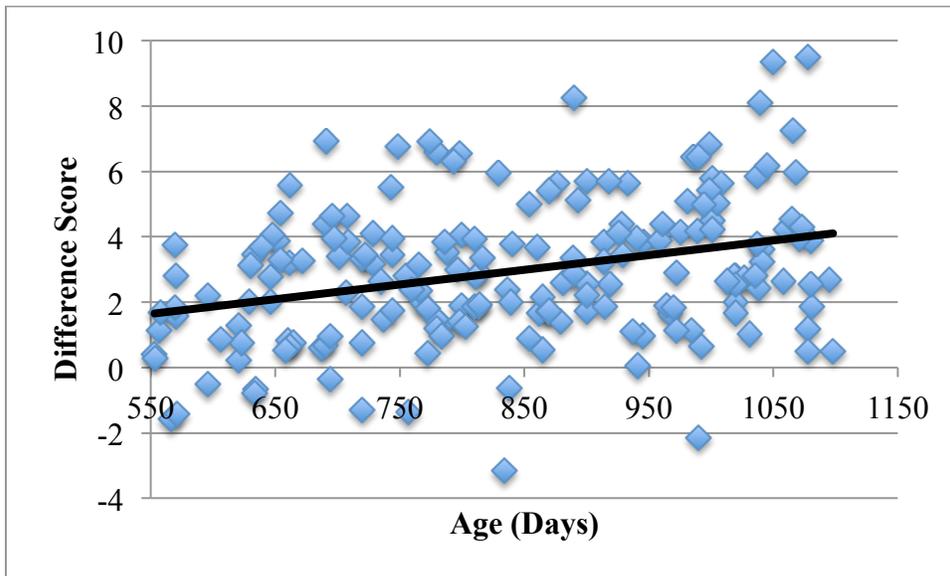


Figure 3.

Experiment 1. Average search time was positively correlated with age (in days), indicating that infants generally searched more across development.

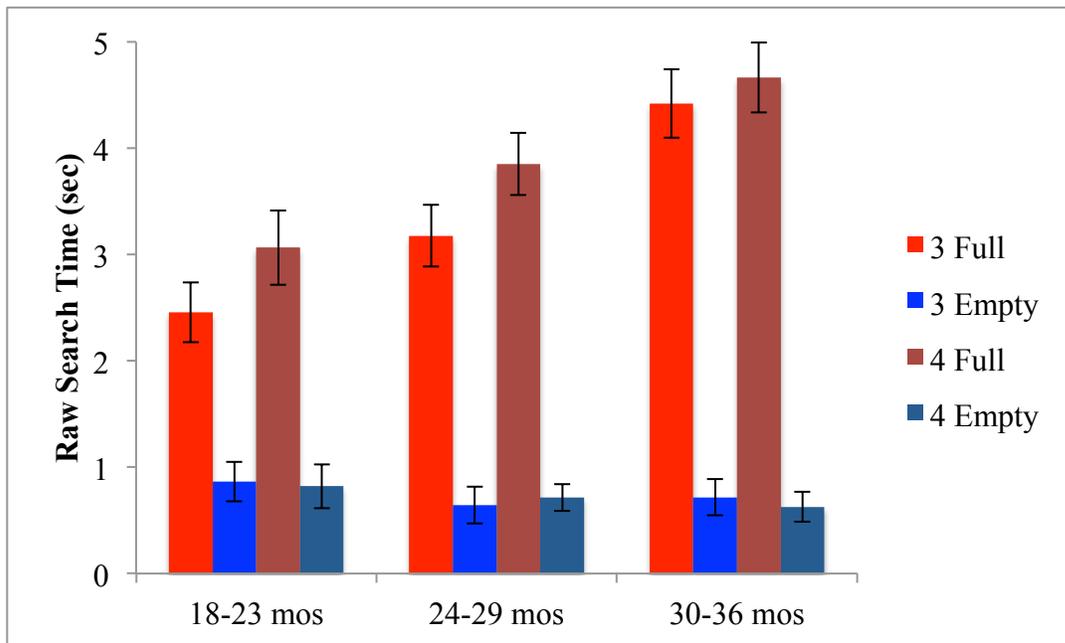
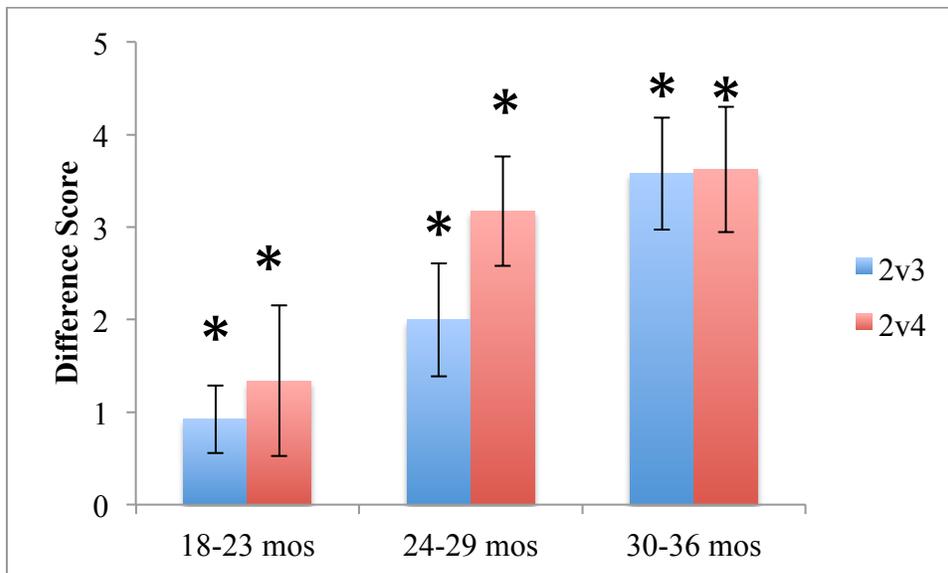
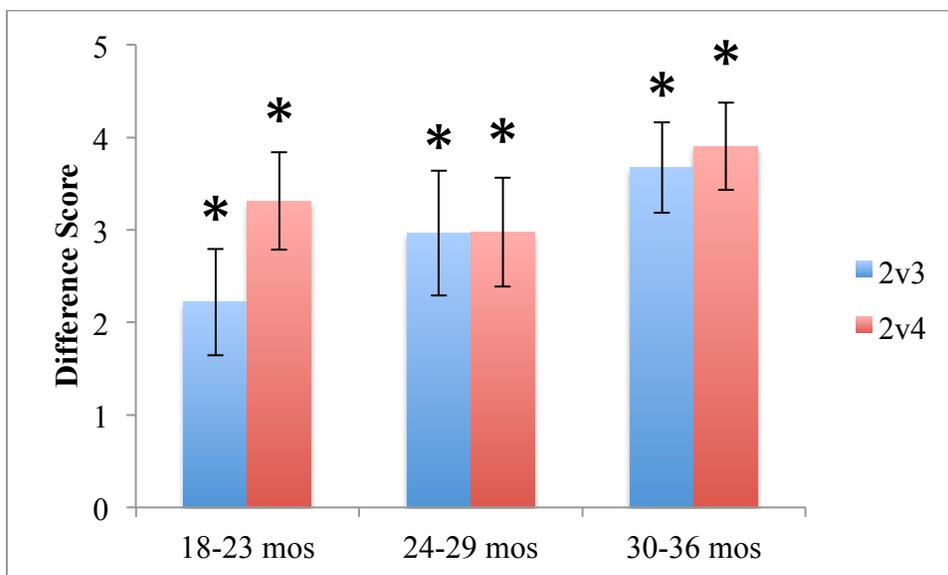
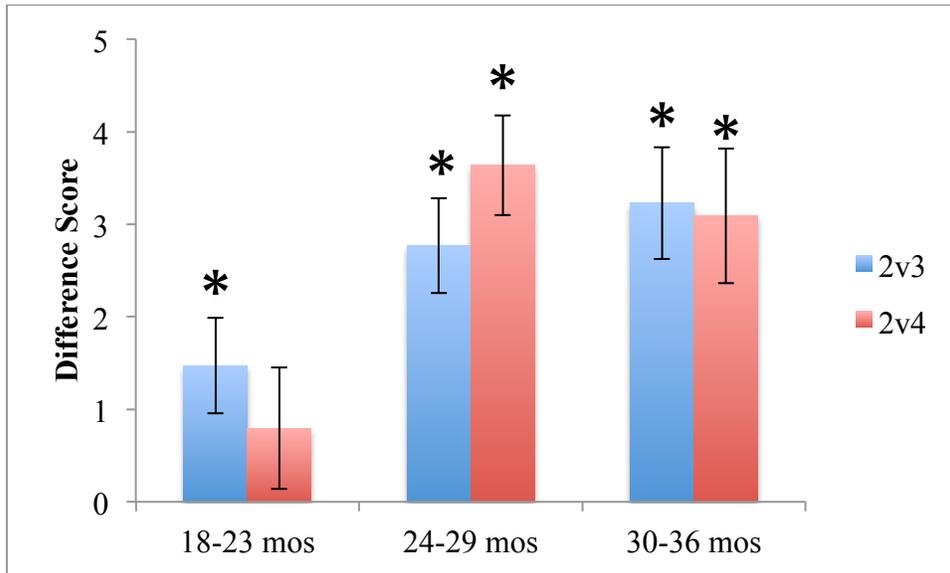


Figure 4.

Experiment 1: Graphed by Trial Type (Expected Empty vs. Expected Full), children searched more with age; however, this was due to increased searching in Expected Full Trials, with comparable and low performance on Expected Empty trials over development.

Standard:**Heterogeneity:**

Language:**Figure 5.**

Experiment 1: Age Group X Comparison X Condition

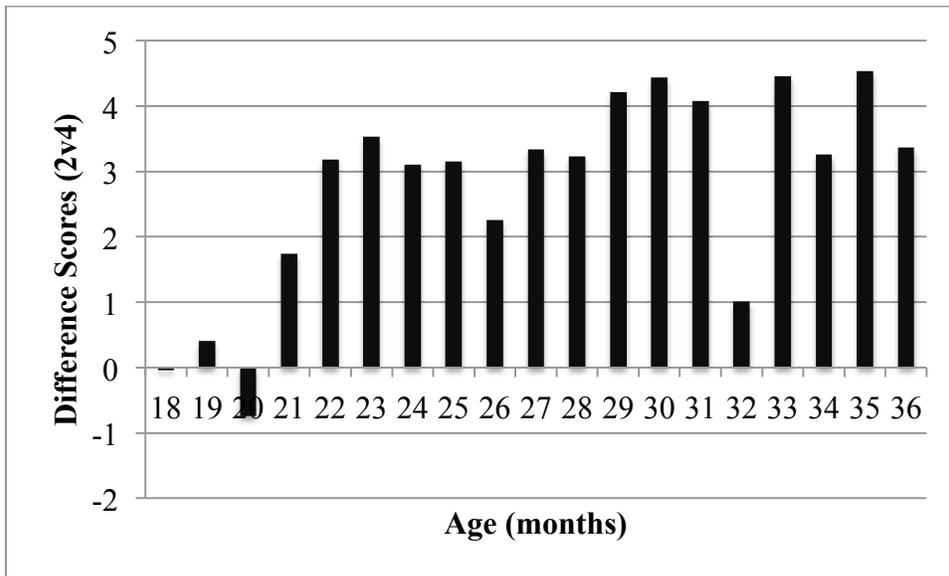


Figure 6.

Experiment 1: Histogram (Means Plot) for the 2v4 Comparison. Infants generally searched more with age, leading to greater success they developed.

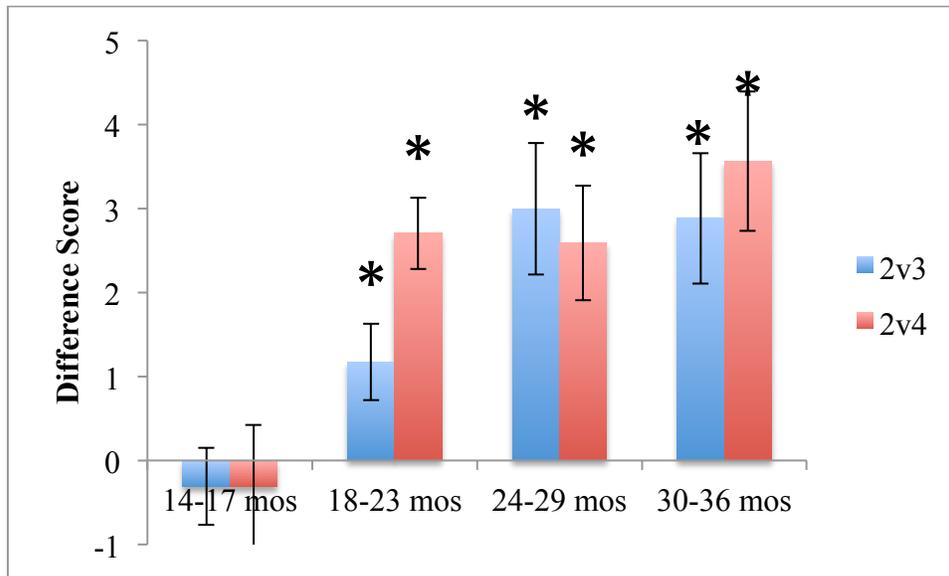


Figure 7.

Experiment 2: Standard Control. Difference score analyses confirm that, although 14-17 month olds fail to compare small and large sets, 18-36 month olds robustly make this distinction.

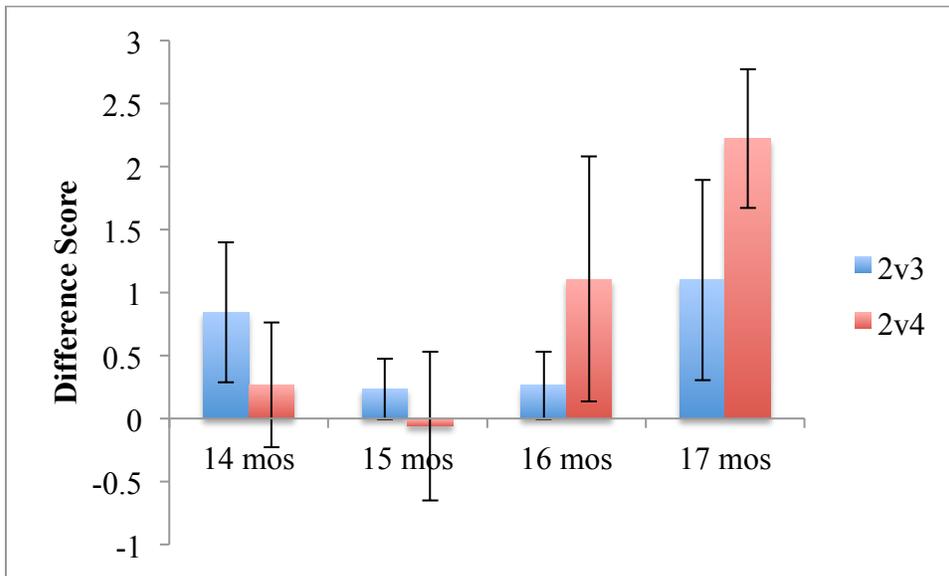


Figure 8.

Experiment 3: Age (in months) X Comparison (difference scores vs. 0)

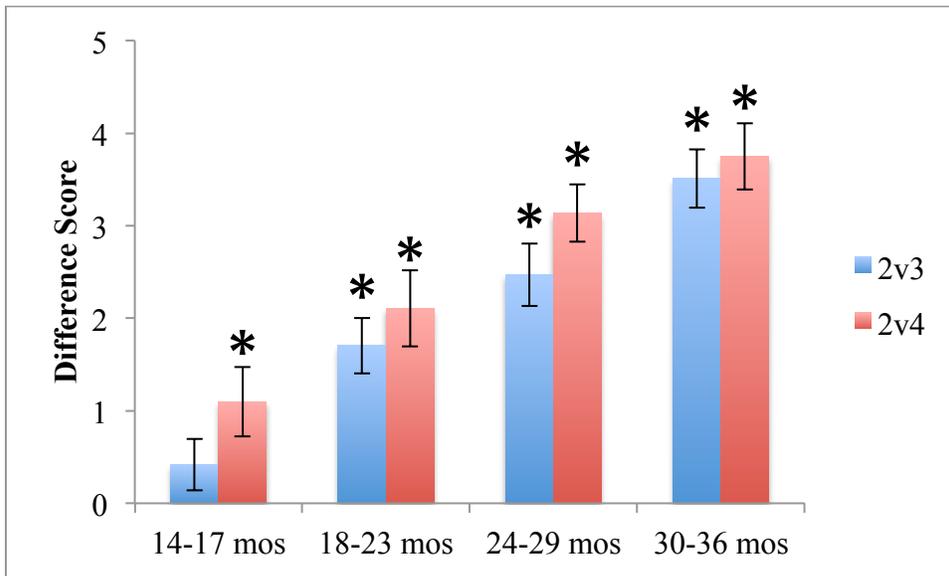


Figure 9.

Cross-Experiment Analysis: Experiments 1 and 3 (Standard Condition only; difference scores vs. 0)