



# Continuities and discontinuities in working memory representations of collections over ontogeny

## Citation

Tuerk, Arin Samantha. 2014. Continuities and discontinuities in working memory representations of collections over ontogeny. Doctoral dissertation, Harvard University.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:13129561>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

© 2014 Arin Samantha Tuerk

All rights reserved.

Dissertation Advisor: Professor Susan Carey

Continuities and discontinuities in working memory representations of collections over ontogeny

## ABSTRACT

Working memory, or the ability to maintain and manipulate information such that it can be used to guide behavior, is known to be severely capacity limited, in most circumstances, to about 3-4 objects. Both infants and adults have the ability to surpass these limits by encoding to-be-remembered items in groups or *collections*, exploiting statistical regularities or conceptual information to devise more efficient coding schema. Despite progress made toward understanding continuities in working memory, little is known about how changes over development interact with the ability to employ maximally efficient mnemonic data structures.

Paper 1 demonstrates that although adults can encode at most three mutually exclusive collections that accrue sequentially over time, they can circumvent this limit when items overlap in features (e.g. red and blue circles and triangles) and statistical regularities are introduced among collections defined by a single visual feature (e.g. most red items are triangular and not circular). Adults' performance suggests they are able to encode items from intersecting collections hierarchically and exploit statistical regularities among collections to reconstruct the numerosities of up to six collections in parallel, exemplifying how efficient coding can radically enhance working memory.

Paper 2 demonstrates that young preschoolers can also represent three mutually exclusive collections that accrue in an intermixed fashion over time. Results show that the ability to

to surpass this capacity limit by hierarchically reorganizing collections and exploiting statistical regularities among them develops between the ages of three and seven. These results are discussed in the context of executive function development.

Paper 3 provides evidence that computations of average size and orientation rely on qualitatively different processes with distinct developmental trajectories. Experiment 1 demonstrates that while the presence of additional identical elements in an array detrimentally impacts 6-month-olds' representations of element size, it improves the precision with which infants represent orientation. Experiment 2 demonstrates that performance is not affected when infants' attention is cued to a single item within arrays. These results are discussed in the context of the development of controlled attention.



## TABLE OF CONTENTS

Acknowledgements	p. vi
Introduction	p. 1
Paper 1: Using Statistical Regularities to Increase Working Memory Representations of Approximate Number	p. 37
Paper 2: Hierarchical Working Memory Representations of Collections Over Ontogeny	p. 74
Paper 3: 7-month-olds Extract Element Orientation But Not Element Size From Homogeneous Arrays	p. 113
Conclusion	p. 145
References	p. 157
Appendices	p. 170

## ACKNOWLEDGEMENTS

I could never have written this dissertation without the love and support of my amazing family. First and foremost, thank you to my parents. Any achievement of mine is truly thanks to you both. Thank you for instilling in me a love of learning, a ceaseless curiosity and the confidence to question everything. You are both my very best friends and coolest people I will ever know. Thank you to Freya and Matt, you are the most generous, loving and caring siblings I could ever ask for, and I truly do not deserve you.

Thank you also to Henry, my soul mate and partner. Thank you for your endless patience and for making me feel like I can do anything I put my mind to. You inspire me to be a better person, to take risks, to try new things, and to learn something new every day. I can't wait to grow together every day for the rest of our lives.

Finally, I thank the members of my intellectual community without whom this dissertation would never have been possible. Thank you to my committee – Susan Carey, Elizabeth Spelke, George Alvarez – and my outside examiner Yaoda Xu, for critical commentary and support. A special thank you to Susan Carey for being an incredible advisor. I'm not sure where this adventure will lead me, but I know that whatever I end up doing, I will do better because of the time we spent together. I am also extremely grateful to Lisa Feigenson for numerous contributions to the work in this dissertation, and for inspiring in me a love of cognitive science and empirical research. I thank Mariko Moher not only for being an incredible mentor and role model, but also for being a wonderful friend. I'd also like to thank all of the amazing friends I have made in graduate school for hours of insightful conversation, but more for all of the hours of not-so-insightful conversation. I love you and will miss you all.

## INTRODUCTION

Over ontogeny, children develop the ability to communicate, cooperate, and reason logically about the world in ways unprecedented in any other species. At the center of this development is the ability to form enduring mental representations of pertinent information in an ever-changing environment. We are constantly thinking about things for which we lack direct perceptual input, as when tracking an object that passes behind another object, or reaching for an object that has become hidden by an intervening surface. Thinking about things unseen requires mnemonic representation, such that processing can continue despite the lack of incoming sensory input.

Working memory, or the ability to maintain and manipulate information such that it can be used to guide behavior, is known to be severely capacity limited, in most circumstances, to about four individuals or so (e.g., Cowan, 2001; Luck & Vogel, 1997; Scholl & Xu, 2001; Sperling, 1960). The signature capacity limits of working memory are shared by a variety of non-human animals and human infants (Non-human animals: Hauser, Carey & Hauser, 2000; Uller & Lewis, 2009; Infants: Feigenson & Carey, 2003, 2005; Feigenson, Carey & Hauser, 2002; Ross-Sheehy, Oakes & Luck, 2003; Adults: Alvarez & Cavanagh, 2004; Broadbent, 1975; Cowan, 2001; Jiang, Olson, & Chun, 2000; Luck & Vogel, 1997; Sperling, 1960; Y. Xu, 2002), suggesting great continuity in working memory processes over phylogeny and ontogeny. In addition to having similar capacity limits on the number of separate items we can store in working memory, both infants and adults have the ability to surpass these limits to some extent by encoding to-be-remembered items in groups or *collections*, exploiting statistical regularities in the input or conceptual information to devise a more efficient coding scheme (adults:

Miller, 1956; Ericsson, Chase, & Faloon, 1980; Chen & Cowan, 2005; Baddeley, Thomson, & Buchanan, 1975; Burgess & Hitch, 1999; Estes, 1973; Zhang & Simon, 1985; Luck & Vogel, 1997; Brady, Konkle, & Alvarez, 2009; infants: Feigenson & Halberda, 2008; Moher, Tuerk & Feigenson, 2012; Rosenberg & Feigenson, 2013).

By 14 months, infants can efficiently reorganize information into hierarchically organized collections called *chunks* to bypass working memory capacity limits. While infants can only remember three individual items hidden at once in the absence of any grouping cues, when given perceptual, conceptual, linguistic or spatial cues to grouping, 14-month olds successfully chunk items together to remember four items in the same paradigm (Feigenson & Halberda, 2008; Rosenberg & Feigenson, 2013). In addition to chunks, which maintain representations of the individuals contained therein, both infants and adults can represent collections of simultaneously presented items in visual working memory called *ensembles* which are selected on the basis of spatial location or shared visual features such as color. Critically, ensemble representations contain summary information about the entire array, but do not maintain representations of the individual items therein (Ariely, 2001; Feigenson, 2011). Research has shown that 6-month-old infants form working memory representations of numerosity defined over entire large ( $> 4$  items) arrays (e.g. Brannon, 2002; Brannon, Abbot, & Lutz, 2004; Cordes & Brannon, 2008; Libertus & Brannon, 2010; Xu & Spelke, 2000) though they can store at most two individual items at once (Moher, et al. 2013). Additionally, when presented with arrays of dots of two colors, both 9-month-old infants and adults are able to remember the numerosities of the two color subsets and the superset of all dots, representing information about three ensembles in total (infants: Zosh, Halberda & Feigenson, 2011; adults: Halberda, Sires & Feigenson, 2006).

Questions addressed in the study of cognitive development fall into three broad categories: characterizing the initial state and continuities of abilities over ontogeny, characterizing important changes in abilities over development, and characterizing the mechanisms that underlie these discontinuities. Much research in the field has addressed the emergence of particular conceptual content such as knowledge of number (e.g. Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004), objects (see Spelke, 1998, for a review), geometry (Dehaene, Izard, Pica, & Spelke, 2006) and the social world (Spelke, & Kinzler, 2007), but the same questions arise in the study of the structure of memory. Despite the extensive progress made toward understanding working memory storage capacity for individual objects, chunks, and ensembles over the lifespan, little is known about how the ability to efficiently deploy these mnemonic structures changes over development. The ability to encode chunks which ground out in representations of the individuals contained therein has been well characterized over development. The current dissertation research instead focuses on summary representations of large collections in working memory. In the first two papers, I extend our understanding of collection (i.e., ensemble) representations in visual working memory to a different kind of working memory—one for dynamic events that unfold over time, and characterize continuities and discontinuities in the processes underlying this type of working memory representation. In the third paper, I discuss developmental discontinuities in ensemble representations of element features in visual working memory and explore possible mechanisms underlying this change.

### *The multi-faceted memory system*

#### *Short-term and Long-term memory*

The empirical study of memory dates back at least to the 19<sup>th</sup> century work of Hermann Ebbinghaus (1885) who taught himself series of nonsense syllables, and measured his forgetting curves over delays ranging from 20 minutes to 31 days. Ebbinghaus observed that he often had a “first fleeting grasp... of the series in moments of special concentration” (p. 33) but that this initial memory did not always lead to a robust memory of the series. Additionally, he found that the number of syllables that could reliably be learned in a single session was limited to seven. Soon after, William James (1890) proposed that memory was composed of two systems, a primary memory which could hold a small amount of “fleeting” information, and a secondary memory, a seemingly boundless long-term store of information accumulated over a lifetime.

Since the time of Ebbinghaus and James, multiple lines of research have produced evidence in support of a multi-faceted memory system. The need for different systems with different specializations is mandated by the range of mnemonic tasks we perform in our every day lives. While long-term memory has a vast capacity for storing enduring mnemonic representations, the amount of information stored by an individual over long durations is too large to be accessed rapidly for immediate comparisons such as those we constantly perform on a moment-to-moment basis, like remembering a phone number until you can dial it. Short-term memory, on the other hand, allows for the temporarily increased availability of information in memory such that it can be used to carry out tasks, but fades quickly when our attention is shifted elsewhere (Cowan et al., 1999).

Behavioral evidence supports the distinction between short-term and long-term memory. Convergent evidence from multiple paradigms demonstrates that short-term memory has a capacity of only about three to four items. For example, in a change detection task, adult

observers were presented with flashing arrays of colored squares and were required to respond when any of the squares changed color from flash to flash. Performance was at ceiling for arrays of one to three items, but declined sharply when the arrays contained four or more items (Luck & Vogel, 1997). A similar capacity limit was observed in a task that required searching a complex scene for the changing item. In this task, adults were shown a rapidly flashing grid of 32 dots of varying luminance values in which all but one of the dots maintained its original luminance value from flash to flash. By analyzing the number of flashes required for an observer to locate the target, researchers determined that adults were succeeding by storing and comparing only approximately 3 dots across successive flashes (Halberda, Simons and Wetherhold, 2006). Unlike long-term memory, the amount of information that can be held in short-term memory is also limited by the duration of time over which representations must be retained. For example, the ability to recall a list of words is affected by word length, with fewer words being recalled when the words have longer spoken duration (Baddeley, Thompson & Buchanan, 1975).

Neurological evidence also supports the distinction between short-term and long-term memory, as tasks that require only the temporary maintenance of information have been shown to engage the prefrontal cortex, and tasks tapping long-term memory often engage the medial temporal lobe (Cabeza & Nyberg, 2000; Talmi, Grady, Goshen-Gottstein, & Moscovitch, 2005). Additionally, amnesic patients may have impaired short-term memory, but retain long-term memory, and vice-versa, demonstrating a double-dissociation between these systems (Baddeley & Warrington, 1970; Drachman & Arbit, 1966; Scoville & Milner, 1957).

Multiple proposals for additional subdivisions of the memory system have been put forward. In one model, proposed by Atkinson and Shiffrin (1968), the mnemonic system is divided into short-term memory, long-term memory, and sensory registers which take input from

the environment and may then transfer that information into short-term memory. If information is not transferred to short-term storage, it quickly decays or is replaced by new incoming information. In Atkinson and Shiffrin's model, information flows from sensory registers to short-term memory to long-term memory, and information may be lost during any of these inter-stage transfers.

In an alternative model of memory, Baddeley and Hitch (1974) proposed that short-term memory itself is divided into multiple components. On this model, short-term memory consists of two modality specific systems: a "visuospatial sketchpad" responsible for visual and spatial information and a "phonological loop" responsible for auditory, linguistic and temporal information. These systems are directed by a third and critical component of short-term memory, the "central executive," that supervises the integration of information and coordinates the two modality-specific "slave" systems. The proposal of a central executive system that could direct attention to relevant information, suppress irrelevant information and coordinate additional cognitive processes introduced the idea that information could not only be held, but also manipulated during temporary storage. With this distinction, short-term memory gained its alternative name, *working* memory (Baddeley, 1998, 2003; Baddeley & Hitch, 1974).

### *Working memory in adulthood*

Though some aspects of the distinction between short-term memory and working memory remain debated (see Cowan, 2008 for review), it is generally accepted that working memory represents an active memory system responsible not only for the temporary maintenance, but for the simultaneous manipulation of information that is typically required in



complex cognitive tasks. Working memory provides a workspace for the transformation of information during cognitive tasks and serves as a bridge between externally and internally generated mental representations.

A task is said to tap working memory processes when it requires both maintenance and transformation of information over a brief period. For example, imagine the following tasks: first, you hear a string of digits presented sequentially, and you are to repeat as many digits as possible after a five-second delay; second, you must mentally compute the sum of 55 and 77. In the first task, which taps short-term memory, success requires the maintenance of the string of digits, information you perceived directly from the environment. In the addition task, you must hold and operate over the addends, remembering to carry the 2, information resulting from the transformation of these representations, from the ones column to the tens column as you go. While the two constructs are inextricably linked, short-term memory and working memory have been shown to cluster onto separate factors in factor analyses of children, adolescents and adults (Alloway, Gathercole, Willis, & Adams, 2004; Gathercole et al., 2004), and some evidence suggests they are linked to distinct neural subsystems (D'Esposito, Postle, & Rympa, 2000; Smith & Jonides, 1999).

Working memory is also a fundamental component of executive function, which has been characterized as a suite of domain general cognitive skills necessary for controlling cognitive actions. Planning and monitoring behavior, switching between tasks and controlling attention to select and update mental representations all rely on executive function (Diamond, 2002; Stuss; 2011). It is nearly impossible to overstate the importance of executive function, which supports a wide array of complex cognitive skills such as creativity, flexible perspective taking, self-control and attentional regulation. In addition to predicting reading and math competencies throughout

the school years (Gathercole, Pickering, Knight, & Stegmann, 2004), executive function remains critical for success throughout life. Children with poorer executive function at ages 3 to 11 tend to have more health problems (Crescioni et al. 2011, Riggs et al. 2010), less job success (Bailey, 2007), and commit more crimes (Denson et al., 2011) in adulthood as compared to their peers with higher executive function (Moffitt, et al., 2011).

Studies capitalizing on individual differences in working memory capacity have found that adults' performance on working memory tasks requiring controlled attention, but not span tasks, correlate with fluid intelligence (Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Bleckley, Conway & Engle, 2001). Thus, the aspects of working memory that correlate most strongly with fluid intelligence are also those aspects most critical for supporting executive function.

### *Individual objects in working memory*

Although working memory is critical for carrying out any complex cognitive task, the system itself, like short-term memory, is severely capacity limited. Behavioral studies attempting to quantify the capacity of working memory for visual information, i.e., visual working memory, have converged on the finding that adults can represent only 3 or 4 individual items at once (Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; Luck & Vogel, 1997; Sperling, 1960; Zhang & Luck, 2008). Neuroimaging and electrophysiological studies have also provided evidence of this signature capacity limit. By parametrically manipulating visual working memory load, researchers have been able to isolate the brain regions associated with it (e.g., Braver et al., 1997; J. D. Cohen et al., 1997; Linden et al., 2003). Evidence from group averages across

multiple participants in functional magnetic resonance imaging (fMRI) studies has demonstrated that activity in the intraparietal sulcus (IPS) and occipital cortices tightly correlates with the number of representations actively held in visual working memory (Todd & Marois, 2004). In this group average study, the maximum number of items maintained in visual working memory was found to vary widely across participants, from 1.74 to 6.37 (Todd & Marois, 2004). In a follow-up study, Todd and Marois (2005) found that IPS activity predicts individual differences in capacity. Additionally, data from fMRI experiments have demonstrated that activations in the IPS increase as set size increases from 1 to 4 during visual working memory encoding and maintenance (Y. Xu & Chun, 2005). Activation in this area peaked around 4 items and did not vary with item complexity suggesting that this capacity limit may originate in the IPS.

### *Chunks in working memory*

In order to determine the number of individual items we can hold in working memory, previous studies have employed stimuli without reliable patterns or redundancies. However, our everyday lives are full of rich contextual information, regularities and structure such that to-be-remembered items are rarely represented independently. One way in which humans circumvent the limits of working memory capacity for individuals is by capitalizing on the structure of input and grouping related individual items into chunks. The representation of chunks of items, where a chunk is a collection of individuals that are more strongly associated with one another than with any extrachunk items, has been shown to increase working memory performance as

retrieving a single chunk retrieves representations of the items it comprises (Cowan, 2001; 2010; Cowen, Chen & Rouder, 2004).

In the first scientific examination of this ability, Miller (1956) observed that we can, under some conditions, bypass the strict 4-item limit in working memory by recoding information into a more manageable hierarchical format. Miller suggested that, “the process of memorization may be simply the formation of chunks, or groups of items that go together, until there are few enough chunks so that we can recall all the items” (Miller, 1956, p. 94).

Cowan et al. (2004) expanded Miller’s 1956 constant capacity hypothesis to investigate the number of chunks that can be maintained in working memory, and the number of items that may be contained within a single chunk. In order to create chunks that varied in size between conditions, Cowan et al. (2004) introduced a training phase during which printed words were presented either as singletons or as members of a pair, and the associative strength of word pairings was manipulated by varying the proportion of singleton and paired presentations for each word. During test, lists of 8 words were presented for serial recall. In their analysis, a pair of words was counted as a single, two-word chunk whenever the two words were both presented together within the list and recalled in the same order in immediate succession. Cowan was able to determine that increase in performance on serial recall was due to chunking of words into pairs and that capacity for total units (chunks+ singletons) was constant at around 3.5 units (Cowan et al., 2004).

If the number of chunks one can store is roughly 3 or 4, and the number of items that can occupy a chunk is also limited to about 3, then the benefits of chunking should only expand working memory capacity so far. However, the ability to reorganize information hierarchically in

memory allows a powerful way to expand beyond this, by nesting chunks within larger chunks. Drawing on our long-term memory, we are able to apply semantic knowledge to form meaningful chunks of WM input and thereby increase our WM span. This is demonstrated by the case study of participant S.F. As a long distance runner, S.F. had substantial knowledge of various meaningful race times, and when presented with a series such as 347, S.F. would encode the string as 3 minutes and 47 seconds: the official race time of a particular individual for a one-mile run. Through this process of chunking, S.F. was able to use a single, familiar concept to unify 3 formerly independent items (Ericsson et al., 1980). This method is an effective strategy for expanding memory capacity because it allows for the creation of a hierarchy wherein a single label has underneath it several individual units. Once encoded, the single label is all that needs to be rehearsed since recall of the label means recall of the individuals beneath it in the hierarchy.

In a recent series of studies, Brady, Konkle and Alvarez (2009) found that observers could store more items in working memory when there are regularities in the training input. In these studies, participants performed a verbal interference task while they were shown displays containing 8 colors in the form of 4 items, each comprised of two concentric circles and two shaded areas. Across trials, experimenters introduced covariance between colors such that some color pairs were more likely to appear than other color pairs. During test trials, participants were shown outlines of the circles they had just seen and asked to recall the color of one of the 8 areas. Participants successfully recalled almost two times as many items from the patterned displays than from those with uniformly paired colors. The authors reason that participants' ability to compress the input and thereby increase the number of items they can hold in WM is similar to the process of chunking. By forming strong associations within the high-probability color pairs, participants were able to encode the inner and outer colors of a circle as a bound unit. Because

chunking frees up other WM resources to encode other individual items, participants were able to encode more items on trials containing more high-probability pairs.

### *Collections in working memory*

We have so far discussed the ways in which both individuals and chunks of individuals may function as units in working memory, but many of the things that we interact with and encode in our everyday lives are neither lone individuals nor small groups of these individuals, but rather large collections of items. There is evidence that large collections, whether they appear simultaneously or collect over time, may also serve as units in working memory.

Multiple ensemble characteristics can be extracted from collections of simultaneously presented items. Adults can quickly and accurately extract ensemble representations such as average orientation or size, or the total number of individuals within an ensemble (orientation: Orban, Vandebussche & Vogels, 1984; Parkes, Lund, Angelucci, Solomon & Morgan, 2001; size: Ariely, 2001; Chong & Treisman, 2003, 2005a, 2005b; numerosity: Barth, Kanwisher & Spelke, 2003). While there has been controversy as to whether participants in average size tasks are computing over all individuals in an ensemble or merely computing over some sampled subset (Simons and Myczek, 2008), numerosity judgments require taking into account all of the individuals within an ensemble, since samples of individuals do not yield cardinal estimates of the total ensemble.

Adults can accurately estimate the numerosities of multiple ensembles in parallel, and given that numerosity is indisputably a property of an ensemble rather than a property of an

individual item, this evidence confirms that an entire ensemble, rather than some sample of the items contained in the ensemble, can serve as a unit in visual working memory. In the first study of this ability, Halberda, Sires and Feigenson (2006), required participants to make numerosity judgments about briefly presented dot arrays, sometimes consisting of intermixed ensembles of dots of different colors. In this experiment, participants saw dot displays of one to six different colors and were asked to estimate the number of dots of a given color as well as the total number of dots of all colors. Trials were either “probe before,” in which participants were told which color dots they were to enumerate before seeing the displays, or “probe after,” in which they were not told which color collection they would be asked about until after the display had flashed. By comparing participants’ accuracy on these two trial types, researchers were able to determine the number of ensembles a participant could enumerate in a single flash. If accuracy was the same on both trial types, participants had to have approximated and stored numerosities of all of the presented color ensembles, independent of when the probe appeared. Worse performance on the “probe after” trials indicated that there were more color subsets than participants could enumerate simultaneously. Results revealed worse performance on the “probe after” trials when there were more than two colored subsets, suggesting that participants could enumerate at most two color subsets in parallel. However, researchers observed that irrespective of the number of colored subsets, participants always accurately reported the numerosity of the superset of all dots, suggesting that adults can encode up to three hierarchically organized ensembles in parallel. Recent research has demonstrated that information encoded about both supersets and subsets can be used to reconstruct information about individual items contained in the ensembles, even when the total number of individuals presented greatly exceeds working memory capacity limits. For example, observers’ judgments of the size of a given dot within an

array of dots is biased both by the mean size of the dots of the same color and by the mean size of all dots in the display, suggesting that the representation of the individual dot is constructed by integrating information across different levels of abstraction (Brady & Alvarez, 2011).

Ensemble representations clearly support the encoding of information in cases where working memory capacity limits would be otherwise exceeded. In our everyday lives, however, parsing the world into distinct collections is not always so straightforward, for at least two reasons. First, collections of individuals are not always perceptually available as arrays in their entirety. Rather, they frequently accumulate over time, as when a stream of people emerges from a crowded theater. This necessitates updating collection representations dynamically over time. Second, since every individual has a large number of represented properties (e.g., gender, age, height), and these do not always cleanly parse the world into non-overlapping units, the criteria of collection membership are potentially unclear.

Feigenson (2008) probed the first of these issues, examining adults' ability to track collections that accumulated over time. Observers saw between one and five types of perceptually contrasting objects (toy pigs, poker chips, cotton balls, batteries, and green candies) placed one by one into each of two hiding locations. Different numbers of each type of object were placed in each location, with object types completely temporally intermixed, and all of the objects were placed in the first hiding location before any objects were placed in the second. To keep them from verbally counting, observers were instructed to perform verbal shadowing throughout the presentation. When the presentation ended, each hiding location contained either five or ten objects of each type, and observers were asked to indicate which location had more of a certain type of object (i.e., "Which bucket has more pigs in it?") and to verbally estimate how many objects of each type were in each location (i.e. "How many pigs would you say are in this



bucket? And in this one?"). Because the object types were intermixed and objects were presented one at a time, participants did not know when they had seen the last member of a given collection in a given location, and therefore had to update their numerosity estimates for each collection in parallel as the information was presented.

This study revealed that, despite the dynamic and temporally interleaved presentation of the collections, observers knew which of the two locations had more of any object type, and their numerosity estimates were in line with characteristics of the non-verbal Approximate Number System (e.g., Dehaene, 1997). However, this was only true when three or fewer collections were presented—observers' performance fell to chance on both the location identification and the estimation measure when presented with four or five collections. Thus, as in some other studies on the limits of working memory (Feigenson and Carey, 2003; 2005; Feigenson, Carey & Hauser, 2002), observers' failure was catastrophic when the demands of the task exceeded working memory capacity limits. When too many object types were presented, observers were unable to track three of the four or five collections (for example, by ignoring one or two of the presented collections in order to successfully represent the remainder).

Although Feigenson's study differed from visual working memory studies in which all of the to-be-remembered items are presented simultaneously and the relevant collections (i.e. ensembles) are immediately apparent in a single glance, it was still similar to previous work on ensemble representations in visual working memory in that each object clearly belonged to only a single collection. Collection membership was unambiguous, both on the bases of early visual features and semantic category. For example, the collections differed from each in color (pink vs. blue vs. white vs. black/silver vs. green), shape (ovoid with articulated parts vs. disc vs. fluffy balls vs. elongated cylinders vs. cubes), and category membership. Attending to any of these

features would lead observers to represent mutually exclusive, i.e., non-intersecting, collections, in which no object was a member of multiple collections at once. But what happens under more complex conditions, when individual objects can be represented as members of multiple different collections or groups?

To illustrate, imagine the mundane case of doing laundry, where there are at least two important dimensions to consider in preparing the wash: color and material. We may first sort items of clothing by color in order to determine whether they should go in with the light or dark wash, and we may then sort the same items along the orthogonal dimension of material to determine whether they should be hung on the clothesline or put in the dryer. One possible solution in such cases where collection membership is multiply determined is to represent four mutually exclusive collections defined by the conjunctions of features from these two dimensions: dark and dryer-friendly, dark and line dry only, light and dryer-friendly, and light and line dry only. A clear drawback to doing this, though, is that even in this simple case in which there are only two orthogonal dimensions, the number of resulting collections already exceeds working memory capacity (Feigenson, 2008; Halberda et al., 2006; Zosh et al., 2011).

This example highlights the intuition that when the bases for collection membership are multiply determined or ambiguous, it may be disadvantageous to set such narrow criteria for membership. Given that previous work shows that only three mutually exclusive collections can be represented at once (Feigenson, 2008, Halberda et al., 2006), the ability to represent collections using more flexible selection criteria may empower the encoding of more information than would be otherwise possible. To illustrate, we can return to the laundry example. The four mutually exclusive conjunction-based collections we discussed above might be more efficiently represented if an observer could first parse the array of laundry into two mutually exclusive

collections along either one of the two critical dimensions (e.g., first parse the pile of laundry into two color-based collections: light-colored and dark-colored), and then represent the subsets within each of the two resulting collections (e.g., within the light-colored collection, dryer-friendly and line-dry). On this encoding scheme, each individual clothing item is a member of two orthogonal collections (e.g., an item is dark and dryer-friendly). Importantly, any statistical regularities among the collections may allow information to be represented in a more compressed, hierarchical format. For example, if we first sort the laundry along the color dimension, we may determine that more of the laundry is light than dark. During this same sorting process, we may detect that most of the light colored laundry is dryer friendly. From these two pieces of information, we may then be able to infer that most of the laundry, in total, is dryer friendly.

Given that representing statistical regularities among individual objects can increase visual working memory capacity almost two-fold (Brady et al., 2009), sensitivity to statistical information among hierarchically organized collection representations might allow for an even greater expansion of memory. In the first study of this dissertation, we explore adults' ability to select and represent statistically structured feature-based collections, and the potential benefits of such hierarchical reorganization on working memory performance.

Borrowing the methodology from Feigenson (2008), we sequentially hide items from multiple collections in a temporally intermixed manner, too quickly to count, into two buckets. We then ask participants to select the bucket containing more items of a given type. In Experiment 1, we seek to replicate Feigenson's (2008) finding that adults can represent three, but not four mutually exclusive collections in this paradigm. In the following experiments we explore whether adults can hierarchically organize statistically structured collections of items

with overlapping features (e.g., red circles, blue circles, red triangles and blue triangles) and combine information across levels of the hierarchy to improve working memory performance.

### *Working memory development*

#### *Continuities in working memory over ontogeny*

The signature capacity limits of working memory are shared by a variety of non-human animals and human infants (Non-human animals: Hauser, Carey & Hauser, 2000; Uller & Lewis, 2009; Infants: Feigenson & Carey, 2003, 2005; Feigenson, Carey & Hauser, 2002; Ross-Sheehy, Oakes & Luck, 2003; Adults: Alvarez & Cavanagh, 2004; Broadbent, 1975; Cowan, 2001; Jiang, Olson, & Chun, 2000; Luck & Vogel, 1997; Sperling, 1960; Y. Xu, 2002), suggesting great continuity in working memory processes over phylogeny and ontogeny. In addition to having similar capacity limits on the number of separate items we can store in working memory, both infants and adults have the ability to surpass these limits to some extent by encoding to-be-remembered items in a variety of working memory data structures.

#### *Individual objects in working memory*

Across multiple methodologies, infants and adults show a similar strict capacity limit on the number of individual items they can hold in working memory. For example, in a change detection task, adults' performance was at ceiling for arrays of one to three items regardless of the items' complexity, but declined sharply when arrays contained four or more items, revealing

that working memory could store 3 or 4 items, but no more. (Luck & Vogel, 1997). In an analogous experiment, 10- and 13-month-old infants saw two flickering streams of arrays of colored squares where one stream displayed the same items across each 500 ms flicker, and the other stream contained one item that alternated in color between flickers. Infants looked reliably longer at the Changing than the Non-Changing Stream when the array contained 1, 2, 3, or 4 squares, however, they failed to look longer at the Changing Stream when the array contained 6 squares (Ross-Sheehy, Oakes, & Luck, 2003), suggesting that infants could remember the features of four, but not six items at once. Infants show similar capacity limits when to-be-remembered items are presented sequentially, successfully searching for one, two or three items they see sequentially hidden in a box, but failing when four items are hidden (Feigenson & Carey, 2003; 2005). The similar capacity limits observed across a wide variety of methodologies in 10- to 20-month-old children and adults suggests remarkable continuity in the structure of working memory over ontogeny. After a brief period of rapid development in the first year of life, working memory capacity for individuals remains incredibly constant over the lifespan.

### *Chunks in working memory*

Like adults, fourteen-month old infants can also efficiently reorganize information into hierarchically organized chunks to bypass working memory capacity limits. While infants can only remember three items hidden at once in the absence of any grouping cues, when given perceptual, conceptual, linguistic or spatial cues to grouping, 14-month olds successfully remember and search for four items in the same paradigm (Feigenson & Halberda, 2008; Rosenberg & Feigenson, 2013). In fact, even at 7-months of age, before working memory has

matured to the three-item limit, infants can already use multiple locations and shared features among items to form chunks and store more total items in working memory (Moher, Tuerk & Feigenson, 2013).

### *Collections in working memory*

In addition to chunks, where individual item representations are maintained, infants can represent summary information about entire ensembles in visual working memory selected on the basis of spatial location or shared visual feature such as color. Six-month-old infants can discriminate arrays of four items from arrays of eight items (Xu & Spelke, 2000) though they can store at most two individual objects at once (Moher, et al. 2012). The ability to select and encode the numerosity of multiple ensembles in parallel has also been demonstrated in 9-month old infants. When presented with arrays of dots of two colors, both infants and adults are able to remember the numerosities of the two color subsets and the superset of all dots. When arrays contain three or more colors of dots, both infants and adults encode the numerosity of the superset of all dots, but do not remember the numerosity of any color subsets (infants: Zosh, et al., 2011; adults: Halberda, Sires & Feigenson, 2006). In addition to individual items and chunks, infants can encode an entire ensemble as a unit in working memory. It appears that like adults, infants and young children have at their disposal a variety of data structures in which to store mnemonic information.

### *Potential discontinuities in working memory processes over ontogeny*

The hierarchically organized architecture of working memory that allows us to surpass some capacity limits appears to be present early in life; however, age-related gains in cognitive abilities may lead to discontinuities in the way we manipulate encoded information and employ the data structures (i.e. individual object representations, chunks and summary representations of collections) afforded by this architecture.

Although both infants and adults can successfully encode ensemble representations when items are presented simultaneously and the basis of selection is obvious, many collections of items we encounter in our everyday lives are not so clearly delineated. For one, collections of individuals are not always perceptually available in their entirety. While Feigenson's (2008) work demonstrated that adults could enumerate up to three large collections of items that accumulated over time, the development of this ability has not yet been studied, and will be addressed in paper 2 of the current dissertation. Additionally, individual items have a large number of represented properties that could potentially be relevant for determining collection membership, but these properties do not always cleanly parse the world into non-overlapping collections. Paper 1 of the current dissertation explores the ability to represent overlapping feature-based collections in adulthood; paper 2 of the current dissertation explores the development of this ability.

#### *Updating working memory representations of collections that accrue over time*

When multiple collections accumulate over time, encoding them necessitates the dynamic selection and updating of relevant collections as new items are presented. Feigenson's (2008) study demonstrated that adults encode summary representations of multiple collections over

prolonged hiding events and can update these representations as new items are presented sequentially. While both infants and adults can represent multiple ensembles of simultaneously presented items in visual working memory, it is possible that the mnemonic processes used in Feigenson's (2008) task are qualitatively different from those available early in development. To succeed in Feigenson's (2008) task, participants must compare the features of each item presented to those of previously presented items held in a working memory buffer, so that they may update the relevant collection representation or form a new collection representation while the others are maintained simultaneously.

Like adults, infants can represent objects in the absence of perceptual contact and to some degree, update these memory representations to reflect changes in a scene that unfolds over time. For example, 5-month-old infants who see a doll hidden behind a screen, then see a second doll placed behind the same screen, correctly expect that two objects, rather than one or three, will be revealed when the screen is lifted (Wynn, 1992). Success in this task requires that infants represent the initial array of a single item, maintain this representation as the item is hidden, then mentally update it to reflect the addition of a second object (see also Feigenson, Carey, & Spelke, 2002; Simon, Hespos, & Rochat, 1995; Uller, Carey, Huntley-Fenner, & Klatt, 1999). Infants can also update a memory representation to reflect an object's departure from a scene. Infants who see two objects hidden simultaneously behind a screen and then see one object removed, correctly expect that only one item will be revealed when the screen is removed (Wynn, 1992).

Additionally, 10- and 12-month infants can update representations of object arrays hidden across two distinct locations. Infants who see a single cracker placed into one bucket and then two crackers sequentially placed into a second bucket, reliably approach the bucket containing



more crackers, and do so for comparisons of one versus two crackers, one versus three crackers and two versus three crackers. However, when the comparison is one versus four crackers, 10- and 12-month old infants fail to approach the bucket containing four crackers (Feigenson & Carey, 2003; 2005; Feigenson, Carey & Hauser, 2002a). Note that the total number of crackers hidden is the same in the one versus four, and the two versus three crackers conditions, but the number of updates an infant must make in a single location, and the total number of crackers in a single location is higher in the one versus four crackers condition.

The aforementioned cracker tasks suggest that infants can mentally update representations of arrays across multiple locations and use the resulting representations to guide behavior. The limits of this ability have only recently been explored. In the previously discussed task, the dynamics of updating were predictable; namely, all crackers were placed sequentially into a single bucket before any crackers were placed into the second bucket. To succeed, infants could update their representations of the entire contents of the first bucket before forming any representation of the contents of the second bucket. As in the violation of expectation studies with younger infants, these studies never required infants to update a representation, switch attention to update another representation, and then return to the first representation to update it again.

To probe the limits of infants' updating abilities, Feigenson and Yamaguchi manipulated the "orderliness" of the cracker hiding task, hiding crackers in either direct succession (i.e. all of the crackers hidden in a single location and then all of the crackers hidden in the second location) or in alternation, (i.e. some crackers hidden in a single location, then some crackers hidden in the second location, then more crackers hidden in the first location). 11-month olds reliably selected the bucket with more crackers in comparisons of one versus two and one versus three when the

crackers were hidden in direct succession, but failed on the same comparisons when the crackers were hidden in alternation (Feigenson & Yamaguchi, 2009). This failure to reupdate is not specific to the modified foraging task, and appears when implicit measures are used. In a violation of expectation looking time task, Moher and Feigenson (2013) hid objects behind two screens, two behind one screen and one behind the other, and then lifted the screens to reveal either all three objects or only two objects. Eleven-month old infants correctly expected three objects when the objects were hidden in direct succession, but looked equally at the two and three object outcomes when the objects were hidden in alternation. When the two screens were connected with a small foam strip, however, infants again successfully represented the three-item outcome. This suggests that infants' representations are robust to updating when an event can be construed as containing multiple updates to a single array, but are not flexible enough to withstand an event requiring updates to a previously attended display.

Although evidence from the aforementioned cracker task suggests that toddlers, whose working memory capacity for individual items, chunks and collections parallels that of adults, are unable to flexibly update multiple collections in parallel, it is possible that the manipulations involved in the cracker study are different from those involved in studies of adult updating. For adults, reupdating working memory representations is still more difficult than updating a currently attended representation, but the effect of this increase is relatively minor. For example, adults presented with sequences of triangles and rectangles and asked to announce the total number of items in each shape-based collection after each item is presented, are slower to respond when they have to switch between the two counts than when they have to update the same count again (Garavan, 1998). It is possible that reupdating in the case of incrementing alternating counters is easier than the reupdating required in the cracker task or the violation of

expectation task previously discussed because it requires alternating between two existing representations (counts for each collection) rather than forming a representation of a new object and adding it to a preexisting array representation. Previous work suggests that infants are representing each cracker as a separate object in the cracker task (Feigenson et al., 2002), and as such, are limited by the number of items they can efficiently represent as a chunk. Adults, contrastingly, need only to increment a single summary representation for each collection in the counter task. In the cracker task, the mnemonic units being employed are chunks, whose representational capacity is limited by the number of individuals they comprise. In tasks that require the updating of counters as collections accumulate, the summary representations serve as the units in working memory. It is possible that the ability to dynamically reupdate multiple representations in working memory is available earlier in life, but that these reupdates may only be carried out over a limited number of preexisting representations. This possibility will be addressed in paper 2 of this dissertation.

Infants' failure to reupdate representations in situations where they must shift attention away from a currently attended array to track objects elsewhere and then return to update the first representation again suggests that their working memory representations are less flexible than those of older children and adults. The ability to shift attention and to update representations in working memory are often discussed as subcomponents of a larger construct called executive function, which is known to develop significantly over the childhood years (for a review see: Diamond, 2002; 2013).

In fact, working memory itself is often discussed as a subcomponent of executive function, though many working memory researchers define the term more broadly so that it actually becomes synonymous with executive function. For example, in Baddeley's working

memory model, the central executive's functions include multitasking, shifting between tasks, and the capacity to inhibit and attend in a selective manner (Baddeley & Hitch, 1994).

Confusion over the definitions of these constructs has made it difficult to tease apart the developmental trajectory of updating in working memory from other executive function subcomponents.

Recently, inhibitory control has been shown to be one aspect of the multi-component executive function construct that clearly separates from the other components such as shifting and working memory in adults (Miyake et al., 2000) and older children (Lehto et al., 2003). In younger children however, the separability of executive function components remains a matter of debate. Attempts to disentangle the contributions of multiple executive function components have been muddled by the difficulty in devising “pure” tasks that tap only the intended executive function capacity (Miyake et al., 2000). The more recent use of Confirmatory Factor Analysis (CFA) has made it possible to test the relative strengths of unitary and multi-factor models of executive function in young children. By extracting the common variance across multiple tasks and measures of the same executive function component, CFA produces a resultant latent variable is assumed to be a purer measure of the executive function construct in question. Studies using CFA have found that a single, undifferentiated executive control factor best describes the executive function construct in early childhood and during the preschool years (Wiebe et al., 2008, 2011; Hughes et al., 2009; Willoughby, Blair, Wirth, & Greenberg, 2010; Fuhs and Day, 2011). However, more recent results have demonstrated that a two factor model, differentiating working memory and inhibition provided a better fit to data collected from a sample of preschoolers between the ages of 3 and 5 years than a single factor model or a three-factor model of working memory, inhibition and shifting (Miller et al., 2012). Similarly, in 5 and

6 year old children, a two-factor model composed of an inhibition factor and a working memory-flexibility factor was found to provide a better fit than a single factor model (Usai, Viterbori, Traverso, & De Franchis, 2014).

A unitary model in early childhood can be reconciled with a diverse executive function construct later in life if the structure of the component mental abilities changes with development. For example, Garon et al. (2008) proposed that the components of executive function emerge sequentially over the preschool years with working memory developing first followed by inhibition, and together these components enable the development of shifting. It remains an open question whether the flexibility involved in shifting among and updating multiple collections in parallel is available early in childhood while other executive function components, namely inhibition, are still developing.

Reorganization of information in working memory appears to be a fundamental and early-emerging solution to the challenge of storing large amounts of information in a strictly limited system. In addition to encoding chunks of items, where information about the individuals is maintained, both infants and adults can also store summary information about large ensembles of items in visual working memory.

In Experiment 1 of paper 2 of the current dissertation, we test 3-and 4-year olds in a task analogous to Feigenson's (2008) adult task, sequentially hiding temporally intermixed collections of items of three or four semantic types (e.g. blocks, balls, toy pigs and keys). If young children are able to represent three, but not four mutually exclusive collections under sequential presentation it will provide the first evidence that the ability to encode multiple collection-based summary representations under these conditions is available in childhood, and

that collections that accrue over time can serve as units in working memory early in development. Further, if 3-year olds succeed in representing the relative numerosities of three collections in parallel in this paradigm, it will demonstrate that the ability to encode, shift among, and update clearly delineated, homogenous collections in working memory emerges early in development, and provide evidence for the continuity of another mnemonic data structure over ontogeny.

In Experiments 2 and 3 of the second paper, we test 3- to 7- year olds and adults in a task analogous to that used with adults in paper 1 of the current dissertation, and explore the development of the ability to increase working memory performance through the representation of overlapping, statistically dependent collections. In Experiment 1 where collections are mutually exclusive (e.g. blocks, balls and pigs), the maximally efficient encoding schema allows for the representation of three collections simultaneously. In Experiments 2 and 3, collections are of four types of items with overlapping features (e.g. large and small blocks and balls) and statistical regularities are introduced among features on orthogonal dimensions (e.g. most blocks are large and not small, most balls are small and not large). With these overlapping collections, the criteria for collection membership is ambiguous, offering the opportunity to hierarchically reorganize what might be construed as four homogeneous collections (e.g. large blocks, large balls, small blocks and small balls) into two collections that are heterogeneous along the orthogonal dimension (e.g. blocks that are more often large than small).

We manipulate the bases of collection membership by instructing participants to label items' features along only one dimension (e.g. balls and blocks regardless of size). If participants are able to represent accumulating collections hierarchically, labeling each sequentially presented item along one dimension should lead them to represent the two feature-based

collections along that dimension as the top level of the hierarchy (e.g. labeling items along the shape dimension will lead to the representation of a collection of balls and a collection of blocks), and they should be better able to judge the relative numerosities of feature-based collections along the labeled dimension (e.g. “which bucket has more balls in it?”) than along the orthogonal, unlabeled dimension (e.g. “which bucket has more big things in it?”).

It is possible that once the four item types (big balls, big blocks, small balls and small blocks) are reorganized into two feature-based collections (e.g. balls and blocks), detecting and encoding the statistical regularities among the orthogonal, unlabeled features comes for free. If this is the case, then participants of any age who are able to answer questions about the labeled collections should also be able to answer questions about the orthogonal, unlabeled collections. However, evidence suggests that the ability to flexibly represent multiple features of objects, a necessity for encoding the statistical regularities in Experiments 2 and 3, develops significantly over the preschool years (e.g. Zelazo et al., 1996; Kirkham, Cruess, & Diamond, 2003). Like stimuli used in traditional task switching tests of cognitive flexibility (e.g. the Dimensional Change Card Sort (DCCS) used by Zelazo et al., (1996)), the items used in Experiments 2 and 3 are bivalent, that is, they carry a feature relevant to each of the two dimensions along which these items can be sorted. If representing a hierarchy of features beyond those that are labeled requires additional EF capacities, it is possible that participants at a certain age will be able to successfully answer questions about congruent feature-based collections, but fail on questions about incongruent feature-based collections. Together the experiments of paper 2 will shed light on how developing EFs interact with the underlying architecture of working memory and may lead to discontinuities in the way we organize and encode collections.

### *Developmental change in ensemble processing in visual working memory*

Research has shown that infants extract numerosity information defined over entire large (> 4 items) arrays (e.g. Brannon, 2002; Brannon, Abbot, & Lutz, 2004; Cordes & Brannon, 2008; Libertus & Brannon, 2010; Xu & Spelke, 2000). This ability appears to belie the strict capacity limits of focused attention, and suggests that infants, like adults, have compensatory processes that allow them to accurately perceive the statistical properties of a collection of objects, forming what is known as an ensemble representation (Alvarez, 2011).

Researchers studying adults' perceptual systems have claimed that these ensemble representations, or statistical summaries, are effectively computed via automatic processes not limited by the bottleneck of focused attention and working memory (Alvarez, 2011; Chong & Treisman, 2003; Chong, Joo, Emmanouil, & Treisman, 2008). Unlike serial processing of individual objects in a scene, which allows us to encode no more than four objects (e.g. Luck & Vogel, 1997), the extraction of summary statistics, such as the mean or distribution of features among a collection of similar objects, appears to employ a parallel mechanism devoid of capacity limitations. It has been claimed that this process relies on distributed attention across an entire array, and is a general mechanism that computes ensemble representations over multiple stimulus attributes including orientation, size, central tendency, and even facial expression and identity (Albrecht & Scholl, 2010; Alvarez & Oliva, 2008; Ariely, 2001; Chong & Treisman, 2003; Dakin, 2001; Dakin & Watt, 1997; de Fockert, & Wolfenstein, 2009; Haberman & Whitney, 2007; 2009; Parkes, Lund, Angelucci, & Solomon, 2001; Robitaille & Harris, 2011).



For adult observers, these ensemble characteristics have been shown to enhance memory for individuals through a process that replaces direct retrieval with inference based on summary statistics. For arrays of elements too numerous to be attended to and encoded via focused attention, adults are able to circumvent capacity limits and use ensemble representations in working memory to reconstruct information about individual elements. For example, when presented with a large array of dots, an adult will use the average size of all of the dots to inform his judgment of the size of a single dot presented therein (Brady & Alvarez, 2011).

Recent evidence suggests, however, that infants may not benefit from ensemble characteristics in a similar way. While adults seem to automatically extract the ensemble characteristic of element size from an array of dots, infants' representations of element size appear to be substantially hampered when items are presented in an array as opposed to in isolation. For example, 6-month-olds habituated to a single item (an Elmo face) detected a twofold change in size during test (Brannon, et al., 2006); however, when habituated to arrays of homogeneously sized dots, infants did not dishabituate to novel test arrays in which all dots underwent a threefold change in size (Cordes & Brannon, 2011). Though infants robustly detected a twofold change in the size of a single item (Brannon, et al., 2006), they required a fourfold change in item size to detect a change in an array of homogeneously sized dots (Cordes & Brannon, 2011).

That infants are less sensitive to uniform size changes in arrays of homogeneous dots than to size changes in a single dot appears to be in direct contrast with findings that adults' threshold for discriminating size change is the same for individual elements, and mean size judgments of homogeneous and heterogeneous arrays (Chong & Treisman, 2003). In fact, at delays of 2 seconds, adults' threshold for detecting a change in the average size of a

homogeneous array is lower than their threshold for detecting the change in the size of a single item or in the average size of a heterogeneous array (Chong & Treisman, 2003). While this suggests that adults automatically compute average element size, infants, who fail to detect a change in element size even under conditions of minimal external noise (e.g. zero variation in element size), clearly do not. Infants' failure to detect a twofold change in element size suggests the possibility that 6-month-olds may be less accurate in representing any element feature when that element is part of an ensemble than when it is presented alone. Alternatively, adults' computations of ensemble element size, specifically, may rely on representations not available in infancy.

When the adult visual system encounters an ensemble of elements, it automatically computes some statistical summary representations using texture processing, in which early feature information is pooled across regions without requiring the segmentation of individual objects (Dakin & Watt, 1997; Malik & Perona, 1990; Parkes et al., 2001). One well-established case of such a computation is the extraction of average orientation from an ensemble of tilted lines where individual elements are too crowded to allow for the discrimination of individual orientations (Dakin & Watt, 1997; Parkes, et al., 2001). It has been argued that average size is computed via a similar mechanism (Ariely, 2001; Chong & Treisman, 2003; 2005a; 2005b), but this claim has been more controversial.

If computations of average size and average orientation rely on similar automatic, parallel, and global processes, then we would expect the developmental trajectories of these abilities to be similar. In Experiment 1 of the third paper of the current dissertation, I compare the difference in infants' acuity for size representations across single element and homogeneous, multi-element arrays with infants' acuity for orientation representations across single element

and homogeneous, multi-element arrays. If infants' discrimination thresholds for single-element and multi-element arrays vary differentially when either size or orientation is tested, it will suggest that these representations are supported by qualitatively different mechanisms and raise the question of why these mechanisms have such different developmental trajectories.

If automatic processes that derive from mechanisms similar to those subserving adults' representations of average orientation are available in infancy, infants' representations of element orientation should be more accurate for homogeneous arrays than for single item arrays because as information is pooled, the representation becomes less noisy. If the same is true for size – that is, if there are similar automatic processes involved in average size computation, then we should see better acuity for element size in homogeneous arrays than in single item arrays. Cordes and Brannon's (2011) results suggest that this is not the case, but does not necessarily refute the possibility. One caveat to previous comparisons of infants' size representations for single elements and elements in homogeneous arrays is that they were drawn across multiple studies using stimuli that varied in a potentially confounding way. For example, it is possible that the single Elmo faces used in Brannon et al.'s (2006) study were more interesting to infants than the homogeneous dot arrays used to test infants' acuity for multi-element size representations (Cordes & Brannon, 2011). Experiment 1 of paper 3 provides the first direct comparison of infants' size representations for single elements and homogeneous arrays using identical stimuli and a within subject design.

Additionally, while previous studies exploring infants' ability to detect a change in ensemble element size have used the habituation paradigm (e.g. Brannon et al., 2006; Cordes & Brannon, 2011), recent evidence suggests that the change detection paradigm may be more sensitive for testing the development of this ability. In an infant change detection task, a display

is briefly presented (e.g. an array of three colored squares for 500 ms), then after a brief retention period (300 ms), a new array that may or may not contain a change is presented (e.g., one of the squares is a different color). Infants' preference for stimulus streams involving some type of change over streams in which there is no change is taken as evidence that the infants have encoded the altered property of the objects in visual working memory (e.g., Ross-Sheehy et al., 2003).

It is possible that some features of arrays are more salient during the prolonged exposure time inherent in habituation tasks, and thus, previous studies may have biased infants to encode other properties of arrays at the expense of ensemble element size. For example, although previous studies using the habituation paradigm have found young infants unable to discriminate small ( $< 4$  items) arrays on the basis of numerosity (e.g., Clearfield & Mix, 2001; Feigenson et al., 2002a; Feigenson et al., 2002 b; Xu, 2003; Xu et al., 2005), recent results from a change detection task by Starr, Libertus and Brannon (2013), demonstrate that 6-month-olds are capable of making purely numerical discriminations over small arrays. Starr et al., suggest that the critical difference between their change detection study and previous habituation studies is the degree of attentional load involved in the two types of tasks.

In a study by Hyde and Wood, (2011), attentional load was shown to affect the way adults attended to and encoded numerosity information from multi-item arrays (Hyde & Wood, 2011). In low attentional load conditions, individual items in an array were spaced within the resolution of spatial attention such that each item could be individuated. Under these conditions, adults performing a numerosity change detection task exhibited neural correlates of object file representations, suggesting that they were using focused attention processes to encode individual elements. In high attentional load conditions, elements were either spatially crowded such that

they could not be individuated (Exp. 1), or spaced within the resolution of spatial attention but presented with a concurrent task in which participants were required to monitor dual rapid serial visual presentation (RSVP) streams while performing the numerosity change detection task (Exp. 2). When attentional load was high, adults exhibited neural correlates of the approximate number system (ANS), the ratio dependent system responsible for discrimination of large arrays (e.g. Dehaene, 1997), suggesting they were using distributed attention to encode the array as a whole. It is possible that while the habituation paradigm promotes the encoding of individual elements, the change detection paradigm, which heavily loads attention, inhibits the representation of individual object files and instead promotes ensemble representations. Given that 7-month olds may only hold one or two individual object representations in working memory simultaneously, (Kaldy & Leslie, 2003; 2005; Moher, et al., 2012; Ross-Sheehy, et al., 2003) the habituation paradigm previously employed to study ensemble element size representations may have biased infants to engage the object file system instead of a more efficient distributed attention process.

If average element size representation relies on a distributed attention process that is available early in development, infants may have failed to detect a twofold change in size in previous studies (Cordes and Brannon, 2011) because the habituation methodology biased them to attempt to represent more object files than allowed under their working memory capacity limits. In Experiment 1 of the third paper of this dissertation, I use a change detection methodology and large arrays (5 items) to explore 7-month-olds' ability to represent homogeneous ensemble element size and orientation. Alternatively, if average ensemble size representations rely on a sampling process that requires focused attention to individual objects, infants may have failed in previous studies because the number of items in the arrays exceeded their working memory capacity and they were unable to select and encode a subsample of

elements within their capacity limits. In Experiment 2 of this paper, I use a discrepant color cue to highlight a single element within homogeneous size and orientation arrays. Given previous evidence that adults represent average orientation via a texture processing mechanism wherein no individual elements are attended, this manipulation is unlikely to affect the accuracy of their ensemble orientation representations. If the process underlying element size representations relies on focused attention to individual items and infants' failure to encode element size in previous studies of this ability was due to an inability to attend to a single element, this manipulation may allow them to successfully detect a twofold change in ensemble element size. Infants patterns of success and failure in paper 3 will shed light on continuities and discontinuities in the mechanisms underlying extraction of ensemble properties over development.

Taken together, the research included in this thesis characterizes continuities in working memory architecture from infancy through adulthood and highlights developmental changes in the way information maintained in working memory may be manipulated. In the first two papers, I extend our understanding of collection representations for dynamic events that unfold over time, and characterize continuities and discontinuities in the processes underlying this type of working memory representation. In the third paper, I review evidence of the continuities in working memory capacity from infancy through adulthood, and highlight potential discontinuities in the mechanisms underlying efficient ensemble processing. In my concluding remarks, I will recap the novel findings of my work and suggest future directions for studies addressing the primary question of how efficient coding of collections changes with the development of domain general cognitive abilities.

## PAPER 1

### Using Statistical Regularities to Increase Working Memory Representations of Approximate Number

#### ***Abstract***

Working memory can maintain a variety of types of data structures—including words, objects, and events. In all cases, sharp limits have been observed on its capacity. The present studies confirm that collections of individuals also constitute units of working memory, and that the limit of mutually exclusive collections that can be maintained in parallel is about three. Our studies also show that one way this limit can be circumvented is through the organization of individual items into hierarchically structured collections, and the exploitation of statistical regularities among these collections. We found that adult observers were able to spontaneously track the approximate numerosities of multiple collections of objects, where the objects were sequentially presented in temporally intermixed order too quickly to count. Observers' pattern of performance suggests that they first parsed the accumulating object sequences into two collections, then mentally embedded subordinate collections within these higher-level collection representations. Each level of the collection hierarchy was defined on the basis of a distinct visual feature. This manner of encoding allowed observers to greatly exceed the typically observed limits of working memory performance, and hence is an example of how efficient coding can radically enhance memory.

#### ***Introduction***

Planning and executing behavior, comprehending and producing language, formulating arguments, and generating causal explanations all require selecting relevant information from massive perceptual input, and maintaining that information in working memory so that it can be manipulated in the moment. As is well known, working memory capacities are severely limited, in most circumstances, to about four items (e.g., Cowan, 2001; Luck & Vogel, 1997; Scholl & Xu, 2001; Sperling, 1960). But the amount of information potentially relevant to our decisions often exceeds these limits.

There are many ways by which thinkers can circumvent the limits of working memory--most of them involving exploiting regularities in the input to devise a more efficient encoding scheme. For example, verbal and visual working memory can capitalize on statistical regularities among individual items to form “chunks,” which recode the stimuli into higher order units yet still preserve representations of the component items, and thus support enhanced retrieval of the individuals to be remembered. Representations of hierarchically coded chunks can be maintained more efficiently than representations of individual items, thereby freeing resources for further items (e.g., Verbal Working Memory: Baddeley, Thomson, & Buchanan, 1975; Burgess & Hitch, 1999; Chen & Cowan, 2005; Ericsson, Chase, & Faloon, 1980; Estes, 1973; Miller, 1956; Zhang & Simon, 1985; Visual Working Memory: Brady, Konkle, & Alvarez, 2009; Luck & Vogel, 1997). Alternatively, observers can represent collections of items in visual working memory that are selected as a group on the basis of spatial location or shared visual features such as color. These “ensemble” representations allow summary statistics to be computed about the entire ensemble as a single unit. For example, observers can represent the approximate number of birds in a flock, or the average size of the dots in an array containing very many (e.g., Alvarez, 2011; Brady & Alvarez, 2011; Chong & Treisman, 2003; Im &



Halberda, under review). Critically, these ensemble representations contain information about the entire array, but do not represent individual items contained therein (Ariely, 2001; Feigenson, 2011)—this is one way in which representations of ensembles are critically different from representations of chunks. Ensemble representations enhance memory for individual items through a process that replaces direct retrieval with inference based on ensemble statistics.

Adults can accurately estimate the numerosities of multiple ensembles in parallel, and given that numerosity is indisputably a property of an ensemble rather than a property of an individual item, this evidence confirms that an entire ensemble, rather than some sample of the items contained in the ensemble, can serve as a unit in visual working memory (Halberda, Sires, Feigenson, 2006). Further, ensemble representations can be hierarchically organized in terms of supersets and subsets. When shown arrays of spatially intermixed dots of two colors, both adults and infants spontaneously encode the approximate number of dots within each color subset, as well as the approximate total number of dots in the entire array—the superset (Halberda, et al. 2006; Zosh, et al., 2011). Additionally, information encoded about both supersets and subsets can be used to reconstruct information about individual items contained in the ensembles, even when the total number of individuals presented greatly exceeds working memory capacity limits. For example, observers' judgments of the size of a given dot within an array of dots is biased both by the mean size of the dots of the same color and by the mean size of all dots in the display, suggesting that the representation of the individual dot is constructed by integrating information across different levels of abstraction (Brady & Alvarez, 2011).

Ensemble representations clearly support the encoding of information in cases where working memory capacity limits would be otherwise exceeded (as shown by the simple “flock of birds” example). In our everyday lives, however, parsing the world into distinct collections is

not always so straightforward, for at least two reasons. First, collections of individuals are not always perceptually available as arrays in their entirety. Rather, they frequently accumulate over time, as when a stream of people emerges from a crowded theater. This necessitates updating collection representations dynamically over time. Second, since every individual has a large number of represented properties (e.g., gender, age, height), and these do not always cleanly parse the world into non-overlapping units, the criteria of collection membership are potentially unclear.

Feigenson (2008) probed the first of these issues, examining adults' ability to track collections that accumulated over time. Observers saw between one and five types of perceptually contrasting objects (toy pigs, poker chips, cotton balls, batteries, and green candies) placed one by one into each of two hiding locations. Different numbers of each type of object were placed in each location, with object types completely temporally intermixed, and all of the objects were placed in the first hiding location before any objects were placed in the second. To keep them from verbally counting, observers were instructed to perform verbal shadowing throughout the presentation. When the presentation ended, each hiding location contained either five or ten objects of each type, and observers were asked to indicate which location had more of a certain type of object (i.e., "Which bucket has more pigs in it?") and to verbally estimate how many objects of each type were in each location (i.e. "How many pigs would you say are in this bucket? And in this one?"). Because the object types were intermixed and objects were presented one at a time, participants did not know when they had seen the last member of a given collection in a given location, and therefore had to update their numerosity estimates for each collection in parallel as the information was presented.

This study revealed that, despite the dynamic and temporally interleaved presentation of the collections, observers knew which of the two locations had more of any object type, and their numerosity estimates were in line with characteristics of the non-verbal Approximate Number System (e.g., Dehaene, 1997). However, this was only true when three or fewer collections were presented—observers’ performance fell to chance on both the location identification and the estimation measure when observers were presented with four or five collections. Thus, as in some other studies on the limits of working memory (Feigenson and Carey, 2003; 2005; Feigenson, Carey & Hauser, 2002), observers’ failure was catastrophic when the demands of the task exceeded working memory limits. When too many object types were presented, observers were unable to track three of the four or five collections (for example, by ignoring one or two of the presented collections in order to successfully represent the remainder).

Feigenson’s (2008) study concerns a form of working memory that makes high executive function demands, as it requires flexible updating and controlled attention. As each object was presented, its features had to be compared to those of previously presented objects so that only the relevant collection representation was selected and updated or a new collection representation could be formed. Participants made these computations with no prior knowledge about what type of event would unfold, what objects would be involved, or how many types of objects or objects of each type there would be, suggesting that adults can flexibly select the bases for collection membership on the fly as events unfold.

Although Feigenson’s study differed from visual working memory studies in which all of the to-be-remembered items are presented simultaneously and the relevant ensembles are immediately apparent in a single glance, it was still similar to previous work on ensemble representations in visual working memory in that each object clearly belonged to only a single

ensemble or collection. Collection membership was unambiguous, both on the bases of early visual features and semantic category. For example, the collections differed from each in color (pink vs. blue vs. white vs. black/silver vs. green), shape (ovoid with articulated parts vs. disc vs. fluffy balls vs. elongated cylinders vs. cubes), and category membership. Attending to any of these features would lead observers to represent mutually exclusive, i.e., non-intersecting, collections, in which no object was a member of multiple collections at once. But what happens under more complex conditions, when individual objects can be represented as members of multiple different collections or groups? To illustrate, imagine the mundane case of doing laundry, where there are at least two important dimensions to consider in preparing the wash: color and material. We may first sort items of clothing by color in order to determine whether they should go in with the light or dark wash, and we may then sort the same items along the orthogonal dimension of material to determine whether they should be hung on the clothesline or put in the dryer. One possible solution in such cases where collection membership is multiply determined is to represent four mutually exclusive collections defined by the conjunctions of features from these two dimensions: dark and dryer-friendly, dark and line dry only, light and dryer-friendly, and light and line dry only. A clear drawback to doing this, though, is that even in this simple case in which there are only two orthogonal dimensions, the number of resulting collections already exceeds working memory capacity (Feigenson, 2008; Halberda et al., 2006; Zosh et al., 2011).

This example highlights the intuition that when the bases for collection membership are multiply determined or ambiguous, it may be disadvantageous to set such narrow criteria for membership. Given that previous work shows that only three mutually exclusive collections can be represented at once (Feigenson, 2008, Halberda et al., 2006), the ability to represent

collections using more flexible selection criteria may empower the encoding of more information than would be otherwise possible. To illustrate, we can return to the laundry example. The four mutually exclusive conjunction-based collections we discussed above might be more efficiently represented if an observer could first parse the array of laundry into two mutually exclusive collections along either one of the two critical dimensions (e.g., first parse the pile of laundry into two color-based collections: light-colored and dark-colored), and then represent the subsets within each of the two resulting collections (e.g., within the light-colored collection, dryer-friendly and line-dry). On this encoding scheme, each individual clothing item is a member of two orthogonal collections (e.g., an item is dark and dryer-friendly). Importantly, any statistical regularities among the collections may allow information to be represented in a more compressed, hierarchical format. For example, if we first sort the laundry along the color dimension, we may determine that more of the laundry is light than dark. During this same sorting process, we may detect that most of the light colored laundry is dryer friendly. From these two pieces of information, we may then be able to infer that most of the laundry, in total, is dryer friendly. Given that representing statistical regularities among individual objects can increase visual working memory capacity almost two-fold (Brady et al., 2009), sensitivity to statistical information among hierarchically organized collection representations might allow for an even greater expansion of memory.

In the current studies, we explore the ability to select and represent statistically structured feature-based collections, and the potential benefits of such hierarchical reorganization on working memory performance. In Feigenson's (2008) study, the collections shown to observers were mutually exclusive and unstructured: each object clearly belonged to only one collection, and the number of objects in any given collection provided no information about the number of

objects in any other. Here we explored collection representations under more complex conditions, asking whether flexible encoding strategies may be applied to intersecting structured collections, and whether these encoding schemas may allow for more than three collections to be represented at once. We then began to explore whether and in what circumstances collection encoding is flexible enough for statistical relations among collections to be discovered and exploited, allowing observers to surpass previously demonstrated working memory capacity limits.

Because our question presupposes that the working memory limit for mutually exclusive collections encoded in parallel is three, in Experiment 1 we first sought to replicate the previously observed failure to represent four collections (Feigenson, 2008) using stimuli similar to those used in our subsequent experiments. Furthermore, to ensure the observed limit is due to working memory capacity for collections, rather than any ambiguity about the basis of collection selection (i.e., the extent to which the objects could be neatly parsed into distinct collections), participants were shown one individual object of each type before the hiding event, and were told they would only see objects of these types. The distribution of collections across the two hiding locations was the same as that in Feigenson (2008). If there is indeed an upper limit of three mutually exclusive unstructured collections that can be represented in working memory, then observers should fail to represent the numerosities of any of the presented collections in Experiment 1, as they did in the four-collection condition of the studies by Feigenson (2008).

***Experiment 1: Capacity Limits for Mutually Exclusive Unstructured Collections.***

## ***Method***

### ***Participants***

Sixteen adults (5 males; mean age: 19.0 years; range: 18 to 21 years) who were native English speakers were recruited. Participants gave informed consent and received small prizes for participating.

### ***Materials***

Observers saw four types of stimulus objects, all cut out of green foam-board: circles, squares, S-shaped squiggles, and stars. The objects were approximately 1.5 inches in length and height. During the critical event, the objects were placed into each of two identical buckets that were opaque and approximately 10 inches in height.

### ***Procedure***

Participants sat at a table across from the experimenter. First, they were instructed in verbal shadowing (this was done to prevent participants from attempting to verbally count the objects as they were presented). For 30 seconds, participants practiced repeating a random letter sequence that had been recorded by a female speaker (0.77 letters/s), played through a stereo system. Participants were told to repeat each letter immediately upon hearing it, and to do so without stopping to correct any errors. All participants were judged to have performed the shadowing task successfully and so continued on to the next phase of the experiment.

Next, the experimenter told participants that they would see a short sequence of events while verbally shadowing, after which they would answer a few simple questions about the events. Participants resumed shadowing while they watched two brief practice events. In the

first, the experimenter placed a different toy under each of three cups, moved the cups around the tabletop (as in a shell game), and then revealed them. In the second event, the experimenter hid three toys behind two small screens, moved them around and then revealed them. Participants watched these practice events, which lasted approximately one minute in total, while continuously shadowing. Participants were not asked any questions about these events and were not told that these were practice events.

Before the critical hiding event, while participants continued verbally shadowing, the experimenter placed one object from each of the four mutually exclusive collections on the table and told participants, “I have objects of four types, some of them are stars, some of them are circles, some of them are squiggles, and some of them are squares.” The experimenter pointed to each object as she labeled it, making sure the participant was visually attending. After this, she cleared away the objects, leaving the table empty.

For the test event, the experimenter placed the two opaque buckets 80 cm apart on the table, then sequentially hid objects in each while participants continuously verbally shadowed. The experimenter took the objects from a container hidden under the table and held each object above a bucket before placing it inside, making sure participants were attending to each object. One object was placed approximately every two seconds, and objects of the four different types were placed in a randomized order so that they were completely temporally intermixed. The experimenter finished placing all of the objects in one bucket before moving on to the second bucket.

By the end of the test event, the two buckets contained an equal number of total objects, but the most frequent shape in each bucket differed. For each mutually exclusive collection, one



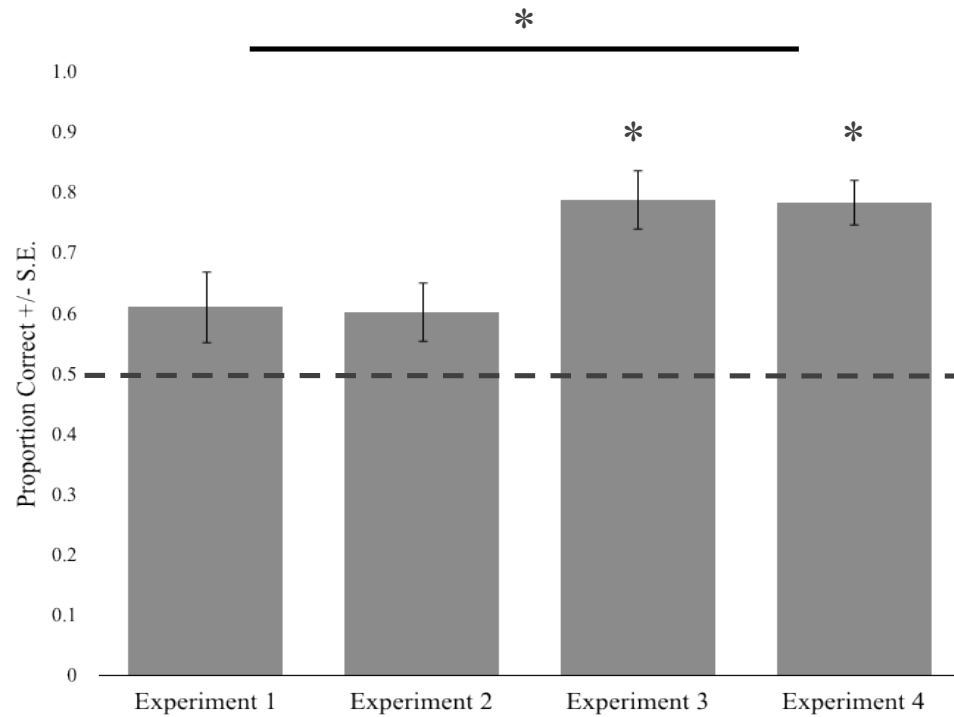
bucket contained twice as many objects as the other bucket (e.g., one bucket had 10 stars; the other had 5; one had 10 circles; the other had 5; See Appendix 1).

After placing all of the objects into the second bucket, the experimenter turned off the shadowing recording and asked participants which bucket had more objects of each of the four mutually exclusive collections (e.g., “Which bucket has more stars in it?”). The order in which the item types were queried was counterbalanced across participants. Participants’ answers were recorded by the experimenter with paper and pencil. Approximately 75% of testing sessions were videotaped to ensure reliability of procedures and coding among experimenters.

## ***Results***

Replicating Feigenson’s (2008) results, we found that participants were at chance in choosing which bucket had more objects from any of the four mutually exclusive collections, 61% across all four queries,  $t(15)=1.815$ ,  $p<.089$ . There was no effect of question order, (Q1: 9/16 correct, Q2: 10/16 correct, Q3: 11/16 correct, Q4: 9/16 correct,  $p=.87$ ), and performance did not vary by the type of object queried (circles: 9/16 correct, squares: 9/16 correct, stars: 11/16 correct, squiggles: 10/16, correct,  $p=.87$ , both Friedman tests). All reported tests are 2-tailed (Figure 1).

**Figure 1**



*Fig. 1. Proportion of correct responses across all questions asked in Experiments 1, 2, 3 and 4. The dashed line indicates chance performance. Participants were above chance in choosing the correct bucket in Experiment 3 (78.8%) and in Experiment 4 (78.3%). Performance in Experiment 2 (60%) was significantly different from performance in Experiment 3 and Experiment 4.*

## ***Discussion***

In Experiment 1, we replicated the finding that adults are unable to represent the relative numerosities of four mutually exclusive collections that accumulate in parallel. It appears that when participants' working memory capacity was overloaded, they were unable to recover information about any of the collections they had seen.

In Experiment 1, we used four mutually exclusive collections each consisting of a different, simple shape, and found that participants failed to encode the relative numerosities of any of these collections. Next, in Experiments 2 - 4, we asked whether presenting participants with structured collections that were statistically co-dependent would increase the amount of information they could store in working memory. In these next experiments, we used stimuli that could be reorganized into intersecting feature-based collections along multiple dimensions (akin to the laundry example given earlier). Consider four mutually exclusive conjunction-based collections: red triangles, blue triangles, red circles and blue circles. These four collections cannot each be selected and encoded as such in working memory, given the capacity constraints demonstrated in Experiment 1 and in the experiments by Feigenson (2008). However, if participants can instead represent these collections hierarchically—e.g., by representing the two color-based collections of red things and blue things, and then representing two subordinate collections within each—e.g., the triangles and the circles, they may be able to represent information about all four collections in a more efficient way.

These are big “ifs.” We do not know how flexible participants will be in devising an efficient encoding scheme on the fly, and in a single trial. No study has explored whether participants can create working memory representations of structured, intersecting collections that accumulate over time, and whether representing nested collections in this way can be exploited to circumvent working memory limits.

Therefore, in Experiment 2, we asked whether participants could spontaneously reorganize mutually exclusive conjunction-based collections with overlapping features into fewer feature-based collections, and use the statistical regularities among these features to increase working memory performance. We altered the procedure from Experiment 1 in two

ways. First, participants were not made aware of the object types before the critical hiding event. Second, instead of querying participants about mutually exclusive conjunction-based collections (e.g., the red triangles), we asked them about the relative numerosities of the feature-based collections along each dimension (e.g. all triangles, all circles, all red shapes, all blue shapes). Notice that if participants can answer questions about all four feature-based collections as accurately as they answer questions about one, two, or three mutually exclusive collections (e.g., pigs, poker chips, etc.) in the studies by Feigenson (2008), then they have exceeded the limits found in that study and in Experiment 1, for they have recovered information about four collections.

### ***Experiment 2: Intersecting Collections***

We asked whether participants could increase working memory performance on a task in which four conjunction-based collections could be reorganized into intersecting feature-based collections using statistical regularities among the features. We explored this possibility across multiple feature dimensions, using intersecting collections of shape and size in Experiment 2a, and of shape and color in Experiment 2b.

In Experiments 2a and 2b, every object presented could be conceived of as a member of two collections—a collection based on shape, or a collection based on either size (Expt. 2a) or color (Expt. 2b)—this was different from the design of Experiment 1 and the experiments by Feigenson (2008), in which each object clearly belonged to just a single collection. In addition, we introduced statistical regularities between the feature dimensions. In Experiment 2a, the most common triangle was large and the most common circle was small; and in Experiment 2b, the

most common triangle was red and the most common circle was blue. This was true across both hiding locations in each experiment. If participants are able to identify and represent feature-based collections and encode the regularities along these feature dimensions, they should be able to form more efficient hierarchical working memory representations, and reconstruct the relative numerosities of all four feature-based collections. However, if participants are unable to reorganize information carried by the conjunction-based collections into feature-based collections or if they are able to represent feature-based collections but not the regular structure within each, they should fail in Experiment 2 since adults are unable to store representations of four independent collections in working memory (Feigenson, 2008; Halberda et al., 2006; Experiment 1).

## ***Method***

### ***Participants***

Participants were 40 adults (Experiment 2a:  $n=20$ , 5 male; mean age 20.0 years; range 18 to 22 years; Experiment 2b:  $n=20$ , 3 male; mean age 20.9 years; range 18 to 25 years). Participants were native English speakers, and gave informed consent and received course credit for participating.

### ***Materials***

See Appendix 1 for the exact structure of objects in each hiding location in Experiments 2a and 2b. The same practice objects and hiding buckets were used as in Experiment 1.

**Experiment 2a:** There were four types of objects: small triangles, large triangles, small circles, and large circles. Each object was cut out of colored foam-board. The small objects were approximately 1.5 inches in diameter, and the large objects were approximately 3 inches in diameter. Sixteen out of 24 triangles were large, and 16 out of 24 circles were small. Sixteen out of 24 objects in bucket A were triangles, and 16 out of 24 objects in bucket A were large. Sixteen out of 24 objects in bucket B were circles, and 16 out of 24 objects in bucket B were small.

**Experiment 2b:** There were four types of objects: red triangles, blue triangles, red circles, and blue circles cut out of colored foam-board. Each object was approximately three inches in height and width. Sixteen out of 24 triangles were red, and 16 out of 24 circles were blue. Sixteen out of 24 objects in bucket A were triangles, and 16 out of 24 objects in bucket A were red. Sixteen out of 24 objects in bucket B were circles, and 16 out of 24 objects in bucket B were blue.

### ***Procedure***

The procedure for Experiments 2a and 2b was identical to that of Experiment 1, except that participants did not see the four object types before the hiding event. Instead, the critical hiding event occurred immediately following the second practice event.

After all of the test objects had been presented, the buckets contained an equal number of total objects, but the most frequent shape and size (or color) in each bucket differed. For each feature-based collection, one bucket contained twice as many objects as the other bucket (e.g., one bucket had 16 circles whereas the other had 8; one bucket had 16 large things whereas the other had 8). Given the statistical regularities introduced among these feature-based collections, in Experiment 2a, the bucket with the greater number of circles, for example, also had the greater

number of small objects, and in Experiment 2b, the bucket with the greater number of circles, for example, also had the greater number of blue items (See Appendix 1).

After the last test object had been presented, the experimenter turned off the shadowing recording and asked participants a series of questions. For each feature dimension present (shape and size in Experiment 2a; shape and color in Experiment 2b), the experimenter queried one of the two possible feature-based collections (e.g., “Please point to the bucket that has more large things”), and then queried one of the two possible feature-based collections along the orthogonal dimension (e.g., “Please point to the bucket that has more circles”). If participants guessed that the buckets contained equal numerosities (which happened rarely), they were told that the buckets contained unequal numbers and were asked to adjust their answers; their final answer was included in the analysis. If participants were uncertain, they were told to give their best guess.

The bucket that contained more of each object type, the bucket with which the presentation began, and the order in which the feature-based collections were queried were all counterbalanced across participants. The stimuli were presented in a randomized order such that objects from the four conjunction-based collections in each experiment were temporally intermixed.

## ***Results***

**Experiment 2a:** Participants were not successful at tracking the approximate numerosities of the feature-based collections-- they were at chance in choosing which bucket had more triangles, circles, small objects, or large objects (53% correct,  $t(19) = .295$ ,  $p = .772$ , one

sample t-test). Because questions were counterbalanced, half of the participants first answered a question about objects of a specific shape (either triangles or circles), and half of the participants first answered a question about items of a specific size (either small or large). There was no difference in performance based on question order (Q1: 8/20 correct, Q2: 13/20 correct,  $p=.096$ ) or dimension queried (shape: 11/20 correct, size: 10/20 correct,  $p=.739$ , both Wilcoxon Signed Rank Tests).

**Experiment 2b:** Unlike participants in Experiment 2a, participants in Experiment 2b were above chance in choosing which bucket had more triangles, circles, red objects, or blue objects when accuracy was collapsed across both queries (67.5% correct,  $t(19)= 2.67$ ,  $p=.02$ , one sample t-test). Because questions were counterbalanced, half of the participants first answered a question about items of a specific shape (either triangles or circles), and half of the participants first answered a question about items of a specific color (either red or blue). There was no difference in performance based on question order (Q1: 13/20 correct, Q2: 14/20 correct,  $p=.763$ ) or dimension queried (shape: 13/20 correct, color: 14/20 correct,  $p=.763$ , both Wilcoxon Signed Rank Tests).

Performance in both Experiments 2a and 2b was lower than that observed in Feigenson's (2008) 1-, 2- and 3-object type conditions, in which adults reliably succeeded at indicating which of two locations contained the greater numerosity of any object type (13/16, 14/6 and 13/16, respectively). Although participants in Experiment 2b performed slightly better than participants in Experiment 2a, the two groups were not significantly different from each other (Experiment 2a: 53% correct, Experiment 2b: 67.5% correct,  $p=.253$ , Mann-Whitney Test). And when we collapsed across the two experiments, performance was not above chance (60% correct,  $t(39)=$



1.84,  $p=.07$ , one sample t-test), and was no better than performance in Experiment 1 ( $p=.993$ , Mann-Whitney Test; see Figure 1).

## ***Discussion***

Participants failed to reliably represent any of the four feature-based collections in either Experiment 2a or 2b, as shown by their inability to judge the relative numerosities of the collections across the two hiding locations. One possible explanation for this failure is that participants attempted to track the four mutually exclusive conjunction-based collections (e.g., small triangles, large triangles, small circles, and large circles), a strategy that would fail because it exceeds working memory limits under these circumstances (Feigenson, 2008). Alternatively, participants may have tried to represent the four feature-based collections (e.g., small objects, large objects, triangles, and circles) without reorganizing them into a hierarchical format. If so, participants would again have been unable to represent the regularities among the four collections because of limits on working memory.

In Experiment 3 we sought to tease these two possibilities apart by explicitly pointing out the feature-based collections before the hiding event began (as we did in Experiment 1). If participants in Experiment 2 failed because they did not notice that the objects could be reorganized into hierarchical feature-based collections, they might succeed if this information were made available before the collections began to accumulate.

## ***Experiment 3: Explicitly Indicated Intersecting Collections***

## ***Method***

### ***Participants.***

Participants were 40 adults (Experiment 3a:  $n = 20$ , 12 male; mean age 19.0 years; range 18 to 21 years; Experiment 3b:  $n = 20$ , 4 male; mean age 20 years; range 18 to 21 years).

Participants were native English speakers who gave informed consent and received course credit for participating.

### ***Materials.***

The materials were identical to those of Experiments 2a and 2b.

### ***Procedure.***

The procedure for Experiments 3a and 3b was identical to that for Experiment 2a and 2b, except that immediately before the critical hiding event, while participants continued verbally shadowing, the experimenter highlighted the intersecting feature-based collections. The experimenter laid out one exemplar object from each of the possible conjunctions (e.g., small triangle, large triangle, small circle, and large circle) and drew participants' attention to each.

For example, in Experiment 3a, the experimenter said, "I have objects of four types. Some of them are triangles, some of them are circles, some of them are big, and some of them are small." As the experimenter mentioned each feature, she pointed to the two objects that shared that feature. This introduction highlighted the feature-based collections, but provided no information as to the statistical dependence among features.

The numbers of objects of each type placed in the two hiding locations were as in Experiments 2a and 2b (see Appendix). As before, for each feature dimension present (shape and size in Experiment 3a; shape and color in Experiment 3b), the experimenter queried one of the two possible feature-based collections (e.g., “Please point to the bucket that has more large things”), and then queried one of the two possible feature-based collections along the orthogonal dimension (e.g., “Please point to the bucket that has more circles”).

## ***Results***

**Experiment 3a:** Participants were significantly above chance in choosing which hiding location had more triangles, circles, small objects, or large objects (72.5% correct collapsed across all queries,  $t(19) = 2.65$ ,  $p = .02$ , one sample t-test). Wilcoxon signed-ranks tests revealed no difference in performance based on question order (Q1: 14/20 correct, Q2: 15/20 correct;  $p = .69$ ) or dimension queried (shape: 14/20 correct, size: 15/20 correct;  $p = .66$ ).

**Experiment 3b:** Participants in Experiment 3b were above chance in choosing which hiding location had more triangles, circles, red objects, or blue objects, 85% correct collapsed across all queries,  $t(19) = 6.66$ ,  $p < .0001$ , one sample t-test (Fig. 1). Wilcoxon Signed Ranks tests revealed no difference in performance based on question order (Q1: 16/20 correct, Q2: 18/20 correct,  $p = .89$ ). However, there was a significant difference in participants’ responses to the two feature dimensions (shape: 14/20 correct, color: 20/20 correct,  $p = .014$ ). It is possible that color was more salient as a basis of collection selection than the other feature dimensions used in our task; participants’ inclination to parse the stream into collections of red and blue, and then to

represent the unequal distribution of triangles and circles within each color-based collection, may have influenced performance in Experiments 2b and 3b.

Overall, performance levels in Experiments 3a and 3b were not significantly different from each other (Experiment 3a: 72.5% correct, Experiment 3b: 85% correct,  $p=.45$ , Mann-Whitney Test). Performance collapsed across Experiments 3a and 3b was significantly above chance (78.8% correct,  $t(39)=5.72$ ,  $p<.0001$ , one sample t-test), and significantly different than performance in Experiment 2, (Experiment 3: 78.8% correct; Experiment 2: 60% correct,  $p=.01$ , Mann-Whitney Test). See Figure 1.

## ***Discussion***

Experiment 3 provides the first evidence that participants can represent feature-based collections in situations where the collections accumulate over time, and can use the statistical regularities among these collections to increase working memory performance. The difference in performance between Experiments 2 and 3 demonstrates that participants better remembered the collections' numerosities when the feature-based collections were explicitly pointed out prior to the hiding event. Even though participants were uninformed about the nature of the task, had no idea when the streams of objects being placed into the two locations would end, and were performing a concurrent verbal shadowing task while they watched the events unfold, they were able to answer with high accuracy which location had more small objects, large objects, triangles, or circles (Experiment 3a) or triangles, circles, red objects, or blue objects (Experiment 3b). Their performance in specifying which bucket had more of the objects in a given feature-based collection (78.8%, collapsed across shape/size and shape/color) is indistinguishable from

that of participants tracking objects of just a single type in a similar task (1-type condition, 81% correct, Feigenson, 2008).

The success observed in Experiments 3a and 3b indicates that participants were representing four collections concurrently. As participants did not know which dimension(s) they would be asked about (Small objects? Large objects? Triangles? Circles?), their ability to accurately answer questions about any two of the four feature dimensions suggests that they were indeed representing all four collections, or were able to reconstruct information about any of the four collections from working memory. Given that this success appears to exceed previously identified capacity limits in this paradigm, participants must have represented the collections in Experiments 3a and 3b differently than in previous experiments (Experiments 1, 2a, 2b, and those by Feigenson (2008)). How was this possible? Experiments 1 and 2 demonstrate that neither the act of seeing all four object types before the hiding event (Experiment 1) nor the mere presence of statistical regularities among the object types themselves (Experiment 2) led to an increase in working memory performance. Instead, it appears that both of these pieces of information were necessary. One way in which participants may have used this information to succeed in Experiment 3 is by recognizing that the mutually exclusive conjunction-based collections could be reorganized into hierarchical feature-based collections. That is, instead of representing the four exemplar objects shown prior to the hiding event as: small triangle, large triangle, small circle, and large circle, participants may have represented, for example: the collection of small objects (which could be further parsed along the dimension of shape) and the collection of large objects (which also could be further parsed along the dimension of shape). Having had the prior experience of viewing the objects and seeing their feature dimension pointed out, despite its brief nature, may then have allowed participants to represent the streams

of objects being placed into the two hiding locations in terms of nested feature-based collections, and then to detect and use the statistical regularities among these to surpass the three-collection capacity limit in working memory.

However, given the nature of the task and the simple binomial measure of accuracy, one potential concern is that participants succeeded in Experiment 3 by representing just three out of the four feature-based collections (e.g., representing the red objects, blue objects, and triangles, but failing to represent the circles)—leading to better than chance performance overall.

Participants' failure in Experiment 1 makes this account unlikely, since participants in that experiment also saw objects of all four types before the hiding event, giving them an equal opportunity to select just some subset of collections to track, and yet they failed to represent the relative numerosities of any of the queried collections.

Still, in Experiment 4 we put this possibility to a strong test, and explored the limits of the capacity to encode statistically structured intersecting feature-based collections in parallel. Given that adults can represent two feature-based collections in parallel, and are sensitive to the statistics among features on another, subordinate dimension (Experiment 3), it is possible that they may be able to track yet a third feature dimension as well among the two collections, as well. To test this, in Experiment 4 we presented participants with eight conjunction-based collections. One way to represent these is as six feature-based collections across three orthogonal feature dimensions (e.g., shape (triangles vs. circles), size (large vs. small objects), and topology (presence vs. absence of a topological feature—a hole)). If participants' reliance on a sampling strategy led to the success observed in Experiment 3, performance in Experiment 4 should clearly be lower. If, however, participants succeeded in Experiment 3 by capitalizing on the structured regularities among the feature-based collections to increase working memory

capacity, they may also succeed robustly in Experiment 4 despite the increase in the complexity of the stimuli.

#### ***Experiment 4: Exploiting the Structure Among Six Intersecting Feature-based Collections***

##### ***Method***

##### ***Participants.***

Participants were 40 adults (Experiment 4a:  $n = 20$ , 3 male, mean age 19.0 years; range 18 to 21 years; Experiment 4b:  $n = 20$ , 5 male; mean age 20 years; range 18 to 21 years). Participants were native English speakers who gave informed consent and received course credit for participating.

##### ***Materials.***

The materials in Experiments 4a and 4b were similar to that of Experiments 3a and 3b, with the addition of a third feature dimension. To create a third feature dimension, we increased the total number of objects in each bucket to 64 and added holes that were approximately .5 inches in diameter to half of the objects (for evidence that the presence / absence of a hole serves as a primary visual feature in object recognition, see Chen, 2005; Chen, Zhang, & Srinivasan, 2003). In Experiment 4a, the objects were small and large triangles and circles with and without holes. In Experiment 4b, the objects were red and blue triangles and circles with and without holes. For any given feature-based collection (e.g., large objects, circles, or objects with holes),

the numerical ratio of the quantities in the two hiding locations was approximately 1:2 (this varied slightly from 1:1.83 to 1:2.5 due to constraints of the design and the necessity of keeping the total number of objects equal across the two hiding locations).

**Experiment 4a:** There were eight types of objects: small triangles with holes, small triangles without holes, large triangles with holes, large triangles without holes, small circles with holes, small circles without holes, large circles with holes, and large circles without holes. There were six intersecting feature-based collections: small objects, large objects, objects with holes, objects without holes, triangles, and circles. All of the shapes were cut out of blue foam-board. All small objects were approximately 1.5 inches in height and width, and all large objects were approximately 3 inches in height and width. Most of the objects in Bucket A were large circles with holes in them, whereas most of the objects in Bucket B were small triangles without holes. Twenty-four out of 64 objects in Bucket A were large circles with holes and 24 out of 64 objects in Bucket B were small triangles that lacked holes. See Appendix 1 for the exact numerosities by feature, and for the distribution of objects across the two hiding locations.

**Experiment 4b:** There were eight types of objects: red triangles with holes, red triangles without holes, blue triangles with holes, blue triangles without holes, red circles with holes, red circles without holes, blue circles with holes, and blue circles without holes cut out of colored foam-board. All of the objects were approximately 3 inches in height and width. Most of the objects in Bucket A were blue circles with holes in them, while most of the objects in Bucket B were red triangles without holes. Twenty-four out of 64 objects in Bucket A were blue circles with holes and 24 out of 64 objects in Bucket B were red without holes. See Appendix 1 for the exact numerosities by feature, and for the distribution of objects across the two hiding locations.



### ***Procedure.***

The procedure was the same as in Experiments 3. As in Experiment 3, before the hiding event, one member of each conjunction-based collection was placed on the table and all of the intersecting feature-based collections were pointed out (e.g., “I have items of eight types. Some things have holes, some things do not have holes; some things are circles and some things are triangles; some things are red and some things are blue,” indicating all four items of each type as each feature was mentioned). As in Experiments 2 and 3, after the hiding event the experimenter queried one of the three possible feature-based collections (e.g., “Please point to the bucket that has more circles”), then queried another feature based collection along an orthogonal dimension (e.g., “Please point to the bucket that has more large things”), and finally queried a the last remaining feature-based collection (e.g., Please point to the bucket that has more objects with holes”). The order in which the feature dimensions were probed, and the particular feature-based collection queried was counterbalanced across participants.

### ***Results***

**Experiment 4a:** Participants in Experiment 4a were significantly above chance in choosing which hiding location had more triangles, circles, red objects, blue objects, objects with holes, and objects without holes, collapsed across all three questions (71.67% correct,  $t(19)=3.90, p=.001$ , one sample t-test). There was no difference in performance based on question order (Q1: 14/20 correct, Q2: 16/20 correct, Q3: 13/20 correct,  $p=.58$ ). There was a marginal effect of dimension queried (shape: 11/20 correct, size: 14/20 correct, topology: 18/20 correct,  $p=.06$ , both Friedman Tests), with participants performing better on questions about

feature-based collections along the topology dimension than those on the shape or size dimensions.

**Experiment 4b:** Participants in Experiment 4b were above chance in choosing which hiding location had more triangles, circles, red objects, blue objects, objects with holes, and objects without holes, collapsed across all three questions (85% correct,  $t(19)=6.84$ ,  $p<.0001$ , one sample t-test). There was no difference in performance based on question order (Q1: 17/20 correct, Q2: 17/20 correct, Q3: 17/20 correct,  $p=.58$ ). In contrast, we did observe a significant effect of dimension queried, (shape: 16/20 correct; color: 20/20 correct; topology: 15/20 correct,  $p=.05$ , Friedman test). Participants performed better on questions about feature-based collections along the color dimension (as in Experiment 2b)) than along the shape or topology dimensions.

Overall, performance in Experiments 4a and 4b did not differ (Experiment 4a: 71.7% correct; Experiment 4b: 85.0% correct,  $p=.10$ , Mann-Whitney test). Collapsed, performance in Experiment 4 was significantly above chance (78.3% correct,  $t(39)=7.31$ ,  $p<.001$ , one sample t-test), was significantly better than performance in Experiment 2 (Exp 4 correct: 78.3%, Exp 2 correct: 60.0%,  $p=.01$ ), and was not different from performance in Experiment 3 (Exp 4: 78.3% correct; Exp 3: 78.8% correct,  $p=.62$ , both Mann-Whitney tests).

## ***Discussion***

If participants in Experiment 3 had used the strategy of representing just three out of the four mutually exclusive conjunction-based collections, and participants in Experiment 4 had used the same strategy (storing and updating representations of three out of eight such collections), we

should have observed a sharp decline in performance in Experiment 4 relative to that in Experiment 3. Similarly, if participants had succeeded in Experiment 3 by sampling three of four feature-based collections, and had attempted the same strategy in Experiment 4 (storing and updating representations of three out of six such collections), we also should have observed a sharp decline in performance. Instead, participants clearly succeeded in Experiment 4, performing at the same level as in Experiment 3-- essentially ruling out any account involving sampling a subset of the available information to reach above-chance performance.

Together, the results of Experiments 3 and 4 show that the working memory limits that constrain the number of mutually exclusive collections that can be represented in parallel do not straightforwardly apply in cases where collections are intersecting and statistically structured. In Experiment 4, 128 items from eight different mutually exclusive conjunction-based collections were placed one at a time-- 64 into one location and then 64 into the other. The order of placement of object types was random. Although participants had some foreknowledge of what and how many feature-based collections they would see, they did not know how many objects would have each feature, nor did they know the statistical regularities among features. To compute these statistics, they either needed to keep track of eight mutually exclusive conjunction-based collections in parallel, or to represent the feature-based collections in a more efficient way that does not exceed known working memory limits. The results of Feigenson (2008), Experiment 1, and Experiment 2 (and, less directly, those of Halberda et al., 2006 and Zosh et al., 2011) appear to rule out the first possibility. This suggests that the successful performance in our experiments rested on the ability to represent intersecting feature-based collections on the basis of binary features along three separate orthogonal dimensions (e.g., shape, color, and topology), and, over time, to detect statistical dependencies among these

features. Doing so appears to have led to participants to indirectly but efficiently represent six feature-based collections (or eight mutually exclusive conjunction-based collections) in parallel.

### ***General Discussion***

Along with the studies by Feigenson (2008), the present studies extend our understanding of collection (i.e., ensemble) representations in visual working memory to a different kind of working memory—one for dynamic events that unfold over time. The primary conclusion of this work is that collections themselves can serve as the units of working memory. In Feigenson’s (2008) study, participants were able to represent the approximate numerosities of one, two and three collections in parallel without any reduction in accuracy as the number of collections increased. Participants seeing events involving collections of toy pigs alone, collections of pigs and batteries, or collections of pigs, batteries and cotton balls, could specify which bucket had more, e.g., pigs, with 80% accuracy (13/16, 14/16 and 13/16 correct, respectively). Since “more pigs” quantifies over collections, and since participants did not know which collection would be probed (or even what question would be asked), this finding requires that participants encoded and updated representations of all three collections and their relative numerosities in each of the two hiding locations.

Feigenson (2008) found that performance fell apart when four or more collections were presented. Experiment 1 in the present series confirms this finding. Even when shown in advance the four types of objects that would later comprise the mutually exclusive collections (stars, circles, squares, and squiggle shapes), participants were unable to represent the four collections in parallel. Thus, collections can be units of working memory-- and as with other

units of working memory (e.g., objects, locations, words, events), there is a sharp capacity limit on the number of units that can be maintained in parallel.

Many studies of working memory show that participants can exploit statistical regularities in the input to recode information in a more efficient manner, thereby freeing resources to encode further items (e.g., Miller, 1956; Chen & Cowan, 2005; Kibbe & Feigenson, under review; Luck & Vogel, 1997; Brady et al., 2009). Experiments 3 and 4 demonstrate one new way in which this is so-- here in the case of memory for collections that are presented over time. In Experiment 3, participants exceeded the three-item limit on working memory, representing the relative numerosities of four feature-based collections (e.g., small objects, large objects, triangles, and circles) in two locations, and at the same level of success as participants presented with one, two or three collections in the studies by Feigenson (2008). Further, participants in Experiment 4 successfully represented six feature-based collections (and see footnote 1 for an extension of this result to eight feature-based collections). Importantly, Experiment 2 showed that merely having statistical regularities present in the input was not sufficient to lead to this improved memory for collections. Therefore, as we consider how participants may have succeeded at this demanding working memory task, we must ask not only how they might have exploited the statistical regularities in the input, but also why first highlighting the stimuli's overlapping features (the variable that distinguished Experiments 2 and 3) helped them do so.

Recall that in Experiment 2, participants were unable to successfully compare the relative numerosities of the four feature-based collections, suggesting that they may have been trying to encode the objects as members of four mutually exclusive conjunction-based collections. As four collections exceeds the capacity limits previously demonstrated in this task, participants' failure

under this encoding scheme would not be surprising. It is possible that participants did not realize that the objects could be represented in terms of hierarchically structured feature-based collections, and thus had no basis for using a more efficient encoding scheme. Below we offer some speculations about what the more efficient encoding scheme might be, and how it may have allowed participants to so dramatically improve their working memory storage.

Examining the stimuli for Experiments 2 and 3 (which were identical) and for Experiment 4 (see Appendix 1), it is clear that if participants encoded only the modal object in each hiding location (e.g., in Experiment 2a, the most common object in location A was a large triangle and the most common object in location B was a small circle), along with the fact that the total number of objects in each location was approximately equal, and that each of the other three conjunction-based collections in each location were roughly equinumerous, they could have correctly affirmed that there were more large objects and triangles in location A and more small objects and circles in location B. This is true, but how were participants able to encode information about the modal object in each location, plus enough information about the equal numerosity of each of the other three conjunction-based collections, to support this inference? Recall that participants never saw the collections all at once, and did not see the objects organized in time as they are organized in space in Appendix 1. The “modal object” is characterized by a conjunction of features. If participants tracked each of the mutually exclusive conjunction-based collections, the modal object can only be identified by comparing the relative numerosities of the four (or eight in Experiment 4) mutually exclusive collections in working memory, a feat that exceeds the capacity limits of working memory as demonstrated by Feigenson (2008), Halberda et al. (2006), and Experiment 1 of this paper.

An alternative strategy, and one we consider more likely, is that participants in Experiments 3 and 4 used top-down information to set up a hierarchical encoding scheme to represent the collections in each hiding location. On this account, after the experimenter highlighted the features of the objects prior to the hiding event, participants would select the one featural dimension to serve as the bases for two feature-based collections that compose the top level of the hierarchy (e.g., could select color, thereby parsing the objects into red objects and blue objects). They then would represent these two feature-based collections in each hiding location (e.g., red objects and blue objects in location A, and red objects and blue objects in location B), and represent the approximate numerosity of each of the two collections in each of the two locations (e.g., about 16 red objects and about eight blue objects in location A; about eight red objects and about 16 blue objects in location B). We know from previous studies using this paradigm that adults can accurately represent the approximate numerosity of up to three collections in each of two locations in parallel (Feigenson, 2008), and therefore that doing so for two collections is well within their capacity. The hierarchy model then requires that participants also represent information about the relative values along the orthogonal dimension(s) that are lower on the hierarchy (e.g., shape), nested within representations of each of the superordinate collections (e.g., red objects, blue objects). Frequencies of the feature-based collections along the secondary (and tertiary) dimension(s) may be represented in terms of average exemplars that are continually updated throughout the hiding sequence (e.g. the average red object is closer in shape to a triangle than to a circle), or by ratio information (e.g., more of the red objects were triangles than circles). Either representation could then be combined with the estimated numerosity of the superordinate collection to compute relative numerosity judgments of the secondary feature-based collections across locations.

That participants in our experiments apparently privileged different features is consistent with such a hierarchical encoding scheme. Performance was better (though not statistically so) in Experiments 2b and 3b than in Experiments 2a and 3a, and this was due to statistically reliable, better performance when collections could be defined using color. It is not just that participants selected collections to encode on the basis of color, however (although this may be the case for Experiment 2b), because they were also well above chance when queried about the other, non-color dimension. Rather, we suggest that the better performance may have been due to color's salience as a sorting cue, which participants then used to select feature-based collections of red and blue items as superordinate collections in each location, while still representing a running average or proportion calculation with respect to the orthogonal features (e.g., proportion of triangular vs. circular red objects).

In Experiments 4a and 4b we also observed a performance difference that may reflect the relative salience of the feature dimensions chosen as the basis of establishing the two collections at the top level of the hierarchy. In Experiment 4a, participants were most accurate at answering questions about feature-based collections of object with or without holes, and in this case color was not a possible basis for collection selection. In Experiment 4b, performance was again most accurate for questions about feature-based collections defined by color, and still above chance for other dimensions as well.

The hierarchical model we offer here makes several predictions. The first concerns the fact that we have not yet determined the upper limit of working memory capacity for fully intersecting structured collections. If the hierarchical model is correct, then there is no a priori reason that performance on our task should fall apart with the addition of a fourth feature dimension. Indeed, this prediction is borne out. Participants in an experiment not reported here,



who saw 16 mutually exclusive conjunction-based collections that varied binomially across the four feature dimensions used in this series of studies (shape, size, color, and topology), were able to correctly judge the relative numerosities of all eight intersecting feature-based collections (Yamaguchi, Tuerk & Feigenson, 2009<sup>1</sup>).

The hierarchical model also predicts that we might be able manipulate performance on this task by influencing participants to set up the hierarchy with a particular feature dimension at the superordinate level. For example, in an analogous study of both children and adults, instead of shadowing, we instructed participants to label each object placed into a hiding location according to the features along one of the dimensions (e.g., “triangle, circle, circle, triangle...” OR “small, small, big, big, ...”). If this manipulation indeed affects the hierarchical encoding scheme, then performance should be best when probed on the labeled dimension, but still above chance on the unlabeled dimension. And indeed this is so: while older children (seven-year olds) and adults, who performed equivalently, successfully answered questions both congruent and incongruent with the labeled dimension (congruent: 28/28 correct,  $p < .001$ ; incongruent: 25/32 correct,  $p = .002$ , both two-tailed binomial tests), their performance was significantly better in the congruent condition ( $p = .01$ , Fisher’s exact test). Collapsed performance across the incongruent and congruent conditions for adults and older children was 88.3%, similar to the accuracy levels observed in Experiment 3b of this paper, consistent with the hierarchical encoding model. Younger children between the ages of four and six, succeeded only in the congruent condition,

---

<sup>1</sup> We have run this experiment twice, because we were surprised at the success. Across a total of 42 participants, overall performance is 73% correct, ( $p < .01$ , sign test), well above chance, with performance best on the color dimension. Q1) 30/42 correct, Q2) 35/42 correct, Q3) 27/42 correct, Q4) 30/42 correct. Shape) 28/42 correct, Size) 28/42 correct, Color) 35/42 correct, Topology) 31/42 correct (Yamaguchi, Tuerk & Feigenson, talk presented at VSS in 2009)

suggesting that there are additional processing demands when numerosity estimates of the subordinate (unlabeled) collections must be computed (Tuerk & Carey, in preparation).

Of course, one important issue to note is that the task given to participants here (and in the studies by Feigenson, 2008) differs important ways from previous studies of parallel enumeration in visual working memory (Halberda et al., 2006, Poltoratski & Xu, 2013), in both the duration of the memory representation, and in the kinds of representations available to participants. The kind of working memory involved in our paradigm likely makes heavy demands on executive function. It obviously involves the updating of currently held working memory representations, because information accumulates over time and in an unpredictable order. In addition, in order to succeed in Experiments 3 and 4, participants needed to use the information they were given prior to the hiding event to set up representations of intersecting feature-based collections, and likely needed to inhibit the alternative, and less efficient encoding scheme of representing four mutually exclusive conjunction-based collections, which Experiment 2 suggests is spontaneously deployed but cannot support success on this task. Developmental studies may shed light on the role of strategic control in the current studies. Young children share much of the basic architecture of working memory with adults (Cowan, 1997; Feigenson & Carey, 2003, 2005; Feigenson & Halberda, 2004, 2008; Moher, Tuerk & Feigenson, 2012; Oakes & Bauer, 2007; Zosh & Feigenson, 2009), but are limited in the ability to strategically and flexibly establish efficient encoding schemes in working memory. The ability to flexibly attend to multiple features of a stimulus (e.g., color and shape), and the ability to sort items based on abstract categories are both necessary for success in this task, and develop significantly over the preschool years (e.g., Zelazo et al., 1996; Kirkham et al., 2003, and Kharitonova, Chien, Colunga & Munakata, 2009, respectively). We have recently found that

whereas like adults, three- to five-year old children can represent three mutually exclusive collections in parallel, the ability to use the statistical regularities among intersecting collections to increase the total amount of remembered information does not appear until about age seven (Tuerk, & Carey, in preparation; poster 2011).

In summary, the present experiments demonstrate that adults are capable of selecting and representing intersecting feature-based collections, and using statistical regularities among these to increase the amount of information maintained in working memory. Much further work remains to be done in order to determine the particular encoding schema participants use to succeed in the given task. Future research exploring the role of top-down information in the selection of optimized encoding scheme will shed light on the flexibility of working memory and its capacity limitations.

## PAPER 2

### Hierarchical Working Memory Representations of Collections Over Ontogeny

#### ***Abstract***

Reorganization of information in working memory appears to be a fundamental and early-emerging solution to the challenge of storing large amounts of information in a strictly limited system. In addition to encoding chunks of items, where information about the individuals is maintained, both infants and adults can also store summary information about large ensembles of items in visual working memory. The present studies confirm that collections of individuals also constitute units of a different type of working memory that makes updating demands on executive function. They demonstrate for the first time that like adults, young preschoolers can represent and update three mutually exclusive collections that accrue in an intermixed fashion over time. They also demonstrate that adults can surpass this capacity limit by reorganizing individuals into hierarchically represented collection structures and exploiting statistical regularities among the collections across levels of the hierarchy. The ability to reorganize individuals in this manner to circumvent capacity limits and track the approximate numerosities of multiple nested collections develops between the ages of three and seven.

#### ***Introduction***

As is well known, working memory capacities are severely limited, in most circumstances, to about four individuals or so (e.g., Cowan, 2001; Luck & Vogel, 1997; Scholl & Xu, 2001; Sperling, 1960). The signature capacity limits of working memory are shared by a variety of non-human animals and human infants (Non-human animals: Hauser, Carey, & Hauser, 2000; Uller & Lewis, 2009; Infants: Feigenson & Carey, 2003, 2005; Feigenson, Carey & Hauser, 2002; Ross-Sheehy, Oakes & Luck, 2003; Adults: Alvarez & Cavanagh, 2004; Broadbent, 1975; Cowan, 2001; Jiang, Olson, & Chun, 2000; Luck & Vogel, 1997; Sperling, 1960; Y. Xu, 2002), suggesting great continuity in working memory processes over phylogeny and ontogeny. In addition to having similar capacity limits on the number of separate items we can store in working memory, both infants and adults have the ability to surpass these limits to some extent by encoding to-be-remembered items in a variety of working memory data structures.

Across multiple methodologies, infants and adults show a similar strict capacity limit on the number of individual items they can hold in working memory. For example, in a change detection task, adults' performance was at ceiling for arrays of one to three items regardless of the items' complexity, but declined sharply when arrays contained four or more items, revealing that WM could store 3 or 4 items, but no more. (Luck & Vogel, 1997). In an analogous experiment, 10- and 13-month-old infants saw two flickering streams of arrays of colored squares where one stream displayed the same items across each 500 ms flicker, and the other stream contained one item that alternated in color between flickers. Infants looked reliably longer at the Changing than the Non-Changing Stream when the array contained 1, 2, 3, or 4 squares, however, they failed to look longer at the Changing Stream when the array contained 6 squares (Ross-Sheehy, Oakes, & Luck, 2003), suggesting that infants could remember the features of

four, but not six items at once. Infants show similar capacity limits when to-be-remembered items are presented sequentially, successfully searching for one, two or three items they see sequentially hidden in a box, but failing when four items are hidden (Feigenson & Carey, 2003; 2005). The similar capacity limits observed across a wide variety of methodologies in 10- to 20-month-old children and adults suggests that working memory capacity for individual items may be consistent across development.

In addition to having similar capacity limits on the number of separate items we can store in working memory, both infants and adults have the ability to surpass these limits to some extent by encoding to-be-remembered items as collections, exploiting statistical regularities in the input or conceptual information to devise a more efficient coding scheme. For example, the letter string ABCCNNCBS is much easier to recall when we recognize the three familiar television acronyms ABC, CNN and CBS that comprise it. This hierarchical reorganization of individual items based on previously acquired conceptual knowledge results in the representation of two nested levels: the chunks (the acronyms) and their components (the individual letters therein). Representations of hierarchically coded chunks can be maintained more efficiently than representations of individual items, thereby freeing resources for further items (e.g., Verbal Working Memory: Miller, 1956; Ericsson, Chase, & Faloona, 1980; Chen & Cowan, 2005; Baddeley, Thomson, & Buchanan, 1975; Burgess & Hitch, 1999; Estes, 1973; Zhang & Simon, 1985); Visual Working Memory: Luck & Vogel, 1997; Brady, Konkle, & Alvarez, 2009).

Fourteen-month old infants can also efficiently reorganize information into hierarchically organized chunks to bypass working memory capacity limits. While infants can only remember three items hidden at once in the absence of any grouping cues, when given perceptual, conceptual, linguistic or spatial cues to grouping, 14-month olds successfully remember and

search for four items in the same paradigm (Feigenson & Halberda, 2008; Rosenberg & Feigenson, 2013). In fact, even at 7-months of age, before working memory has matured to the three-item limit, infants can already use multiple locations and shared features among items to form chunks and store more total items in working memory (Moher, Tuerk & Feigenson, 2012).

In addition to chunks, which maintain the individuals within a collection, both infants and adults can represent collections of items in visual working memory called “ensembles,” which are selected on the basis of spatial location or shared visual features such as color. Critically, ensemble representations contain summary information about the entire array, but do not represent individual items contained therein (Ariely, 2001; Feigenson, 2011). For example, adult observers can represent the average size of the dots in an array containing very many dots without explicit knowledge of any individual member dot’s size (e.g., Alvarez, 2011; Brady & Alvarez, 2011; Chong & Treisman, 2003; Im & Halberda, under review). Ensemble representations are also available in infancy, as six-month-old infants can discriminate arrays of four items from arrays of eight items (Xu & Spelke, 2000) though they can store at most two individual objects at once (Wynn, 1992, Moher, et al. 2012). The ability to select and encode the numerosity of multiple collections in parallel has also been demonstrated in 9-month old infants. When presented with arrays of dots of two colors, both infants and adults are able to remember the numerosities of the two color subsets and the superset of all dots. When arrays contain three or more colors of dots, both infants and adults encode the numerosity of the superset of all dots, but do not remember the numerosity of any color subsets (infants: Zosh, et al., 2011; adults: Halberda, Sires & Feigenson, 2006). In addition to individual items and chunks, an ensemble may serve as a unit in working memory, enhancing memory for individuals through a process that replaces direct retrieval with inference based on ensemble statistics.

Although both infants and adults can successfully encode ensemble representations of multiple collections when they are presented simultaneously and the basis of selection is obvious, many of the collections we encounter in our everyday lives are not so clearly delineated. For one, collections of individuals are not always perceptually available in their entirety. When multiple collections accumulate over time, it necessitates the dynamic selection and updating of relevant collections as new items are presented. Additionally, individual items have a large number of represented properties that could potentially be relevant for determining collection membership, but these properties do not always cleanly parse the world into non-overlapping collections.

Without the presence of obvious grouping cues, the criteria of collection membership must be determined while information is still accumulating. Feigenson (2008), found that adults can accurately represent up to three collections of items that accumulate in this fashion. In these studies, participants each saw one hiding event, involving between one and five types of objects (pigs, poker chips, cotton balls, batteries, and green Starburst candies). Items were presented one at a time, for two seconds each, in a randomized order, and hidden into two buckets. Participants were engaged in verbal shadowing during the event, preventing them from counting, or even verbally encoding the items, and they did not know what questions would be asked. One stream of items was hidden in one bucket before the experimenter moved on to hide the second stream in the other bucket. At the end of this event, each bucket contained either 5 or 10 items of a particular type. After all objects were hidden, participants were asked to indicate which bucket had more of a certain type of object in it (i.e. “which bucket has more pigs in it?”). Since the types of items were intermixed and presented one at a time, participants did not know when they had seen the last member of a particular collection in each bucket, and thus were required to



update numerosity estimates for each collection in parallel. For presentations of up to three types of objects, participants' accuracy on the question of which bucket had more objects of a given type was above chance, but performance fell to chance when participants were presented with four or five object types.

Feigenson's (2008) study offers convergent evidence that a large collection of items may be stored as a unit in working memory, and suggests that for adults, the ability to encode summary representations of multiple collections is robust to prolonged hiding events where items are presented sequentially over time. While both infants and adults can represent multiple large collections of simultaneously presented items in visual working memory, it is possible that the mnemonic processes used in Feigenson's (2008) task are qualitatively different from those available early in development. To succeed in this task, participants must compare the features of each item presented to those of previously presented items held in a working memory buffer, so that they may update the relevant collection representation or form a new collection representation while the others are maintained simultaneously.

The cognitive abilities required to succeed in a task like Feigenson's, namely set shifting and updating, are often discussed as subcomponents of a larger construct called Executive Function (EF) which is known to develop significantly between the ages of three and seven (Diamond, 2013; Garon, Bryson, & Smith, 2008; Munakata et al., 2011; Zelazo, et al., 2003). It is possible that the working memory flexibility involved in shifting among and updating multiple collections in parallel is not available early in childhood while EF is still developing. That is, young preschoolers may fail in a task analogous to Feigenson's (2008) task where collection-based summary representations must be repeatedly selected and updated.

In Experiment 1, we test 3-and 4-year olds in a task analogous to Feigenson's (2008) adult task. In order to succeed in this task, children had to maintain representations of three collections in each of two locations, and flexibly shift among and updating the relevant collection by comparing each item presented to exemplars held in working memory. Some EF researchers (e.g. Diamond, 2006) have theorized that mental set shifting in childhood relies on working memory and inhibitory control processes that are undeveloped in three and four year olds. If young children are able to represent three mutually exclusive collections under sequential presentation, it will provide the first evidence that children can encode collection-based summary representations under these conditions and further, that the ability to encode, shift among, and update clearly delineated, homogenous collections in working memory emerges early in EF development.

### ***Experiment 1:***

#### ***Methods***

##### ***Participants.***

**1a:** Thirty-four three- and four-year olds ( $M = 45.1$  months; range: 36.2 – 58 months; 14 boys) participated. Four additional children were tested but excluded from analyses because they looked into the buckets before the hiding event was completed.

**1b:** Thirty three- and four-year olds ( $M = 44.7$  months; range: 36.7 – 56.6 months; 16 boys) participated. Three additional children were tested but excluded from analyses because they looked into the buckets before the hiding event was completed. All participants were recruited through a departmental participant pool. Informed consent was obtained for all children. Children received a small prize and their parents were paid \$5 for travel expenses.

## ***Materials***

In Experiment 1a, there were three types of familiar items: pink rubber pigs, wooden balls and yellow painted wooden blocks. In Experiment 1b, we added silver keys as a fourth familiar item type. All items were approximately one inch tall and one inch wide, except the keys, which were approximately two inches in length. The items were hidden into two identical opaque buckets. In Experiment 1a, at the end of the hiding event, bucket A had 10 pigs, 10 balls, and five blocks, and bucket B had five pigs, five balls, and 10 blocks. In Experiment 1b, bucket A had 10 pigs, 10 balls, five blocks and five keys, and bucket B had five pigs, five balls, 10 blocks and 10 keys (See Appendix 2).

## ***Procedure***

Children sat at a table across from the experimenter. Two opaque buckets were placed 80cm apart on the table. The experimenter told each child that they were going to play a hiding game and the child's job was to remember which of the experimenter's toys went into which bucket. The experimenter then began the warm up event. She placed three toy animals (a stuffed animal frog, a stuffed animal dog and a rubber ducky) on the table and had the child name each item. The experimenter then held up each item and placed it into one of the buckets in an alternating manner. After the third item was hidden, the experimenter placed one hand on top of each bucket and asked the child to point to the bucket where each toy was hidden, referring to each toy by the name the child spontaneously gave (e.g. "froggy," or "duck"). After the child answered all three questions, the experimenter emptied out each bucket and gave the child feedback about his or her answers, saying for example, "That's right! The frog was in this

bucket!” or “Oops! The frog was in this bucket!” If the child failed to correctly locate any of the three toys, the procedure was repeated up to two more times. Successful completion of the warm up task was required for inclusion in the data analysis.

For the test event, the experimenter told the child that she had some other toys to show him or her. She then placed one pink plastic pig, one wooden ball and one yellow painted block on the table on the table and said, “Here are three kinds of toys that I have.” The experimenter then held up each item and asked the child to name it. Children correctly named each item type, with small variations such as referring to the block as a square or a cube. The experimenter then said, “But I don’t just have these three, I have a whole bunch of them, and I’m going to hide them into these buckets. As I hide each item, I want you to tell me what it is.” The experimenter then practiced holding up each item and having the child name it before taking the three exemplars off the table. The experimenter then began the hiding event, taking the objects from a container hidden under the table with her right hand and holding each object above the bucket before placing it inside, making sure the child was correctly naming each object. One object was placed approximately every two seconds. The experimenter finished placing all of the objects in one bucket before moving on to the second bucket.

For each of the three collections of objects, one bucket contained twice as many items as the other bucket (e.g., one bucket has 10 pigs, the other has 5; one has 10 blocks, the other has 5) (See Appendix 2). Objects were presented in a randomized order such that items from each of the three collections were temporally intermixed.

After placing all of the objects into the second bucket, the experimenter placed one hand on top of each bucket and asked the child to indicate the bucket containing more items of a given

type, using the name that the child had used for labeling (e.g. “Which bucket has more piggies in it?”). If the child did not spontaneously choose a bucket by pointing, a rare occurrence, the experimenter asked, “Can you point to it?” Which item type was queried, which bucket contained more items of a given type and the starting side of the hiding event was counterbalanced across participants. Each child was asked about only one item type. Because the object type queried was counterbalanced across children, and children did not know which type they would be asked about, analyzing only responses for a single item provides an accurate reflection of their abilities.

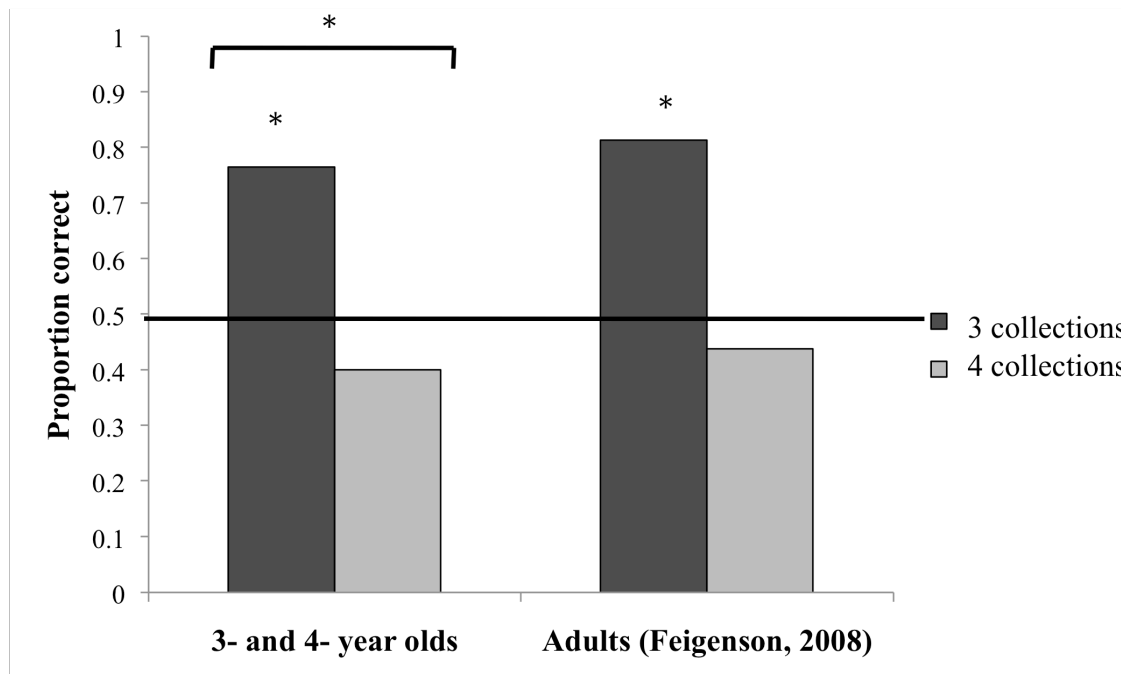
## ***Results***

**1a:** Three- and four-year old children were above chance in choosing which bucket had more items of any of the three types (Fig. 2). 26/34 children successfully indicated the bucket containing more items of a given type,  $p = .003$ , sign test. There was no difference in performance between the three-year olds (15/21 correct) and four-year olds (11/13 correct)  $p = .44$ , Fisher’s exact test, nor was there a difference based on item type queried, (pigs: 9/13 correct; balls: 9/10 correct; blocks: 8/11 correct)  $p = .49$ , Kruskal-Wallis test, all tests reported in this paper are 2-tailed.

**1b:** Like Feigenson’s (2008) adult participants, three- and four- year olds performed at chance when asked to judge the relative numerosities of four collections of items across two locations (Fig. 2). 12/30 children correctly indicated the bucket containing more items of a queried type,  $p = .36$ , sign test. There was no difference in performance between the three-year olds (8/23 correct) and four-year olds (2/7 correct)  $p = 1.0$ , Fisher’s exact test, nor was there a

difference based on item type queried, (pigs: 3/7 correct; balls: 3/8 correct; keys: 4/9 correct; blocks: 2/6 correct)  $p = .39$  Kruskal-Wallis test. The difference in performance with three collections in Experiments 1a and with four collections in 1b was significant  $p < .01$ , Fisher's exact test.

**Figure 2**



*Fig. 2. Proportion of correct responses across all questions asked in Experiments 1a and 1b of the current study, and proportion of correct responses across all questions asked in the 3 and 4 collection conditions of Experiment 1 in Feigenson's (2008) adult study. The black line indicates chance performance. Both young children and adults were above chance in choosing the correct bucket for 3 mutually exclusive collections, but performance for both age groups fell to chance when 4 mutually exclusive collections were presented.*

## **Discussion**

Experiment 1a provides the first evidence that by three years of age, children can encode summary representations of multiple large collections that accrue one by one over time, and further that they can flexibly update these representations over the course of a hiding event. Like adults, young children can represent the relative numerosities of three, but not four collections in parallel, suggesting that working memory capacity for summary representations of independent collections reaches adult-like levels during the preschool years. Children's performance on this task was no worse than adults' performance in Feigenson's (2008) original task using this same paradigm (Children: 76% correct, Adults (experiment 1): 81% correct)  $p=1.00$ , Fisher's exact test. These data suggest that the components of EF needed for this task, namely shifting attention among mental collections and holding and updating up to three collections in working memory, are in place by three years of age.

The units young children represent in working memory can be composed of individual items, chunks or summary representations of large collections of items. Like adults, children can encode information differently depending on the type of event and the structure of the input; but unlike adults, children may lack the ability to advantageously select an encoding schema via top-down control. In Experiment 1, the optimal organization of to-be-remembered information was clear and the basis for collection membership was obvious; however, many of the collections we encounter in our everyday lives are not so clearly delineated.

For example, imagine if rather than items of four different semantic types, the four item types used in Experiment 1b had overlapping features of similar salience, like red and blue circles and triangles. In this case, attempting to encode and update mutually exclusive collections of items defined by the conjunctions of color and shape features (e.g. red circles, red triangles, blue circles and blue triangles) would overwhelm participants' working memory capacities and

they would fail as did the adults in Feigenson (2008) and the children in Experiment 1b of this paper. However, if participants could instead encode the items as members of just two feature-based collections along one dimension, for example, red items and blue items in bucket A and red items and blue items in bucket B, they would be able to answer questions about the relative numerosities of these two feature-based collections. Additionally, once participants are tracking just two collections of items, in this case red and blue items, they may then detect and encode any regularities along the orthogonal dimension, in this case, the relative frequencies of circles and triangles within each color-based superordinate collection. From this kind of hierarchically organized representation of collections and knowledge of the statistical regularities among them, participants would be able to reconstruct information about the relative numerosities of all four feature-based collections (red things, blue things, circles and triangles). We recently explored adults' ability to efficiently reorganize collections like these in working memory, and found that adults can apply hierarchical structure to ambiguous collections of items to improve working memory performance (Tuerk, Moher, Feigenson & Carey, submitted).

In a task analogous to the one used in Feigenson (2008), adult participants saw four item types that overlapped in features like the red circles, red triangles, blue circles and blue triangles described above. The stimuli were structured such that, for example, the most common red item was a triangle (not a circle) and the most common blue item was a circle (not a triangle), thereby creating a situation where orthogonal features were statistically co-dependent. We found that adults were able to successfully judge the relative numerosities of all four feature-based collections (in this case red things, blue things, circles and triangles) across the two hiding locations, but only did so reliably when the overlapping features of the stimuli were explicitly pointed out by the experimenter before the hiding event began (e.g. "some things are blue, some



things are red, some things are circles and some things are triangles”). This suggests that adults can use previously presented information to reorganize mutually exclusive, conjunction-based collections into fewer superordinate feature-based collections, and use the statistical regularities among these features to increase working memory performance.

On this model, after the experimenter highlighted the features of the objects prior to the hiding event, participants would select the one featural dimension to serve as the bases for two feature-based collections that compose the top level of the hierarchy (e.g., could select color, thereby parsing the objects into red objects and blue objects). They then would represent these two feature-based collections in each hiding location (e.g., red objects and blue objects in location A, and red objects and blue objects in location B), and represent the approximate numerosity of each of the two collections in each of the two locations (e.g., about 16 red objects and about eight blue objects in location A; about eight red objects and about 16 blue objects in location B). We know from previous studies using this paradigm that adults can accurately represent the approximate numerosity of up to three collections in each of two locations in parallel (Feigenson, 2008), and therefore that doing so for two collections is well within their capacity. The hierarchy model then requires that participants also represent information about the relative values along the orthogonal dimension(s) that are lower on the hierarchy (e.g., shape), nested within representations of each of the superordinate collections (e.g., red objects, blue objects). Frequencies of the feature-based collections along the secondary (and tertiary) dimension(s) may be represented in terms of average exemplars that are continually updated throughout the hiding sequence (e.g. the average red object is closer in shape to a triangle than to a circle), or by ratio information (e.g., more of the red objects were triangles than circles). Either representation could then be combined with the estimated numerosity of the superordinate

collection to compute relative numerosity judgments of the secondary feature-based collections across locations.

## ***Experiment 2***

Both preschool children and adults can remember up to three large collections of items that collect over time by encoding and updating multiple summary representations in parallel, but adults can surpass even this capacity limit by representing hierarchically organized collections with overlapping, statistically dependent features (Tuerk et al., submitted). In Experiment 2, we seek to replicate this finding with adult participants using the same stimuli and procedures we will use with child participants in Experiment 3. Additionally, we test the hierarchical model hypothesis more directly by manipulating the relative salience of the orthogonal feature dimensions, leading participants to select features along a specific dimension as the bases for superordinate collections.

In a task analogous to that of our previous adult study (Tuerk et al., submitted), participants see items of four types: large balls, small balls, large blocks and small blocks, hidden sequentially across two locations. Instead of shadowing random letter strings during the hiding event, participants are instructed to label each item presented along the size dimension (size-labeling condition) or shape dimension (shape-labeling condition). Such labeling should interfere with counting, but at any rate, participants do not know they will be probed about number. Labeling each sequentially presented item along one dimension should lead participants to represent the two feature-based collections along that dimension as the top level of the hierarchy (e.g. labeling items along the shape dimension will lead to the representation of a collection of

balls and a collection of blocks). If participants are indeed organizing stimuli hierarchically, they should be better able to judge the relative numerosities of feature-based collections along the labeled dimension (e.g. “which bucket has more balls in it?”) than along the orthogonal, unlabeled dimension (e.g. “which bucket has more big things in it?”), but should also be able to reconstruct answers about the relative numerosities of subordinate collections.

## ***Methods***

### ***Participants.***

Twenty-eight adults ( $M = 20.2$  years; range: 18 – 25 years; 7 male) who were native English speakers with normal or corrected to normal vision were recruited from Harvard University. Fourteen participants (five male) were in the size-labeling condition, and 14 participants (two male) were in the shape-labeling condition. Participants gave informed consent and received course credit for participating.

### ***Materials.***

There were four types of items: small balls, small blocks, large balls and large blocks. The small items and the large block were made out of wood and the large ball was made out of pliable foam. All items were painted the same shade of yellow. The small items measured 1 inch wide x 1 inch tall x 1 inch deep. The large items measured approximately 3 inches wide x 3 inches tall x 3 inches deep. Sixteen out of 24 balls were small, and 16 out of 24 blocks were large. At the end of the hiding event, 16 out of 24 items in bucket A were balls, and 16 out of 24

items in bucket A were small. Sixteen out of 24 items in bucket B were blocks, and 16 out of 24 items in bucket B were large (See Appendix 1).

### ***Procedure***

The “warm up” trials were identical to those in Experiment 1.

For the test event, the experimenter told the participant that she had some other toys to show him or her. She then placed one small ball, one small block, one large ball and one large block on the table, one at a time, having the participant label each item as it was presented. Once all four items were on the table, the experimenter highlighted the overlapping features of the items. The experimenter held up the big and small balls and said, “These two are the same because they’re both balls.” She then held up the big and small blocks and said, “These two are the same because they’re both blocks.” She then held up the small ball and small block and said, “These two are the same because they are both small.” She then held up the large ball and large block and said, “These two are the same because they’re both big.” The experimenter then arranged the four items on the table and said, “Some of my toys are balls, and some of them are blocks, some of them are big, and some of them are small,” pointing to the relevant items as she spoke. She then said, “But I don’t just have these four, I have a whole bunch of them, and I’m going to hide them into my buckets.”

If the participant was in the size labeling condition, the experimenter then said, “As I hold up each item, I want you to tell me whether it is big or small.” If the participant was in the shape labeling condition, the experimenter instead said, “As I hold up each item, I want you to tell me whether it is a ball or a block.” The experimenter then said, “Let’s practice!” and held up each one of the four items for the participant to label according to his or her labeling condition

before taking the four exemplars off the table. The experimenter then began the hiding event, taking the objects from a container hidden under the table with her right hand and holding each object above the bucket before placing it inside, making sure the participant was correctly labeling each object. One object was placed approximately every two seconds. The experimenter finished placing all of the objects in one bucket before moving on to the second bucket.

For each of the four feature-based collections (balls, blocks, big items and small items), one bucket contained twice as many items as the other bucket (e.g., one bucket has 16 balls, the other has 8; one bucket has 16 big items, the other has 8) (See Appendix 1). Objects were presented in a pseudo-randomized order such that items from each of the four conjunction-based collections were temporally intermixed.

After placing all of the objects into the second bucket, the experimenter placed one hand on top of each bucket and asked the participant to indicate the bucket containing more items of a given type (e.g. “Which bucket has more balls in it?” or “Which bucket has more small things in it?”). Each participant was asked two questions, one along the shape dimension, and one along the size dimension. Which feature-based collection was queried first, which bucket contained more items of a given type and the starting side of the hiding event was counterbalanced across participants.

The feature-based collection queried was either congruent or incongruent with regards to the dimension along which participants had been labeling each item. For example, if participants had been labeling items along the shape dimension, labeling each item as “big” or “small” as it was presented, they could then be asked a congruent question, such as, “Which bucket has more big things in it?” or an incongruent question such as, “Which bucket has more balls in it?”

## ***Results***

Participants were significantly above chance in choosing which bucket had more balls, blocks, small items or large items, 89%,  $t(27)=7.32$ ,  $p < .0001$ . Because questions were counterbalanced, half of the participants first answered a question about objects of a specific shape (either balls or blocks), and half of the participants first answered a question about items of a specific size (either small or large). There was no difference in performance based on question order, (Q1: 26/28 correct, Q2: 24/28 correct,  $p=.688$ ) or dimension queried, (shape: 26/28 correct, size: 24/28 correct,  $p=.688$ , both McNemar tests) There was no difference in performance based on labeling condition,  $p = .54$ , independent samples Mann-Whitney test. There was a significant effect of congruency (congruent: 28/28 correct; incongruent: 22/28 correct)  $p = .031$ , McNemar test), but participants were significantly above chance in answering questions both along the congruent dimension,  $p < .0001$ , and the incongruent dimension,  $p = .004$ , both sign tests.

## ***Discussion***

Experiment 2 offers convergent evidence that adults can exceed working memory capacity limits and represent a total of four hierarchically organized, feature-based collections in situations where items collect over time and regularities are introduced among features. Even though participants were uninformed about the nature of the task, and had no idea when the streams of objects being placed into the two locations would end, they were able to answer with

high accuracy which location had more balls, blocks, large items, and small items. Their performance in specifying which bucket had items in a given feature-based collection (89.3%) was indistinguishable from that of participants tracking objects of a single type (e.g., toy pigs) in a similar task (1-type condition, 81% correct, Feigenson, 2008).

Additionally, the labeling manipulation introduced in Experiment 2 allowed us to directly test the hierarchical model proposed in the previous study of this ability (Tuerk et al., submitted). On this model, participants encode numerosity information for collections along a superordinate dimension and combine it with information about the relative frequencies of features along the orthogonal dimension to represent all four feature-based collections. We predicted that the labeling manipulation would lead participants to select superordinate collections along the labeled dimension, a hypothesis borne out by participants' significantly higher performance on questions about these collections as compared to collections along the unlabeled dimension.

As participants did not know which dimension(s) they would be asked about, their ability to accurately answer questions about any two of the four feature-based collections suggests that they were indeed representing all four collections, or were able to reconstruct information about any of the four from working memory. Given that this success appears to exceed previously identified capacity limits in this paradigm, participants must have represented the collections in Experiment 2 differently than when representing mutually exclusive collections (as in Feigenson, 2008, Exp.1).

Participants in Experiment 2 could not have been representing the four exemplar objects shown prior to the hiding event as large block, small block, large ball and small ball, because four mutually exclusive collections so defined would have exceeded working memory capacity

and resulted in failure. Instead, participants likely represented two collections along the labeled dimension that could further be parsed along the orthogonal dimension. For example, a participant in the shape labeling condition may have represented the items as a collection of blocks (which could be further parsed along the dimension of size) and a collection of balls (which also could be further parsed along the dimension of size). It is possible that the prior experience of seeing all four exemplars together before the hiding event, despite its brief nature, may have led participants to create a hierarchically organized representation of the items' features before any item was hidden. With the appropriate hierarchical structure in place, participants were then able to update summary representations of collections at both the superordinate and subordinate levels, encoding both the numerosities of the collection along the labeled dimension and the relative frequencies of the features along the orthogonal dimension. While performance was better on congruent questions, participants were also significantly above chance in judging the relative numerosities of collections along the incongruent, unlabeled dimension. The additional noise in judgments about the relative numerosities of collections along the unlabeled dimension indicates that participants are reconstructing these representations by combining numerosity information at the superordinate level of the hierarchy with information about the relative frequencies of features along the orthogonal dimension.

### ***Experiment 3***

Adults are able to represent collections using more flexible selection criteria to empower the detection of statistical regularities among collections and the encoding of more information than would be otherwise possible. The development of this ability has yet to be explored, and



will be the focus of Experiment 3 of this paper. While young children demonstrate the same working memory capacity for mutually exclusive collections as adults (Experiment 1), there are a few reasons to believe they may fail in the task used in Experiment 2.

Although working memory capacity limits and at least some computations that allow us to surpass those limits are continuous over ontogeny, it is possible that the type of reorganization necessary to represent overlapping feature-based collections of items that accumulate over time requires additional, domain-general capacities that develop during childhood. Identifying the common features among items that serve as the bases for collection membership, representing multiple collections in parallel and switching among them to update numerosity information as new items are presented requires a suite of EF skills including inhibition, control over interference, working memory and cognitive flexibility.

Our previous work suggests that when overlapping dimensions are not explicitly highlighted, adults' inclination is to track stimuli like those used in Experiments 2 and 3 as four conjunction-based collections, thereby overloading their working memory capacity (Tuerk, et al., submitted). To succeed in Experiment 2, adults had to inhibit this tendency and instead track items as members of two superordinate feature-based collections, heterogeneous along the orthogonal dimension. Previous studies have demonstrated that significant development of inhibitory processes occurs in the preschool years leading to age related gains in performance on complex inhibition tasks that tax working memory (e.g. Gerstadt, Hong, & Diamond, 1994; Carlson, 2005). It is possible that young preschoolers will be unable to inhibit the tendency to select conjunction-based collections, and thus may fail in Experiment 3 as they do in Experiment 1b with four mutually exclusive collections.

In order to detect and use the statistical regularities among features in Experiment 2, adults had to first reorganize the four conjunction-based collections of small balls, small blocks, large balls and large blocks, into feature-based collections along the two dimensions of shape and size. To accomplish this, adults used their knowledge of the relations among stimuli attributes (namely the overlapping feature dimensions) to select wider criteria for collection membership and represent feature-based rather than conjunction-based collections.

The stimuli used in Experiment 3 are the self-same stimuli from Experiment 1, and like stimuli used in traditional task switching tests of cognitive flexibility, items in our four conjunction-based collections are bivalent, that is, they carry a feature relevant to each of the two dimensions along which these items can be sorted. In Zelazo's classic task switching test, the Dimensional Change Card Sort Test (DCCS), three-year old children successfully sorted bivalent cards by either color or shape, but failed to switch the bases for their sorting when the other dimension became relevant (Zelazo et al. 1996, 2003). Evidence suggests a link between cognitive flexibility like that employed in the DCCS and the ability to represent abstract sorting rules, as only those children who flexibly switched in the DCCS were able to apply their sorting behavior to novel cards (Kharitonova & Munakata, 2011). For example, children who successfully switched from sorting blue trucks and red flowers by shape were able to apply a more general "shape" sorting rule to novel cards (e.g. sorting a green TV with a similarly boxy purple truck, and a yellow tennis ball with a similarly round red apple). This link between cognitive flexibility and abstraction also held across DCCS performance and performance in an "odd-one-out" task where children were asked to select the one picture that did not belong with the others. The finding that only switchers succeeded in the odd-one-out task where the target depends not on the attributes of the stimulus itself, but rather on the relations among stimuli

attributes, suggests that the link between cognitive flexibility and abstraction is quite general (Kharitonova & Munakata, 2011).

Labeling items along either the shape or size dimension may help children encode the four item types presented as members of two feature-based collections, but it need not be so. While a few studies have shown that labeling facilitates children's performance on EF tasks (Kirkham, Cruess & Diamond, 2003), the exact role of labeling in promoting performance remains uncertain. One possibility is that applying shared labels allows children to generate more abstract relations among items (Gentner & Loewenstein, 2002). If this is the case, then labeling should allow children to successfully reorganize four conjunction-based collections into two feature-based collections. Alternatively, labeling may just strengthen a particular bottom-up response by directing attention to it without fostering any change at a deeper conceptual level. (Yerys & Munakata, 2006). If children can reorganize the four conjunction-based collections into two feature-based collections, they should succeed robustly when asked about a feature-based collection congruent with the labeling condition, for two collections is fewer than their three collection capacity limit on working memory representations under these conditions. If applying shared labels to the four conjunction-based collections makes these superordinate collection representations available, then even three-year old children should successfully judge the relative numerosities of feature-based collections in the congruent condition.

It is possible that once the four conjunction-based collections are reorganized into two feature-based collections, detecting and encoding the statistical regularities among the orthogonal features comes for free. If this is the case, then children able to answer questions about the congruent feature-based collections should also be able to answer questions about the incongruent feature-based collections, like the adults in Experiment 1. However, evidence

suggests that the ability to flexibly represent multiple features of objects, a necessity for encoding the statistical regularities in Experiments 2 and 3, develops significantly over the preschool years (e.g. Kirkham, et al., 2003; Zelazo, Frye, & Rapus, 1996). If representing a hierarchy of features beyond those that are labeled requires additional EF capacities, it is possible that participants at a certain age will be able to successfully answer questions about congruent feature-based collections, but fail on questions about incongruent feature-based collections.

## ***Methods***

### ***Participants***

Thirty-four three-year olds (M = 42.1 months; range: 36.5 – 47.1 months; 17 boys), 32 four-year olds (M = 55.1 months; range: 49.6 – 59.7 months; 14 boys), 31 five-year olds (M = 65.5 months; range: 60.0 – 70.2 months; 17 boys), 30 six-year olds (M = 78.5 months; range: 72.6 – 83.5 months; 13 boys) and 31 seven-year olds (M = 88.8 months; range: 84.0 – 95.1 months; 13 boys) participated in Experiment 3. One-hundred seventeen children were recruited and run in a private room at the Boston Children's Museum, and 35 children were recruited from the departmental database and run in the lab. Additional children were excluded from analyses because they did not pass the warm up trials (2) or they looked into the buckets before the hiding event was completed (3). Informed consent was obtained for all children. All children received a small prize and parents whose children participated in the lab were paid \$5 for travel expenses.

### ***Materials***

The materials were the same as those in Experiment 2.

### ***Procedure***

The procedure was the same as in Experiment 2, except we analyze only children's responses to the first question asked. As discussed in Feigenson (2008), analyzing multiple responses per participant may introduce unwanted sources of variance such as memory decay between the first and last query and interference from earlier responses on later responses. We were particularly concerned with the latter source of unwanted variance as we found young child participants tended to select each bucket once regardless of the correct answer. For example, 52 out of 66 three- and four- year old participants switched buckets in their responses to the first and second questions regardless of whether the correct bucket was the same or different for both queries (same: 25/33; different: 27/33) and regardless of whether they answered the first question correctly or incorrectly (first response correct: 31/38; first response incorrect: 21/28). Because participants did not know what collection(s) they would be asked about, and both the feature-based collection queried, and the order of dimensions queried was counterbalanced across subjects, analyzing participants' first responses provides an accurate measure of their ability.

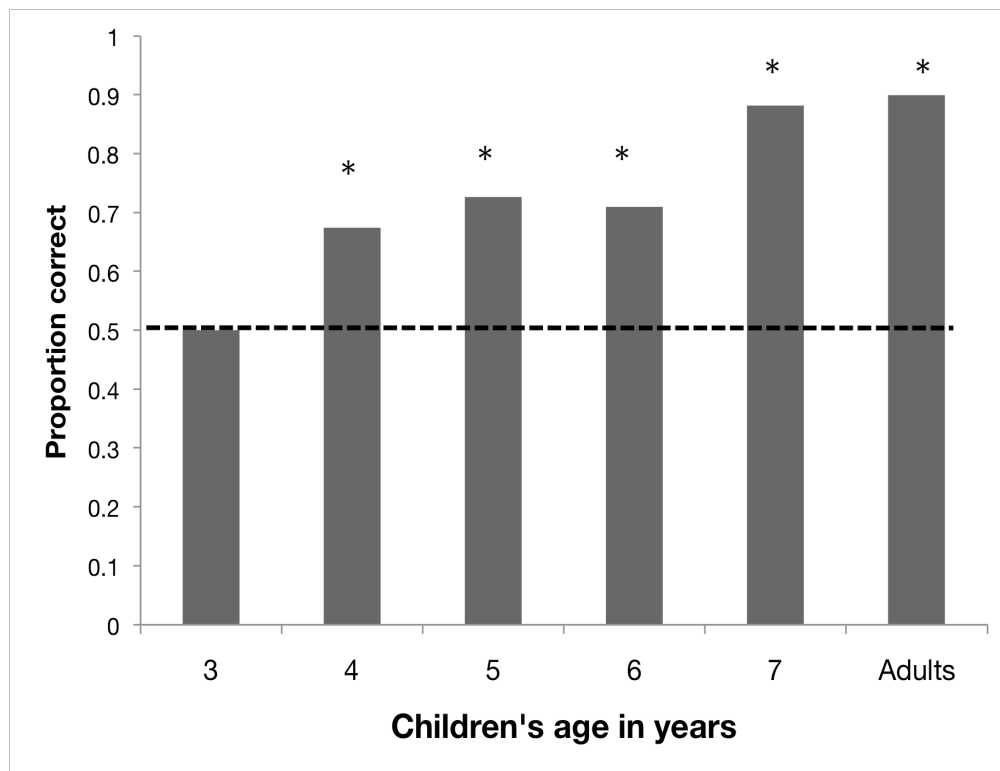
### ***Results***

Three-year olds were at chance in selecting the bucket with more balls, blocks, big items or small items: 17/34 correct,  $p = 1.14$ . Four-, five-, six- and seven- year olds were all significantly above chance in choosing which bucket had more items of a given type: four-year olds: 22/32 correct,  $p = .05$ ; five-year olds: 23/31 correct,  $p = .01$ ; six-year olds: 21/30 correct,  $p$

= .04; and seven-year olds: 27/31 correct,  $p < .001$ , all sign tests (fig. 2). Overall there was no difference based on the dimension queried (shape queried: 56/78 correct; size queried: 54/80 correct),  $p = .61$ , or the dimension labeled (shape labeled: 56/77 correct; size labeled: 54/81),  $p = .49$ , all Fisher's exact tests.

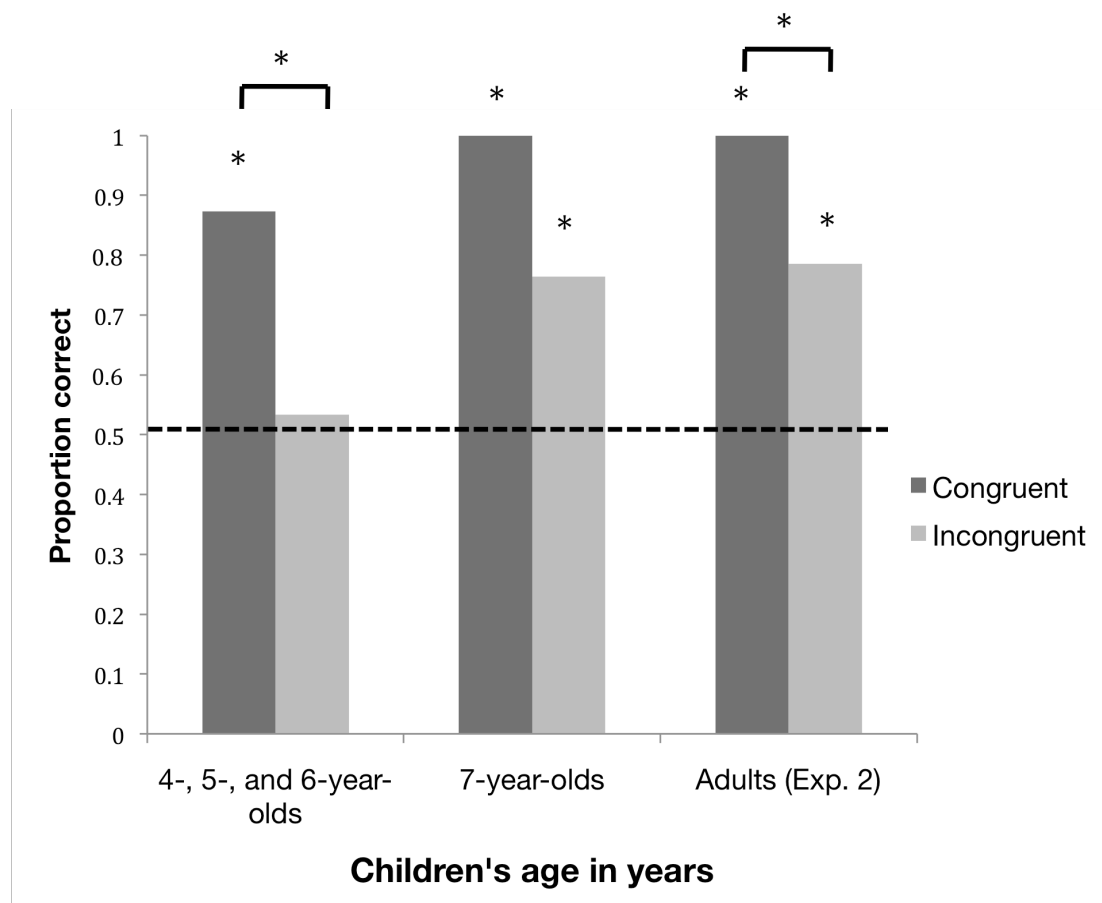
In sum, these results display three patterns of performance. Three-year olds performed at chance across congruent and incongruent questions (50% correct). Four- five- and six-year olds' overall success was driven by above chance performance on congruent questions, and they performed at chance on incongruent questions (four-year olds: congruent: 15/17 correct,  $p < .01$  incongruent: 7/15 correct,  $p = 1.0$ ; five-year olds: congruent: 15/17 correct,  $p < .01$ , incongruent: 8/14 correct,  $p = .79$ ; six-year olds: congruent: 12/14 correct,  $p < .02$ , incongruent: 9/16 correct,  $p = .80$ , all sign tests). Aggregating over these three ages, children performed significantly better on congruent questions (87.5%) than on incongruent questions (53%), ( $p < .001$ , Fisher's exact test) (fig. 3). Seven-year olds' overall success, like adults', resulted from above change performance on both congruent and incongruent questions (congruent: 14/14 correct,  $p < .0001$ , incongruent: 13/17,  $p < .05$ ), and although their performance was higher on congruent questions than incongruent questions, it was not significantly so ( $p = .11$ , Fisher's exact test).

**Figure 3**



*Fig. 3. Proportion of correct responses on all questions about the relative numerosities of four feature-based collections queried in Experiments 2 (Adults) and 3 (3- to 7-year olds). Chance performance, indicated by the dashed line, is .5.*

**Figure 4**



*Fig. 4. Proportion of correct responses on feature-based collections queried in Experiments 1 (Adults) and 3 (4-7- year olds) by congruence. Congruent queries were those of feature-based collections along the same dimension participants labeled, and incongruent queries were those of feature-based collections along the unlabeled, orthogonal dimension. Chance performance, indicated by the dashed line, is .5.*

## **Discussion**

Experiment 3 demonstrates that although working memory capacity for mutually exclusive collections asymptotes at adult-like levels by three years of age (Experiment 2), the ability to efficiently encode overlapping collections to improve working memory performance



develops between the ages of three and seven. Three-year olds are unable to reorganize four mutually exclusive conjunction-based collections into two feature-based collections even when the overlapping features of the stimuli are explicitly highlighted before the hiding even and their attention is continually drawn to the features that define these collections (congruent condition). It appears that labeling each presented item along one feature dimension was not enough to lead three-year olds to represent the more abstract feature-based collections along that dimension. The fact that three-year olds performed like adults in Experiment 1 but failed even in the congruent condition of Experiment 3 suggests that representing overlapping collections in this paradigm makes additional demands on EFs that have yet to develop by this age. Given findings of the relationship between cognitive flexibility and abstraction (Kharitonova & Munakata's 2011), we can interpret the failure of 3-year-olds, who fail to switch sorting rules in DCCS tasks, as a failure to form feature-based collections based on the similar attributes among perceptually dissimilar items (e.g. large block and large ball).

Four-, five- and six- year olds successfully represent the relative numerosities of feature-based collections along the labeled dimension but are at chance on questions about the dimension incongruent with the labeled dimension. This suggests that unlike three-year olds, four- to six-year olds are able to inhibit the more salient parsing of the stimuli as four conjunction-based collections and instead represent two feature-based collections along the labeled dimension. Their failure to judge the relative numerosities of feature-based collections along the unlabeled dimension suggests that they are not detecting the statistical regularities among the orthogonal feature dimensions.

Seven-year olds perform like adults, and are above chance on both congruent and incongruent questions, suggesting that they are able to inhibit the tendency to represent

conjunction-based collections, and are able to flexibly attend to both dimensions of each item presented and update the relevant feature-based collections. The pattern of results across age groups in Experiment 3 suggests that there is an additional component involved in the seven-year olds' and adults' representations of the hiding events. While four to six year olds children are able to reorganize four conjunction-based collections into two feature-based collections, only at seven years of age are they able to increase working memory capacity and represent four overlapping collections via hierarchical reorganization.

### ***General discussion***

Our ability to represent multiple collections in parallel is not limited to ensemble representations in visual working memory, but extends to dynamic events that unfold over time (Feigenson, 2008; Tuerk et al., submitted). The current study demonstrates that young preschoolers' ability to represent and flexibly update multiple, mutually exclusive collections in working memory mirrors that of adults (Experiment 1). Both three- to four-year olds and adults successfully represent the relative numerosities of three, but not four, mutually exclusive collections that accrue over time across two locations. Our work (Experiment 2), along with previous work from our lab (Tuerk, et al., submitted) demonstrates that adults are able to surpass this limit by reorganizing collections with overlapping features hierarchically and detecting statistical regularities among the features that define collections at the superordinate and subordinate levels. We find that while even three-year-old children represent three mutually exclusive collections in parallel (Experiment 1), the ability to surpass this capacity limit develops in stages over the next four years of childhood (Experiment 3).

The first stage is evidenced by three-year olds' performance; in Experiment 1a, three- and four- year olds were able to simultaneously represent and dynamically update the relative numerosities of three mutually exclusive collections that accumulated over time. Children watched as an experimenter hid streams of items of three types (balls, blocks and toy pigs) in an intermixed fashion into two buckets, hiding all items in the first bucket before moving on to hide items in the second bucket. At the end of the hiding event, one bucket contained twice as many items of a given type as the other (e.g. one bucket had 10 pigs, the other had 5). Although children had no idea when they had seen the last of any item type, and did not know what questions they would be asked, their performance on this task was no worse than adults' performance on a comparable task (Feigenson, 2008). Like adults, children's performance fell to chance when a fourth item type was added to the hidden streams (children: Experiment 1b; adults: Feigenson, 2008).

While even young infants can select and represent multiple collections in a visual working memory task, that young preschoolers are able to represent collections that accrue one by one over time indicates a novel continuity in updating abilities over the lifespan. Three- and four- year olds' ability to update multiple collection representations in an unpredictable, alternating manner provides a stark contrast to the updating failures of infants in even the most basic tasks. For example, when crackers are hidden in direct succession such that all crackers in one location are hidden before any crackers are hidden in the second location, 11-month olds reliably crawl to the bucket with more crackers in comparisons of one versus two and one versus three. However, when crackers are hidden in alternation, 11-month olds select a bucket at random in the same comparison conditions (Feigenson & Yamaguchi, 2009).

Although evidence from the aforementioned cracker task suggests that toddlers are unable to flexibly update multiple collections in parallel, it is possible that the manipulations involved in the cracker study are different from those involved in the updating of large collections like those used in the current study. Previous work suggests that infants are representing each cracker as a separate object in the cracker task (Feigenson et al., 2002), and as such, are limited by the number of items they can efficiently represent as a chunk. Adults, contrastingly, need only to increment a single summary representation for each collection in the counter task. In the cracker task, the mnemonic units being employed are chunks, whose representational capacity is limited by the number of individuals they comprise. In tasks that require the updating of counters as collections accumulate, the summary representations serve as the units in working memory. Results of the current experiments suggest that the ability to dynamically reupdate multiple representations of collections in working memory is available in childhood and largely continuous over the lifespan, but that these updates may only be carried out over a limited number of preexisting representations.

While three-year olds successfully select and update three mutually exclusive collections in Experiment 1, they fail to reorganize four item types (small balls, small blocks, large balls and large blocks) into feature-based collections along either size or shape dimensions in Experiment 3. In Experiment 3, children's attention is continually drawn to a single dimension as they label each item. For example, in the shape-labeling condition, children label each item as a block or a ball regardless of its size. Although three-year olds successfully follow labeling directions and spend approximately five minutes labeling the shape of each item presented into each bucket, they are completely at chance in their judgments of which bucket contains more blocks or balls. Our findings in Experiment 1 suggest that encoding the relative numerosities for just two

collections in this paradigm should be trivially easy for three-year old children, and yet, they fail under these conditions. We argue that the bivalent nature of the items used in Experiment 3 interferes with three-year olds' ability to reorganize the collections into a more manageable format of two feature-based collections.

The ability to flexibly represent multiple features of objects develops significantly over the preschool years, and young preschoolers often have trouble switching between feature dimensions on the DCCS, perseverating on the first sorting rule (e.g. Zelazo et al., 1996; Kirkham et al., 2003). In Experiment 3 we use only one labeling rule throughout the entire task and continually draw children's attention to one dimension, significantly decreasing demands on inhibitory control, and yet three-year olds fail to represent the congruently labeled feature-based collections in Experiment 3 of our study. It remains possible that our labeling manipulation was not enough to prevent three-year olds from trying to represent four conjunction-based collections, offering additional evidence that the effects of labeling in EF tasks are not happening at the conceptual level.

Sorting by a single rule in the pre-switch trials of the DCCS is akin to labeling sequentially presented items in our task, but both sorting and labeling may be done without representing any abstract relations among items and without representing any categories or collections. It has been proposed that children who persevere, or fail to switch rules on post-switch trials of the DCCS, have the same underlying representation of the rules as children who switch, but lack the ability to inhibit behavior associated with the previous sorting rule (e.g. Zelazo et al., 1996; Kirkham et al., 2003). An alternative account posits that children who successfully switch and children who persevere have different underlying representations of the competing rules. While switchers rely on "active" memory representations of the rules to

support top-down control of attention to relevant features, perseverators rely on “latent” memory representations, built up through repetitive experience. (Munakata, 1998; Morton & Munakata, 2002; Kharitonova, Chien, Colunga & Munakata, 2009). Active memory representations, which rely on later developing regions of PFC, are thought to code for more abstract information, whereas latent representations are thought to code for more stimulus specific representations (Ashby & Maddox, 2005; Bunge & Zelazo, 2006; Rougier, Noelle, Braver, Cohen, & O'Reilly, 2005; Wallis, Anderson, & Miller, 2001). While all items within a collection were identical in Experiment 1 and children could select collections based solely on stimuli similarity, the feature-based collections in Experiment 3 were based on features shared across items, a more abstract, relational concept that three-year olds were unable to use as the basis for selection.

The second stage in the development of the ability to represent hierarchically organized collections is evidenced by four- to six-year olds’ performance; in Experiment 3, four- five- and six- year olds successfully judged the relative numerosities of feature-based collections along the labeled dimension, but performed at chance on questions about feature-based collections along the orthogonal, unlabeled dimension. In order to successfully answer questions about all four feature-based collections in our study, participants must be able to represent multiple orthogonal dimensions in parallel. While they must be attending to one dimension in order to label correctly and form superordinate collections, they must also be able to represent regularities along the orthogonal dimension. Four- to six-year olds are able to reorganize items into two abstract feature-based collections, but appear to only represent information along the labeled dimension, ignoring the orthogonal dimension entirely. Older children and adults, however, appear to represent both feature dimensions. It is possible that seeing both feature-dimensions highlighted before the hiding event leads seven-year olds and adults to set up representations of both

dimensions in PFC, actively maintaining both while using the labeled dimension to select superordinate collections. Modeling work exploring the role of the PFC in learning abstract rules over time has shown that the PFC contextualizes the mapping between sensory inputs and outputs by representing relevant goals and prior information. Additionally, the PFC is known to support the active maintenance of multiple nested rules at once during tasks like the stroop task and the Wisconsin Card Sort Task (Rougier, et al., 2005).

The third stage of development in the ability to increase working memory capacity by representing hierarchical, structured collections is evidenced by seven- year olds' performance in Experiment 3 which mirrors adults' performance in Experiment 2. Both seven-year olds and adults correctly judge the relative numerosities of all four feature-based collections, suggesting that their representations are qualitatively different from representations of mutually exclusive collections like those used in Experiment 1, and in Feigenson's (2008) work.

We have previously argued that adults bypass capacity limits on collections by hierarchically reorganizing items with overlapping features and encoding statistical regularities among features to represent four collections at once (Tuerk et al., submitted), and the developmental data presented in the current paper are in line with the predictions of this argument. In order to use the proposed hierarchical model to increase working memory performance, a participant must first be able to select feature-based collections along a single dimension. At three years of age, children are able to represent and update three mutually exclusive collections in parallel (Experiment 1), but fail to reorganize four conjunction-based collections into two superordinate feature-based collections even along the labeled dimension (Experiment 3). Three-year olds' inability to select feature-based collections leaves them unable to get the hierarchical model off the ground. Between the ages of three and four, children make

gains in inhibitory control (Zelazo, 1996; Kirkham et al., 2003) and the ability to form categories based on abstract features (Kharitonova & Munakata, 2011). As such, four- to six- year olds are able to select feature-based collections along the dimension they are labeling and successfully answer questions about feature-based collections along the congruent dimension (Experiment 3). Their failure on questions about feature-based collections along the unlabeled, incongruent dimension suggests that they are only attending to a single feature dimension and not using the overall structure of the stimuli as highlighted by the experimenter before the hiding event to guide their representations.

Seven year olds' success on questions along both the congruent and incongruent feature dimensions suggests that, like adults, they are using the information given by the experimenter before the hiding event (highlighting the overlapping feature-based collection) to organize efficient mnemonic structures while the younger children are not. The ability to maintain this kind of goal-relevant information across distractions and delays such that it can be used to support and guide flexible behavior develops significantly between the ages of three and eight. While many developmental theories have posited that children are simply less skilled in their ability to *proactively* maintain goal-relevant information to guide behavior (e.g. Diamond, 1991), more recent work suggests that younger children use a qualitatively different, *reactive* form of cognitive control wherein goal-relevant information is only recruited on an as-needed basis (Chatham, Frank, & Munakata, 2009; Munakata, Snyder, & Chatham, 2012). Seven year olds may be succeeding where younger children fail because they are able to apply previously obtained knowledge to prepare for the hiding event by representing a hierarchy of features that will scaffold their representations and allow them to detect and encode regularities.



In addition to charting its development, the current study expands our understanding of how we represent overlapping collections and increase working memory by testing the proposed hierarchical model directly. We predicted that labeling items along one dimension (e.g. labeling items as “block” and “ball” regardless of size) would lead participants to select superordinate collections along this dimension and represent the numerosities of these two feature-based collections directly. Once participants represented these two feature-based collections, they would be in a position to detect and encode the regularities along the orthogonal, unlabeled dimension within each superordinate collection. From this information, participants would be able to answer questions about the relative numerosities of any of the four feature-based collections – either by reading out the numerosity information directly from the superordinate collections, or by reconstructing this information by combining the numerosity information from the superordinate collections with the statistical regularities encoded along the orthogonal dimension. We predicted that participants would be more accurate on questions about the labeled, superordinate feature-based collections for which numerosity was directly encoded than the unlabeled, orthogonal feature-based collections, which had to be reconstructed. This prediction was borne out by the pattern of adults’ performance in Experiment 2. While they were above chance at selecting the bucket with more items of any of the four feature-based collections, adults performed significantly better on questions about collections along the labeled dimension.

In summary, the present experiments demonstrate that the ability to represent multiple collections in parallel extends to dynamic events that unfold over time and is available by three-years of age. We find that while young preschoolers’ ability to represent and flexibly update multiple, mutually exclusive collections in working memory mirrors that of adults, the ability to

surpass this limit by reorganizing collections hierarchically and detecting statistical regularities among features develops in stages over the next four years of childhood. Further work remains to be done in order to determine the particular encoding schema adults and children use to succeed in the presented tasks. Future research exploring the interaction of EF and the selection of an optimized encoding scheme will shed light on the development of flexibility in working memory and the processes by which we surpass capacity limitations.

## PAPER 3

7-month-olds extract element orientation but not element size from homogeneous arrays

### *Abstract*

A debate in the adult literature has long waged over whether the processes underlying two well studied ensemble representations, average size and average orientation, rely on similar mechanisms of distributed attention. In the current study, we provide evidence that these representations rely on qualitatively different processes with distinct developmental trajectories. In Experiment 1, we demonstrate an asymmetry in infants' discrimination thresholds for size and orientation in single-element and multi-element arrays. Infants notice a twofold change in element size when presented with single element- but not homogeneous arrays and contrastingly, infants detect a 9° change in element orientation when presented with homogeneous arrays but not single-element arrays. In Experiment 2 we explore the basis of this discrepancy, testing the hypothesis that infants' attention is automatically spread across multiple elements within an array, and that this mode of processing supports ensemble representations of element orientation, but not element size. Taken together, our results suggest that the process underlying the extraction of ensemble element size relies on focused attention sampling strategies that are not available in infancy while the process underlying representations of ensemble orientation is pre-attentive, automatic, deployed in parallel across all elements in an array, and in place at 7 months.

## ***Introduction***

Infants' visual worlds are full of collections of similar elements – a box full of toys, a bowl of cheerios, a pile of blocks; but previous research has demonstrated that the number of individuated elements they can represent at once is strictly limited (Kaldy & Leslie, 2003; 2005; Moher, Tuerk, & Feigenson, 2012; Ross-Sheehy, Oakes, & Luck, 2003), precluding their ability to encode each toy, cheerio, and block directly. What then, do infants remember about these sets of elements when direct perceptual access is no longer available? Research has shown that infants extract numerosity information defined over entire large ( $> 4$  items) arrays (e.g. Brannon, 2002; Brannon, Abbot, & Lutz, 2004; Cordes & Brannon, 2008; Libertus & Brannon, 2010; Xu & Spelke, 2000). This ability appears to belie the strict capacity limits of focused attention, and suggests that infants, like adults, have compensatory processes that allow them to accurately perceive the statistical properties of a set of objects, forming what is known as an ensemble representation (Alvarez, 2011).

Researchers studying adults' perceptual systems have claimed that these ensemble representations, or statistical summaries, are effectively computed via automatic processes not limited by the bottleneck of focused attention and working memory (Alvarez, 2011; Chong & Treisman, 2003; Chong, Joo, Emmanouil, & Treisman, 2008). Unlike serial processing of individual objects in a scene, which allows us to encode no more than four objects (e.g. Luck & Vogel, 1997), the extraction of summary statistics, such as the mean or distribution of features among a set of similar objects, appears to employ a parallel mechanism devoid of capacity limitations. It has been claimed that this process relies on distributed attention across an entire array, and is a general mechanism that computes ensemble representations over multiple stimulus attributes including orientation, size, central tendency, and even facial expression and

identity (Albrecht & Scholl, 2010; Alvarez & Oliva, 2008; Ariely, 2001; Chong & Treisman, 2003; Dakin, 2001; Dakin & Watt, 1997; de Fockert, & Wolfenstein, 2009; Haberman & Whitney, 2007; 2009; Parkes, Lund, Angelucci, & Solomon, 2001; Robitaille & Harris, 2011). For adult observers, these ensemble characteristics have been shown to enhance memory for individuals through a process that replaces direct retrieval with inference based on summary statistics. For arrays of elements too numerous to be attended to and encoded via focused attention, adults are able to circumvent capacity limits and use ensemble representations in working memory to reconstruct information about individual elements. For example, when presented with a large array of dots, an adult will use the average size of all of the dots to inform his judgment of the size of a single dot presented therein (Brady & Alvarez, 2011).

Recent evidence suggests, however, that infants may not benefit from ensemble characteristics in a similar way. While adults seem to automatically extract the ensemble characteristic of element size from an array of dots, infants' representations of element size appear to be substantially hampered when elements are presented in an array as opposed to in isolation. For example, 6-month-olds habituated to a single element (an Elmo face) detected a twofold change in size during test (Brannon, et al., 2006); however, when habituated to arrays of homogeneously sized dots, infants did not dishabituate to novel test arrays in which all dots underwent a threefold change in size (Cordes & Brannon, 2011). Though infants robustly detected a twofold change in the size of a single element (Brannon, et al., 2006), they required a fourfold change in element size to detect a change in an array of homogeneously sized dots (Cordes & Brannon, 2011).

That infants are less sensitive to uniform size changes in arrays of homogeneous dots than to size changes in a single dot appears to be in direct contrast with findings that adults'

threshold for discriminating size change is the same for individual elements, and mean size judgments of homogeneous and heterogeneous arrays. In fact, at delays of 2 seconds, adults' representations of the average element size of a homogeneous array are more precise than their representations of the size of a single item or the average element size of a heterogeneous array (Chong & Treisman, 2003). Conversely, infants' element size representations are less precise under conditions where averaging models would predict higher accuracy (e.g. zero variation in element size), demonstrating that they are not benefiting from an automatic average size processing mechanism.

One possibility is that infants' fail to detect a twofold change in homogeneous ensemble element size because infants at 6-months are less accurate in representing any element feature when that element is part of an ensemble than when it is presented alone. Alternatively, it may be that the computations specific to adults' accurate average size judgments rely on representations not available in infancy.

### ***Ensemble processing in adults***

When the adult visual system encounters an ensemble of elements, it automatically computes some statistical summary representations using texture processing, in which early feature information is pooled across regions without requiring the segmentation of individual objects (Dakin & Watt, 1997; Malik & Perona, 1990; Parkes et al., 2001). One well-established case of such a computation is the extraction of average orientation from an ensemble of tilted lines where individual elements are too crowded to allow for the discrimination of individual orientations (Dakin & Watt, 1997; Parkes, et al., 2001). It has been argued that average size is

computed via a similar mechanism (Ariely, 2001; Chong & Treisman, 2003; 2005a; 2005b), but this claim has been more controversial.

In attempts to determine whether representing average size requires mechanisms similar to or distinct from those processes underlying individual object segmentation or texture processing, researchers have focused on the known signatures of visual processing that distinguish these modes of processing. For example, if the number of individual element sizes observers use in the computation of the mean size of an array surpasses the well-established 3-4-element limit of parallel attention (Pylyshyn & Storm, 1988; Scholl, 2001), it could be taken as evidence for a separate, global process. Indeed the precision of mean size judgment is typically found to be invariant over changes in array numerosity (Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2005b; Fouriezos, Rubinfeld & Capstick, 2008), a finding that some researchers have taken as evidence of an automatic process of perceptual averaging carried out by parallel “computers” (Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2005b; Chong, et al., 2008). The argument for a massively parallel, involuntary average size processing mechanism has gained additional support from multiple findings that observers can report the average size of an array of circles with fairly high accuracy even when they cannot recall the size of individual circles from the array (Ariely, 2001; Chong & Treisman, 2003; 2005a). The successful computation of average size in lieu of explicit knowledge of individual elements has been taken as additional evidence for a truly global, parallel average size processing mechanism.

Though some properties of average size processing have been identified, the computations underlying the ensemble representation of element size remain opaque. The claim that average size, like average orientation, is computed via a global process by which all elements in an ensemble are attended in parallel, and information is pooled across them, should

be viewed with some reservation. Advocates of this view have posited that all or nearly all elements in an array are taken into account when calculating average size, and thus the mechanism underlying average size processing must be separate from the one underlying individual object processing. However, it has been demonstrated that all published evidence construed as supporting this claim can be explained, at least in principal, through various focused-attention strategies and does not necessitate a special mechanism for global average size processing (Myczek & Simon, 2008; Simons & Myczek, 2008; Solomon, Morgan, & Chubb, 2011). For example, Myczek & Simons (2008) demonstrated that an ideal observer sampling just 2-3 elements from an array, can detect a 4-7% difference between the average size of an entire array of circles and a given test circle, levels of accuracy similar to those observed in ensemble average size tasks with human participants.

Given this evidence, findings that observers are unable to identify any individual member of an array (Ariely, 2001; Chong & Treisman, 2003; 2005a) need not be attributed to a global processing system in which no individual elements are selected. If observers are only selecting some subsample of elements within an array to enter into their computation of average size, it logically follows that they would be unable to identify most of the elements in a large array, as they were never attended to in the first place. Additionally, while the computation of average orientation can be understood in terms of a mechanism that pools across a set of receptors specialized for different orientations via a largely pre-attentive and parallel process, no individual receptors in the early visual system have been found to respond selectively to absolute sizes of objects (Myczek & Simons, 2008; Simons & Myczek, 2008). Thus, it is unclear what information could be pooled to produce a truly global average size percept.

### ***Ensemble processing of size and orientation in infancy***



The primary goal of this research is to determine whether the processes underlying average size and orientation ensemble processing rely on similar mechanisms with the same developmental trajectories. If automatic processes that derive from mechanisms similar to those subserving adults' representations of average orientation are available in infancy, infants' representations of element orientation should be more accurate for homogeneous arrays than for single item arrays because as information is pooled, the representation becomes less noisy. If the same is true for size – that is, if there are similar automatic processes involved in average size computation, then we should see better acuity for element size in homogeneous arrays than in single item arrays. Cordes and Brannon's (2011) results suggest that this is not the case, but does not necessarily refute the possibility. One caveat to previous comparisons of infants' size representations for single elements and elements in homogeneous arrays is that they were drawn across multiple studies using stimuli that varied in a potentially confounding way. For example, it is possible that the single Elmo faces used in Brannon et al.'s (2006) study were more interesting to infants than the homogeneous dot arrays used to test infants' acuity for multi-element size representations (Cordes & Brannon, 2011). Experiment 1 of the current paper will provide the first direct comparison of infants' size representations for single elements and homogeneous arrays using identical stimuli and a within subject design.

Additionally, while previous studies exploring infants' ability to detect a change in ensemble element size have used the habituation paradigm (e.g. Brannon et al., 2006; Cordes & Brannon, 2011), recent evidence suggests that the change detection paradigm may be more sensitive for testing the development of this ability. In an infant change detection task, a display is briefly presented (e.g. an array of three colored squares for 500 ms), then after a brief retention period (300 ms), a new array that may or may not contain a change is presented (e.g., one of the

squares is a different color). Infants' preference for stimulus streams involving some type of change over streams in which there is no change is taken as evidence that the infants have encoded the altered property of the objects in visual working memory (e.g., Ross-Sheehy et al., 2003).

It is possible that some features of arrays are more salient during the prolonged exposure time inherent in habituation tasks, and thus, previous studies may have biased infants to encode other properties of arrays at the expense of ensemble element size. For example, although previous studies using the habituation paradigm have found young infants unable to discriminate small ( $< 4$  items) arrays on the basis of numerosity (e.g., Clearfield & Mix, 2001; Feigenson et al., 2002a; Feigenson et al., 2002 b; Xu, 2003; Xu et al., 2005), recent results from a change detection task by Starr, Libertus and Brannon (2013), demonstrated that 6-month-olds are capable of making purely numerical discriminations over small arrays. Starr et al., suggested that the critical difference between their change detection study and previous habituation studies is the degree of attentional load involved in the two types of tasks. In the change detection paradigm, infants see two streams of rapidly changing information and must maintain representations of the previously displayed images in working memory to compare to the currently presented images. The habituation paradigm, which involves only a single stream of images with longer presentation durations, is considerably less dynamic.

In a study by Hyde and Wood, (2011), attentional load was shown to affect the way adults attended to and encoded numerosity information from multi-item arrays (Hyde & Wood, 2011). In low attentional load conditions, individual items in an array were spaced within the resolution of spatial attention such that each item could be individuated. Under these conditions, adults performing a numerosity change detection task exhibited neural correlates of object file

representations, suggesting that they were using focused attention processes to encode individual elements. In high attentional load conditions, elements were either spatially crowded such that they could not be individuated (Exp. 1), or spaced within the resolution of spatial attention but presented with a concurrent task in which participants were required to monitor dual rapid serial visual presentation (RSVP) streams while performing the numerosity change detection task (Exp. 2). When attentional load was high, adults exhibited neural correlates of the approximate number system (ANS), the ratio dependent system responsible for discrimination of large arrays (e.g. Dehaene, 1997), suggesting they were using distributed attention to encode the array as a whole. It is possible that while the habituation paradigm promotes the encoding of individual elements, the change detection paradigm, which heavily loads attention, inhibits the representation of individual object files and instead promotes ensemble representations. Given that 7-month olds may only hold one or two individual object representations in working memory simultaneously, (Kaldy & Leslie, 2003; 2005; Moher, et al., 2012; Ross-Sheehy, et al., 2003) the habituation paradigm previously employed to study ensemble element size representations may have biased infants to engage the object file system instead of a more efficient distributed attention process.

If average element size representation relies on a distributed attention process that is available early in development, infants in previous studies may have failed to detect a twofold change in size (Cordes and Brannon, 2011) because the habituation methodology biased them to attend to individual objects. Previous studies have demonstrated that infants tasked with encoding more than 3 objects exhibit catastrophic memory failure, such that they appear to remember *none* of the objects in the array (Feigenson, & Carey, 2003; 2005). Thus, infants attempting to encode all of the individual elements in the large arrays presented in Cordes and Brannon's (2011) study would have surely failed to detect a twofold change in the size of any

element. In Experiment 1, I use a change detection methodology and large arrays (5 items) to explore 7-month-olds' ability to represent homogeneous ensemble element size and orientation.

Alternatively, if average ensemble size representations rely on a sampling process that requires focused attention to individual objects, infants may have failed in previous studies because the number of items in the arrays exceeded their working memory capacity and they were unable to select and encode a subsample of elements within their capacity limits. In Experiment 2 of this paper, I use a discrepant color cue to highlight a single element within homogeneous size and orientation arrays. Given previous evidence that adults represent average orientation via a texture processing mechanism wherein no individual elements are attended (e.g., Parkes et al., 2001), this manipulation is unlikely to affect the accuracy of their ensemble orientation representations. If the process underlying element size representations relies on focused attention to individual items and infants' failure to encode element size in previous studies of this ability was due to an inability to attend to a single element, this manipulation may allow them to successfully detect a twofold change in ensemble element size.

### ***Experiment 1***

Previous studies have demonstrated that 6-month-old infants habituated to a single element (an Elmo face), can detect a twofold change in size (Brannon, et al., 2006); however, when habituated to arrays of homogeneously sized dots, infants of the same age do not dishabituate to test arrays in which all dots undergo a threefold change in size (Cordes & Brannon, 2011). Though 6-month-olds robustly detect a threefold change in the size of a single element (Brannon, et al., 2006), they require a fourfold change in element size to detect a change

in an array of homogeneously sized dots (Cordes & Brannon, 2011). These findings do not necessarily refute the claim that average size processing is an automatic, parallel and global process akin to the texture processing that underlies judgments of average orientation. It is possible that for any feature property, infants' representation of the average of a multi-element array will be less precise than their representation of the value for a single element.

This possibility is addressed directly in Experiment 1 of this paper by comparing the difference in infants' acuity for size representations across single element and homogeneous, multi-element arrays with infants' acuity for orientation representations across single element and homogeneous, multi-element arrays. Though 7-month-olds' threshold for detecting a change in the orientation of a single element has not been determined, we hypothesize that this threshold will be lower for multi-item homogeneous arrays than for single element arrays. At four months of age, infants can detect a 10° change in the orientation of a symmetrical shape (Bornstein, Krinsky, & Benasich, 1986); also, at four months the horizontal connections linking neurons with similar orientation preferences in visual cortex (Gilbert & Wiesel, 1989) start to emerge (Burkhalter et al., 1993). If the automatic process underlying adults' representations of average orientation are available in infancy, 7-month-olds in our study should benefit from the redundancy of homogeneous orientation arrays. If infants' discrimination thresholds for single-element and multi-element arrays vary differentially when either size or orientation is tested, it will suggest that these representations are supported by mechanisms with distinct developmental trajectories.

## ***Method***

## ***Participants***

Thirty-two 6- and 7- month-old infants ( $M$  age = 6.83 months,  $SD$  = .49; 20 male) participated, 16 in each of two conditions. Data from an additional **10** infants were excluded because of fussiness ( $n = 8$ ) or because infants did not look to both screens during at least one trial ( $n = 2$ ). Families were recruited from the Cambridge area and parents gave written informed consent to a protocol approved by the local institutional review board. Families received small gifts for their participation and \$5 in travel reimbursement.

## ***Design***

Infants were randomly assigned to either the size or orientation condition. In both conditions, infants were shown two streams of images, one on each of two monitors located peripherally (see Figure 5). Infants in both conditions were shown three trials of single element arrays and three trials of multi-element, homogeneous arrays. Whether infants saw the single element or multi-element arrays first was counterbalanced across participants and completion of at least one trial of each type was required for inclusion in analyses. On each trial, infants were shown one stream that alternated between two different images, and one stream that did not alternate. In the size condition, the area of the dots in the changing stream changed by a 1:2 ratio. In the orientation condition, the orientation of the elements in the changing stream changed by  $9^\circ$ , between  $40^\circ$  and  $49^\circ$ . In both conditions, neither stream varied in the numerosity or configurations of elements.

## ***Stimuli***

In all conditions, the visual streams consisted of images containing black elements on a white background that were presented for 500 ms followed by 300 ms of blank screen. A thin black rectangular border measuring 16-in across and 10-in in height was present in each image in all conditions and trials. There was no border present on the blank screens. In the size condition, stimuli were dots with diameters of 5cm (area= 19.6 cm<sup>2</sup>) in the small size arrays and 7.2 cm (area= 40.7 cm<sup>2</sup>) in the large size arrays. In single-element trials, the images in each stream contained a single dot presented in the center of the screen, and in multi-element trials, the images in each stream contained 5 identical dots in fixed locations that did not alternate between flashes. In the orientation condition, stimuli consisted of solid black bowling pin shapes, 12.5 cm in length and 3 cm wide, symmetrical over a central axis oriented either 40° or 49° from vertical. In single-element trials, the images in each stream contained a single bowling pin shape presented in the center of the screen, and in multi-element trials, the images in each stream contained 5 identical pins in fixed locations that did not alternate between flashes. Across single and multi-element trials in both conditions, the center points of all elements were fixed (*see figure 5*).

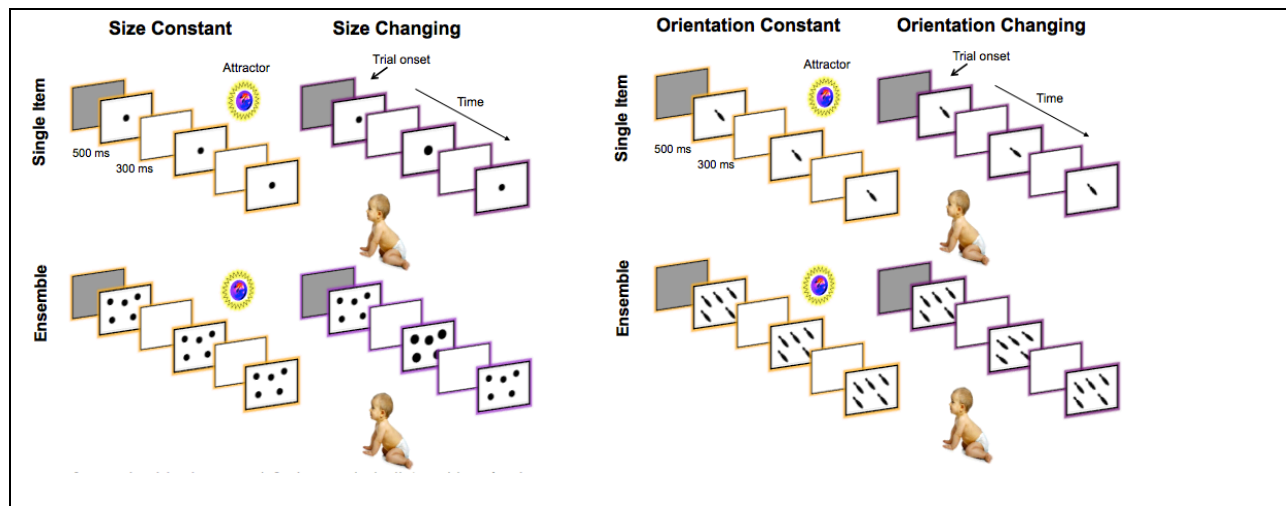
### ***Procedure***

Infants were seated on their parents' laps in a chair approximately 100 cm away from the middle of two 20-in. computer screens placed. The edges of screens were 10-in. apart. The room was darkened and classical music played at a low level throughout the experiment. At the beginning of each trial, the experimenter squeezed a noise-making toy and turned on a revolving light placed between the two screens. The experimenter manually initiated each trial when the infant looked at the attractor. Each trial lasted 60 s and each infant was tested with up to three

trials of single element arrays and three trials of multi-element homogeneous arrays. The image streams alternated sides between trials and the order was counterbalanced between infants.

Infants' looking behavior was digitally recorded for later offline coding. An experienced, naïve observer coded infants' looking to the screens using Supercoder software (Hollich, 2005). For each frame, (.033 s) of the trial, the coder identified whether the infants' eyes were oriented to the left, right, or neither image stream. A primary observer coded all infants. A second observer independently coded 25% of the infants and reliability between coders was extremely high (93% agreement, Cohen's  $K = .9$ ).

**Figure 5**



### ***Data Analysis***

In order to normalize the data and eliminate the effects of individual differences in overall attention to the stimuli, we analyzed preference scores calculated as the proportion of



time each infant spent looking at each of the two image streams as a function of each infant's total looking time to both screens. Preference scores are reported as a preference for the changing stream (i.e., percentage of looking time to the changing stream minus percentage of looking time to the no-change stream). An average preference score was calculated for each participant across a maximum of three trials. A positive score in each condition indicates a preference for the changing stream. For each condition, we performed a one- sample,  $t$  test (two-tailed) comparing preference scores with 0 and binomial statistics on the number of infants showing positive or negative preference scores.

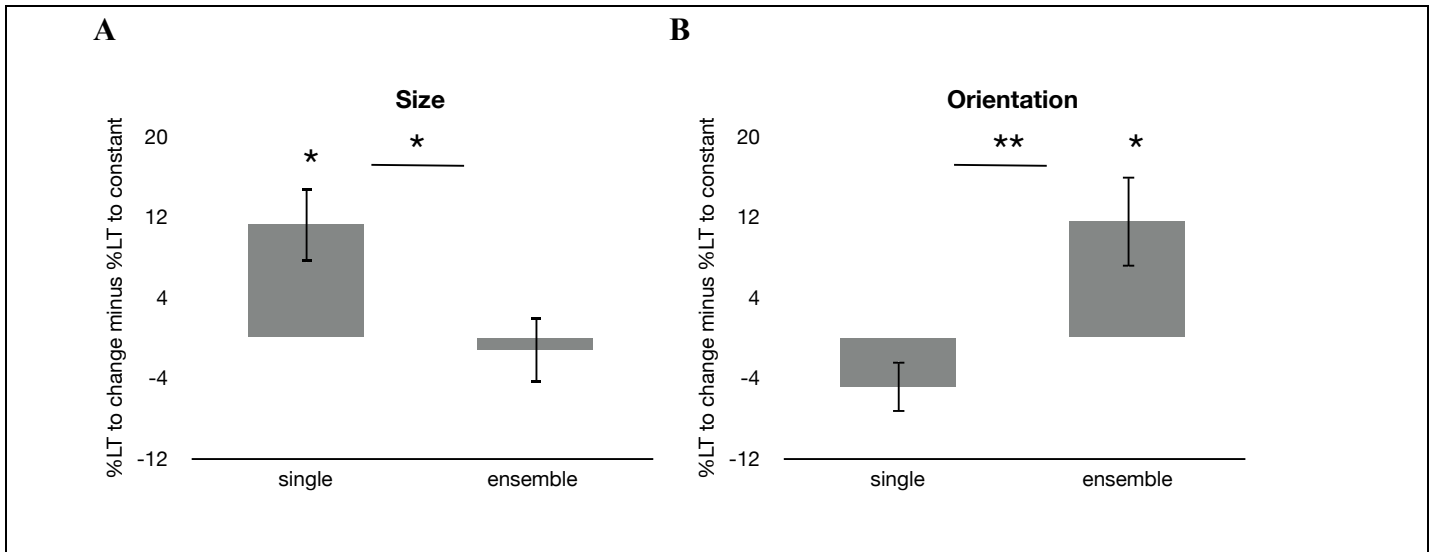
## ***Results***

In the size condition, on single element trials, infants preferred to look at the changing stream as compared to the constant stream,  $t(15) = 2.99, p < .01$ , and 12 out of 16 infants showed this preference ( $p < .05$ , one tailed binomial), but on homogeneous trials, in which all dots underwent a twofold change in size, infants showed no preference  $t(15) = -.36, p = .72$ ., with only 9 of the 16 infants looking longer to the changing stream ( $p = .40$ ) (see Fig. 6).

Contrastingly, in the orientation condition, on single element trials, infants showed no preference for the changing stream as compared to the constant stream,  $t(15) = 1.90, p = .08$ , with only 5 out of 16 infants looking longer to the changing stream ( $p = .11$ ), but on homogeneous trials, in which all elements shifted  $9^\circ$ , infants preferred to look at the changing stream,  $t(15) = 2.53, p < .05$ , with 12 out of 16 infants showing this preference ( $p < .05$ ). Paired samples tests revealed significant differences in preference scores between the single element and homogeneous ensemble trials in both the size condition,  $t(15) = 2.47, p < .05$ , and the orientation condition  $t(15) = -3.47, p < .005$ . Across subjects, independent t-tests revealed that preference scores for the

changing stream in single item trials of the size condition were significantly higher than preference scores for the changing stream in single item trials of the orientation condition  $t(30) = -3.544, p < .01$ . Additionally, preference scores for the changing stream in homogeneous trials of the orientation condition were significantly higher than preference scores for the changing stream in homogeneous trials of the size condition  $t(30) = 2.256, p < .05$ . In the size condition, there was no effect of whether the elements in the constant stream were large or small, and this was true for both single element trials,  $F(1,14) = 2.22, p = .16$ , and homogeneous array trials  $F(1,14) = .04, p = .84$ . In the orientation condition, there was also no effect of whether the elements in the constant stream were oriented at  $40^\circ$  or  $49^\circ$ , and this was true for both single element trials,  $F(1,14) = .09, p = .771$ , and homogeneous array trials,  $F(1,14) = 2.50, p = .14$ .

**Figure 6**



## Discussion

In Experiment 1, we demonstrate an asymmetry in the effect of introducing additional identical elements on 7-month-olds' acuity for size and orientation representations. We also provide the first behavioral evidence that 7-month-olds can detect 9° changes in the oblique orientation of an ensemble. Our finding that infants fail to detect a 9° change in single element trials but successfully discriminate 40° from 49° in multi-element, homogeneous arrays suggests that it is not the case that infants' ensemble representations of any feature are less precise than their representations of that feature on a single element. In fact, it appears that the introduction of additional identical elements improves 7-month-olds' acuity for representations of orientation, suggesting that the automatic pooling of orientation information across a multi-item scene that underlies adults' accurate average orientation representations, is online early in infancy.

Experiment 1 provides the first direct comparison of infants' size representations for single elements and homogeneous arrays using identical stimuli and a within subject design. Our finding that infants prefer the changing stream for single element arrays in the size condition replicates previous findings from the habituation paradigm showing that infants detect a twofold change in an element's size (Brannon et al., 2006). Additionally, our finding that infants do not prefer the changing stream for homogeneous arrays in the size condition replicates previous findings from the habituation paradigm (Cordes & Brannon, 2011), and rules out the possibility that the difference in infants' discrimination threshold for detecting size change across previous studies of single item and homogeneous arrays was due to confounds in the stimuli.

By employing a change detection methodology Experiment 1, we increase attentional load as compared to previous habituation studies (Cordes & Brannon, 2011). It has been suggested that this increase promotes global attention processes and suppresses object file representations, which rely on focused attention processes (Hyde & Wood, 2011; Starr, et al.,

2013). If the perception of average size relies on automatic, global processing mechanisms similar to those underlying representations of average orientation, it appears that these mechanisms are not functional at 7-months of age.

The difference in precision of infants' element size representations for single element and homogeneous arrays is somewhat surprising given that success in Experiment 1 requires attention to only a single element in each array. If infants were able to select just a single element, the task demands in the homogeneous trials would have been identical to those encountered in the single element trials. Previous research demonstrates that infants can benefit from redundant cues in discrimination tasks (e.g., Bahrick & Lickliter, 2000), but in the case of size, it appears that the presence of more than one element is detrimental to the precision of infants' representations.

Although infants in Experiment 1 did not need to attend to more than one element to detect the changing stream in the single element or homogeneous conditions, when presented with a large array of elements, it appears that they spread their attention across the entire image. The extraction of ensemble orientation appears to employ a parallel mechanism whereby information is pooled and averaged automatically. Because the precision of the sample mean improves with the square root of the sample size, the spread of attention over multiple identical elements actually benefits infants in representing ensemble orientation. Unlike ensemble orientation, however, it appears that ensemble element size is not computed under these conditions. Rather, it is possible that extracting element size requires focused attention, in a serial process with strict capacity limits. As such, the spread of attention across all five elements in the homogeneous arrays may have detrimentally impacted the precision with which infants represented element size in Experiment 1.

If infants attempted to encode more elements than could be processed under the capacity limits of focused attention (Moher, et al., 2012; Ross-Sheehy, et al., 2003), highlighting a single element within the array may help infants attend to and encode a single element, allowing them to detect a twofold change in size as they did in the single element trials. In Experiment 2, we test this possibility, highlighting a single element within a homogeneous array by changing its color to red (Fig. 7). Before selective attention can be deployed to select any individual element within an array, pre-attentive processing abstracts a representation of the image that can be used to guide attention (e.g. Treisman, 1985). Color singletons or patches of elements with a color distinct from surrounding elements have been shown to “pop-out” pre-attentively, and guide the subsequent deployment of focused attention (Bertin & Bhatt, 2001; Bhatt, Bertin, & Gilbert, 1999).

If the pre-attentive cue of a distinct color can capture and hold infants’ visual attention in Experiment 2, they should be able to encode the cued element’s size under focused attention just as they did in the single element arrays in Experiment 1. Alternatively, if the addition of a pre-attentive “pop-out” cue is not enough to counteract the spread of attention across all elements in an array, performance on size trials should mirror that of the homogeneous trials in Experiment 1. Unlike element size, which may be computed via focused attention (Myczek & Simon, 2008; Simons & Myczek, 2008; Solomon, Morgan, & Chubb, 2011), ensemble orientation appears to be computed via a textural analysis wherein there are no representations of individual elements. If orientation, like color, is processed pre-attentively (i.e. before selective attention), the highlighting of a single element should not alter infants’ processing of ensemble orientation, and their performance should not differ from that on homogeneous trials in the orientation condition of Experiment 1.

## ***Experiment 2***

### ***Method***

#### ***Participants***

Sixteen 6- and 7- month-old infants ( $M$  age = 6.59 months,  $SD$  = .50; 7 male) participated. Data from an additional 3 infants were excluded because of fussiness. Families were recruited from the Cambridge area and parents gave written informed consent to a protocol approved by the local institutional review board. Families received small gifts for their participation and \$5 in travel reimbursement.

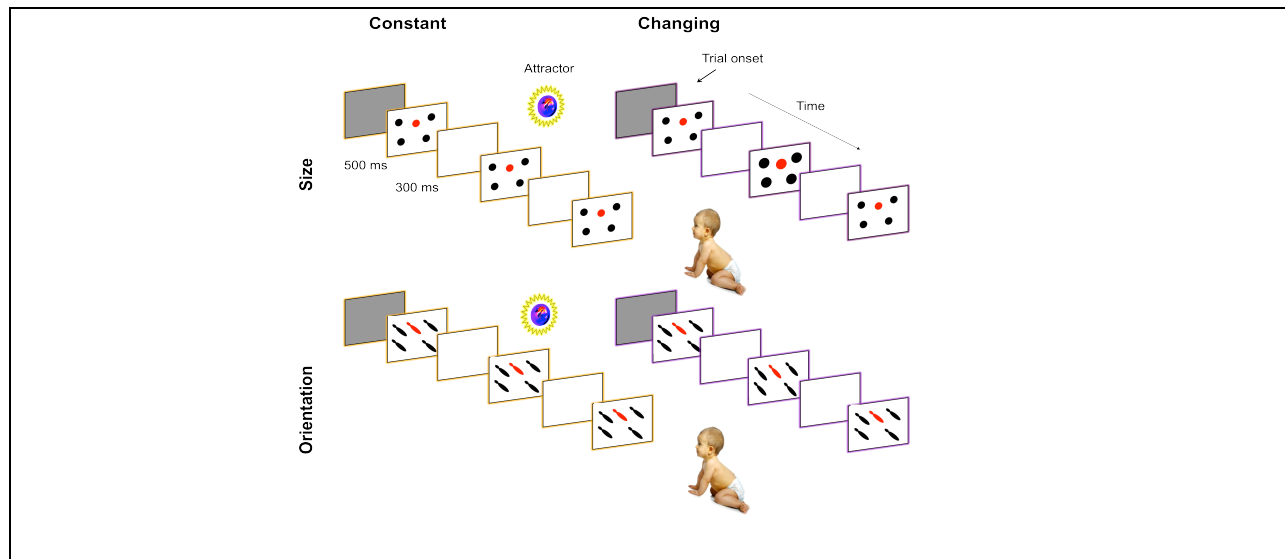
#### ***Design***

The design was the same as in Experiment 1, except we used only multi-element, homogeneous arrays, and each infant saw three trials of size comparisons and three trials of orientation comparisons. As in Experiment 1, trial types were blocked together, and whether infants first saw size or orientation trials was counterbalanced across participants.

#### ***Stimuli***

The stimuli were identical to those used in the homogeneous size and orientation trials of Experiment 1 with the color of a single element changed from black to red. The red element was always the same across all trials and did not change position from flash to flash within trial (see Fig. 7).

**Figure 7**



### ***Procedure***

The procedure was the same as Experiment 1.

### ***Data Analysis***

Preference scores were calculated as in Experiment 1. For each condition, we performed a one- sample *t* test comparing preference scores with 0 and binomial statistics on the number of infants showing positive or negative preference scores.

### ***Results***

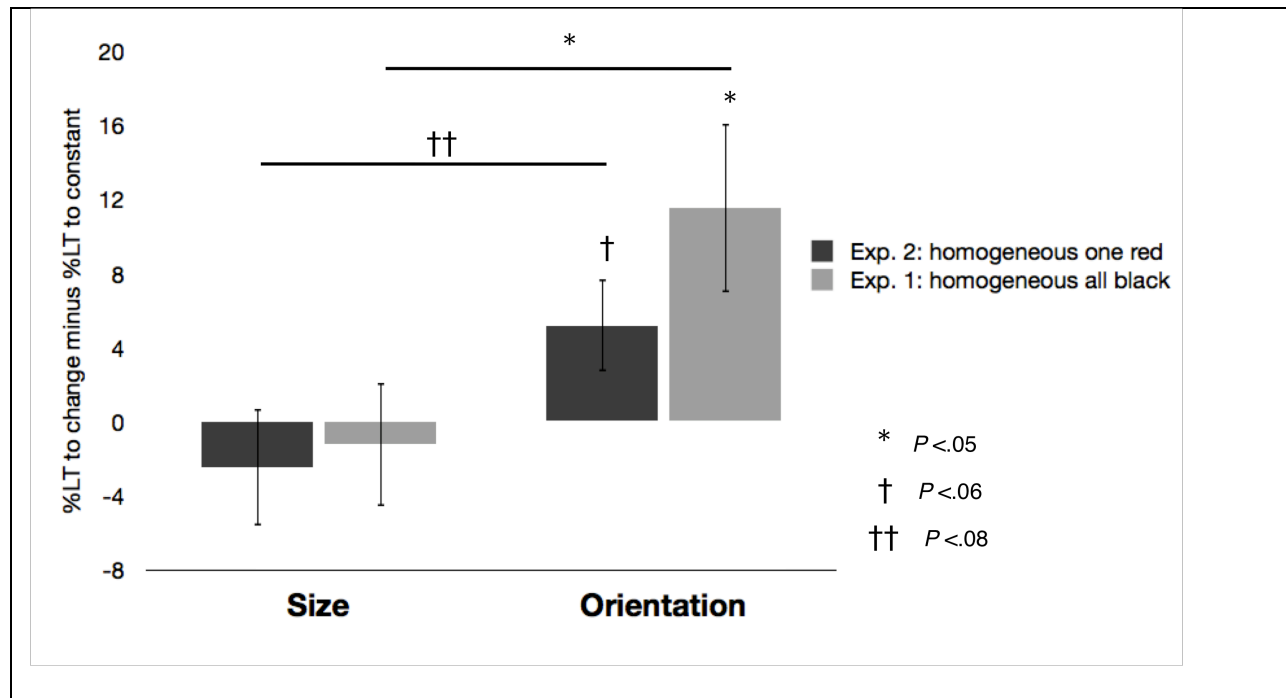
Infants failed to detect a twofold change in element size, but successfully discriminated 9° changes in the orientation of homogeneous arrays. On size trials, infants showed no preference for the changing stream as compared to the constant stream of images,  $t(15) = -.770, p$

= .45, with only 6 out of 16 infants looking longer to the changing stream ( $p = .23$ , one tailed binomial) (see Fig. 4). Contrastingly, on orientation trials, infants showed a marginally significant preference for the changing stream,  $t(15) = 2.08$ ,  $p = .055$ , with 13 out of 16 infants looking longer to the changing than the constant stream ( $p < .05$ ). A paired samples test revealed that the difference between infants' preference for the changing stream on size trials as compared to orientation trials was not significant,  $t(15) = 1.88$ ,  $p = .079$ . There was no effect of whether the images in the constant stream contained large or small elements,  $F(1,14) = .095$ ,  $p = .76$ , or elements oriented at  $40^\circ$  or  $49^\circ$   $F(1,14) = .08$ ,  $p = .77$ .

As depicted in Figure 8, the pattern of infants' preference for the changing homogeneous array on orientation, but not on size trials was the same in Experiments 1 and 2. Infants' preference scores for the changing stream on size trials in Experiment 2 were not different from infants' preference scores on homogeneous trials in the size condition of Experiment 1,  $F(1,30) = .07$ ,  $p = .79$ . Similarly, infants' preference scores for the changing stream on orientation trials in Experiment 2 were not different from infants' preference scores on homogeneous trials in the orientation condition of Experiment 1,  $F(1,30) = 1.50$ ,  $p = .23$ . Collapsed across Experiments 1 and 2, infants' showed a robust preference for the changing stream in homogeneous orientation trials  $t(31) = 3.196$ ,  $p < .005$ , and showed no preference for the changing stream in homogeneous size trials  $t(31) = -.802$ ,  $p = .43$ .



**Figure 8**



## Discussion

Infants' preferential looking patterns were not affected by the introduction of a color singleton in Experiment 2. As in Experiment 1, infants in Experiment 2 failed to discriminate a twofold change in element size in homogeneous multi-element arrays, and successfully detected a 9° change in element orientation in homogeneous multi-element arrays.

Given that orientation and color may both be processed pre-attentively (i.e. before selective attention) it is not surprising that this manipulation did not affect infants' ability to discriminate 9° changes in ensemble orientation. The orientation trials in Experiment 2 however, do offer a replication of our unprecedented initial finding in Experiment 1 that 7-month-olds can detect 9° changes in the orientation of homogeneous arrays.

Even though the configuration of elements was fixed, and the single colored element did not switch locations from flash to flash, infants did not encode the size of the single discrepant element. It is possible that infants in Experiment 2 performed like infants in Experiment 1 because they failed to detect the pop-out element all together. However, previous studies of discrepancy processing in infancy suggest that this is unlikely. When presented with arrays of a single + among Ls, 3-month-olds' saccades to the discrepant target item are equally fast (on the order of milliseconds) regardless of the number of distracters, suggesting that they are relying on global pre-attentive visual processes to guide their looking to the pop-out feature (Adler & Orprecio, 2006).

We suggest that while the pre-attentive cue of discrepant color attracted infants' attention in Experiment 2, it did not hold it long enough to lead infants to encode the pop-out element. Although some types of exogenous feature and spatial cues effectively capture infants' attention to result in a gaze shift (e.g., Bhatt, et al., 1999; Johnson & Tucker, 1996), this attentional capture does not always influence the encoding of cued items or lead to the enhanced binding of visual features into stable object representations. Previous studies have shown that while very young infants' attention is drawn to discrepant elements within large arrays, the presence of a pop-out feature differentially affects encoding at 3- and 5.5-months of age (Bhatt, et al., 1999). In one study on the effect of pop-out features on infant memory, researchers strung a ribbon from 3-month-old infants' ankles to an overhead crib mobile and trained them to kick their feet to make the mobile move. The training mobile was composed of seven blocks with the same black symbol (+, L, or T) on all faces. After 24 hours, infants were presented with a test mobile composed of one block that displayed the target symbol and six blocks that displayed novel distracter symbols. When the target on the test mobile had appeared on the training mobile, 3-

month-olds produced the trained kicking response; however, when the target on the test mobile was novel, infants' did not kick, despite the fact that all of the distracters on the test mobile were the familiar training symbols (Rovee-Collier, Hankins, & Bhatt, 1992). In a novelty preference procedure using looking time measures, Bhatt et al. (1999) replicated the finding that at 3-months, infants' recognition of a large visual array is solely determined by a single target's pop-out feature; however, they found that at 5-months, infants' recognition behavior was determined by the features of the more numerous distracter items.

Though discrepant color has been shown to attract visual attention at 5.5-months (Catherwood, Skoien, & Holt, 1996), it is possible that infants in Experiment 2 processed the visual array globally despite the pop-out item, or focused their attention on the pop-out item and then disengaged and attempted to encode additional items as individuals, overloading their working memory capacity. Future studies employing an eye-tracking methodology will be necessary to determine whether infants' attention was captured by the discrepant item at all. If ensemble element size processing is subserved by object file representations relying on focused attention mechanisms, the potential benefit of the pop-out item in Experiment 2 may have been counteracted by the use of a change detection paradigm. Future research should employ both the habituation paradigm, which has been shown to promote the encoding of object file representations (Hyde & Wood, 2011; Starr, et al., 2013), and the pop-out phenomenon to determine whether infants can successfully segment and encode a single element from a homogeneous array.

The results in Experiment 2 replicate those from the homogeneous trials in Experiment 1 and together these data suggest that the spread of attention across an array of similar elements, either via global attention or serial processing, detrimentally impacts the acuity with which

infants represent element size. Given that this distribution of attention appears to have an opposite effect on their ability to represent ensemble orientation, our findings suggest that computations of ensemble element size are supported by a qualitatively different mechanism.

### ***General Discussion***

Our results suggest that the processes by which infants extract size and orientation from elements in an array are distinct and follow different developmental trajectories. Specifically, in Experiment 1, 7-month-old infants detected a twofold change in the size of a single element but failed to detect the same change undergone by all elements within a homogeneous array. Contrastingly, infants failed to detect a 9° shift in a single element's orientation, but successfully discriminated homogeneous arrays where all elements shifted 9° in orientation. In Experiment 2, we explored whether infants' failure to detect a size change in the homogeneous arrays was due to an inability to select and encode an individual element. Including a single red element within the otherwise homogeneous arrays used in Experiment 1 did not alter infants' preferential looking patterns to multi-element arrays changing in size or orientation. As in Experiment 1, infants in Experiment 2 failed to discriminate arrays based on a twofold change in element size, and successfully discriminated arrays based on a 9° change in ensemble element orientation.

Compared to their acuity for representing the size or orientation of a single element, the presence of additional identical elements decreases infants' sensitivity to changes in element size and enhances infants' sensitivity to changes in orientation. These findings suggest that ensemble representations of size and orientation rely on different processes with distinct developmental trajectories. If representations of ensemble orientation are computed via a global, textural

analysis, then the single element and homogeneous ensemble trials in the orientation condition of Experiment 1 demonstrate the acuities of separate computational processes: one relying on focused attention and another relying on a parallel process of distributed attention.

If representations of element size are computed only via focused attention processes, then the single element and homogeneous ensemble trials in Experiment 1 show the relative accuracy of the same processing system under minimal (single item) and excessive (homogeneous ensembles) load. When presented with more items than can be maintained in working memory, one strategy is to select some subset of items to attend to and maintain representations of just these items. In fact, some evidence suggests that the ability to selectively ignore distracting or competing items is related to individual differences in adults' visual working memory capacity, with low capacity individuals being less able to weed out irrelevant sources of information (Fukuda & Vogel, 2009).

The ability to direct attention in the context of multiple competing inputs such that individual items can be selected and encoded from multi-item arrays appears to develop significantly between 6- and 10-months. For example, in a change detection paradigm, 6-month-old infants show no preference for changing over constant streams of arrays of three colored squares even though the colors of all three squares in the changing stream change color on each cycle (e.g., first they are red, blue and yellow, then they are green, white and pink) (Oakes, Messenger, Ross-Sheehy, & Luck, 2009). Note that, as in the case of Experiments 1 and 2 of the current paper, infants in this study could have succeeded by attending to any single item within the changing array. That they fail to do so suggests they were unable to remember any of the three items. Just a few weeks later, however, at 7.5-months, infants show a robust preference for

the changing stream in this same task, evidencing rapid development in working memory performance.

Previous studies have demonstrated that 6-month-olds succeed in a change detection task where arrays contain three objects that are all the same color, and all change to a different color on each cycle. Additionally, when these homogeneous changing arrays are pitted against changing arrays in which each of three heterogeneously colored squares changes color on each cycle, 6-month-olds prefer to look at the heterogeneous changing stream (Oakes, Ross-Sheehy, & Luck, 2006). This pattern of performance suggests that although 6-month-olds are unable to direct attention to specific individuals within arrays that presumably exceed their working memory capacity, they are sensitive to some global changes in these arrays.

It is possible that older infants, who are better able to direct their attention in the context of multi-item arrays, would be able to succeed in Experiment 2 of the current paper, and future research should explore this possibility to determine whether element size can be extracted from homogeneous arrays under these conditions. However, it is likely that for arrays of 5 elements, far more than can be handled by infants' working memory for individuals, even older infants would be unable to use the cue of discrepant color to select and encode a single item's size.

Previous findings have demonstrated that 6-month-olds, who do not encode the color of any items in change detection tasks with multi-item arrays (Ross-Sheehy, et al., 2003), show no difference in performance when a colored box is drawn around the changing item (valid cue) or around a non-changing item (invalid cue). Contrastingly, 10-month-olds, who already demonstrate a preference for a changing stream of three colored squares, benefit from this attentional capture manipulation on valid cue trials (Ross-Sheehy, Oakes, & Luck, 2011). This

suggests that in infancy, useful cues to a single valid item can only be exploited when working memory is not otherwise overloaded.

Recent studies of adult visual working memory have suggested that the computation of average ensemble element size may rely on a hybrid of focused attention processes and obligatory pooling of information across multiple elements where no individuals are represented. In a series of psychometric studies Allik and colleagues (2014) presented adults with a single reference dot and then four test dots, and asked observers to judge whether the mean of the four test dots was larger or smaller than the reference dot. On each trial, a constant number of pixels was added to or subtracted from the diameters of the four test dots, which initially had the same size as the reference dot. Across trials, the same summary size change was applied to one out of the four test dots (+4 units, +0 units, +0 units, +0 units), or spread across all four test dots (+1 unit, +1 unit, +1 unit, +1 unit). Adults were no more accurate in their mean comparisons on trials where 4 size units were added to the diameter of just a single dot than on trials where 1 size unit was added to each of the four test dots, suggesting that they were unable to use the size of a single discrepant dot to inform their judgments. In a second experiment, Allik et al. (2014) probed adults' representations of the individual dots' sizes directly, and found that when asked to recall the size of the single deviant dot, adults judged it as having the average size of all four test dots. Adults appear to obligatorily pool size information across multiple elements in an array, as they do in the case of orientation (Parkes, et al., 2001); but note that in this study, adults only averaged the size of four dots, a number of elements within their limits of focused attention and visual working memory (e.g., Cowan, 2001; Luck & Vogel, 1997).

Those who argue that average size and average orientation are computed via similar mechanisms under distributed attention often take as evidence for their claim findings that adults

report average size without explicit knowledge of any individual elements' sizes (e.g., Ariely, 2001). However, the question of whether an observer maintains access to representations of individual elements is orthogonal to the question of whether an observer computes average size via focused attention or distributed attention processes. Although adults appear to obligatorily average size information from multiple dots in an array (Allik, et al., 2014), they likely do so over a sampled subset of elements. Previous findings that adults' mean size judgments do not decrease in accuracy with an increase in array numerosity has been taken as evidence for a parallel process that extracts average size under distributed attention (Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2005b; Chong, et al., 2008). However, given that statistical aggregation should lead to improvements in mean size judgments, findings that accuracy remains flat as the number of judged elements increases actually suggest a decrease in precision (Allik et al., 2013). Additionally, previous studies in this vein that have demonstrated flat performance levels over increases in array size have increased array numerosity by adding more copies of dots of only four sizes rather than adding dots of new sizes (e.g. Ariely, 2001). Recent evidence suggests that when all dots in an array are heterogeneous in size, accuracy of mean judgments does in fact decrease as set size increases (Marchant, Simons, & deFockert, 2013).

Although the arrays of elements in our study were always homogeneous, our data are consistent with the argument that average element size is not computed under distributed attention. Unlike the previous study of infants' ensemble element size representations which employed a habituation methodology (Cordes & Brannon, 2011), the currently reported experiments use a change detection paradigm, which has been shown to promote global array processing, and still find that infants are unable to detect a twofold change in homogeneous array element size. In Experiment 2, however, infants were also unable to use focused attention



processes to encode the size of a single pop-out element within homogeneous arrays. It is possible that infants, like adults, (Allik et al., 2013; 2014), compulsorily pool information across multiple elements in an array even when they need only to attend to a single element; however, unlike adults, 7-month-olds lack the ability to select a subsample of elements within their focused attention and visual working memory capacity limits.

It is possible that the extraction of both ensemble orientation and element size are subserved by automatic processes available in infancy and continuous over ontogeny; but while ensemble orientation is calculated under distributed attention, the computation of element size relies on capacity limited focused attention processes. What develops, possibly, between infancy and adulthood is not the ability to average elements' size, but rather the ability to direct attention to a subset of elements within the capacity limits of focused attention and short-term memory. Additional studies varying array size and attentional cues in both the habituation and change detection paradigms are needed to determine whether infants and adults are processing ensemble element size in a qualitatively similar way. Future work exploring the development of top-down attentional control and the emergence of sampling abilities may help us better understand why the addition of identical elements detrimentally impacts infants' size representations and may shed light on the processes underlying average size computation later in life.

Our findings that infants detect a twofold change in an individual element's size but fail to detect the same change within homogeneous arrays replicates and extends previous findings from habituation studies exploring infants' representations of element size (Brannon, et al., 2006; Cordes & Brannon, 2011). Data from Experiments 1 and 2 of this paper show definitively that it is not the case that for any feature representation, infants' acuity decreases with the addition of identical exemplars; demonstrating that infants' ensemble representations of orientation are more

accurate than their representations of a single element's orientation. Taken together, our results are consistent with Allik et al.'s (2014) hybrid model in which the process underlying the extraction of ensemble element size automatically pools information across multiple elements, but relies on focused attention sampling strategies that are not available in infancy. Conversely, the process underlying representations of ensemble orientation is pre-attentive, automatic, deployed in parallel across all elements in an array, and present at 7 months.

## CONCLUSION

The primary goal of this dissertation was to characterize the way in which collections are efficiently encoded and maintained in working memory over ontogeny. Results provide evidence for remarkable continuities both in the ways in which collections that accrue over time can be hierarchically organized to increase working memory capacity and in the mechanisms underlying the efficient pooling of information across simultaneously presented collections (i.e. ensembles). This research also provides evidence that age related gains in cognitive flexibility and controlled attention lead to discontinuities in the ways various mnemonic data structures (i.e. individual object representations, chunks and summary representations of collections) afforded by the underlying architecture of working memory are employed.

The results reported in this dissertation bear on two main lines of enquiry regarding working memory representations of collections. First, they extend our understanding of collection (i.e., ensemble) representations in visual working memory to a different kind of working memory—one for dynamic events that unfold over time. Papers 1 and 2 demonstrate that in both young children and adults, this kind of summary representation can serve as a unit in working memory, evidencing a new continuity in the structure of the working memory system over the lifespan. Additionally, these papers show that when collections represented by this type of working memory overlap in features and statistical regularities arise, these collections can be hierarchically organized to increase working memory capacity. However, the ability to manipulate information in this manner requires executive function, and emerges with age-related gains in this suite of domain general cognitive capacities between the ages of three and seven.

The second line of enquiry concerns the processing of collections that are simultaneously available in visual arrays (i.e. ensembles). Paper 3 of this dissertation addresses an active debate in the adult visual working memory literature regarding the mechanisms underlying computations of average element size and orientation from multi-element arrays. My findings demonstrate an asymmetry in infants' discrimination thresholds for size and orientation in single-element and multi-element arrays, suggesting that these representations are subserved by qualitatively different mechanisms with distinct developmental trajectories.

Together, the results of papers 1, 2 and 3 of this dissertation provide evidence for continuities in working memory structure over ontogeny, but also highlight the ways in which age related gains in cognitive development lead to the emergence of discontinuous mnemonic representational capacities.

### *Main findings*

#### *Collections that accrue over time may serve as units in working memory*

Paper 1 confirms that collections of individuals that accrue over time in an intermixed fashion also constitute units of working memory, and that the limit of mutually exclusive collections that can be maintained in parallel is about three. Paper 2 demonstrates, for the first time, that this type of working memory structure is also available in young children, representing a novel contribution to the literature on continuities in working memory capacity over the lifespan.

While both infants and adults can represent multiple large collections of simultaneously presented items in visual working memory (i.e. ensembles), tasks like those used in papers 1 and 2, where collections accumulate one by one in an intermixed fashion over prolonged hiding events, present unique challenges for the working memory system. To succeed in these tasks, participants must compare the features of each item presented to those of previously presented items held in a working memory buffer, so that they may update the relevant collection representation or form a new collection representation while the others are maintained simultaneously. The finding that by three years of age children can encode three such collections demonstrates that continuities in working memory structures for collections are not limited to the realm of visual working memory, but extend to dynamic events that unfold over time.

*Statistical regularities among hierarchically organized collections can be used to increase working memory capacity*

Many studies of working memory show that participants can exploit statistical regularities in the input to recode information in a more efficient manner, thereby freeing resources to encode further items (e.g., Miller, 1956; Chen & Cowan, 2005; Kibbe & Feigenson, under review; Luck & Vogel, 1997; Brady et al., 2009). Paper 1 of the current dissertation demonstrates one new way in which this is so – here in the case of memory for collections that are presented over time. This series of studies explored whether adults could increase working memory performance on a task in which four conjunction-based collections (collections defined by a conjunction of features, e.g., blue triangles, or red circles) could be reorganized into intersecting feature-based collections (collections defined by only a single visual feature, e.g.,

triangles, or red items) using statistical regularities among the features. Paper 2 builds on the findings of paper 1 and explores the development of the ability to employ this efficient coding schema.

In these experiments (paper 1: Exps. 2-4; paper 2: Exps. 2 & 3), every object presented could be conceived of as a member of two collections— for example, a collection based on shape, or a collection based on color. This represents an important departure from previous studies in which each object clearly belonged to just a single collection (Feigenson, 2008, Experiment 1 of paper 1). In addition, we introduced statistical regularities between the feature dimensions such that, for example, the most common triangle was red and the most common circle was blue. Under these conditions, both adults and seven year old children exceeded the three-collection limit on working memory, representing the relative numerosities of four feature-based collections; however, children younger than seven failed to surpass previously identified capacity limits, suggesting a discontinuity in the ability to organize overlapping collection representations efficiently in working memory.

Additional experiments in paper 1 showed that the mere presence of statistical regularities in the input was not sufficient to improve adults' memory for collections. This suggests that seeing the experimenter highlight the stimuli's overlapping features before the hiding event (the variable that distinguished Experiments 2 and 3 in paper 1) played a critical role in adults' ability to organize information in a maximally efficient format. We propose that both adults in Experiments 3 and 4, and older children in paper 2 (who also saw the features highlighted before the hiding event) used top-down information to set up a hierarchical encoding scheme to represent the collections in each hiding location. On this account, after the experimenter highlighted the features of the objects prior to the hiding event, participants would

select the one featural dimension to serve as the bases for two feature-based collections that compose the top level of the hierarchy (e.g., could select color, thereby parsing the objects into red objects and blue objects). They then would represent these two feature-based collections in each hiding location (e.g., red objects and blue objects in location A, and red objects and blue objects in location B), and represent the approximate numerosity of each of the two collections in each of the two locations (e.g., about 16 red objects and about eight blue objects in location A; about eight red objects and about 16 blue objects in location B).

We know from previous studies using this paradigm that both adults (Feigenson, 2008) and children (Experiment 1 of paper 2 of the current dissertation) can accurately represent the approximate numerosity of up to three collections in each of two locations in parallel (Feigenson, 2008), and therefore that doing so for two collections is well within their capacity. The hierarchy model then requires that participants also represent information about the relative values along the orthogonal dimension(s) that are lower on the hierarchy (e.g., shape), nested within representations of each of the superordinate collections (e.g., red objects, blue objects). Frequencies of the feature-based collections along the secondary (and tertiary) dimension(s) may be represented in terms of average exemplars that are continually updated throughout the hiding sequence (e.g. the average red object is closer in shape to a triangle than to a circle), or by ratio information (e.g., more of the red objects were triangles than circles). Either representation could then be combined with the estimated numerosity of the superordinate collection to compute relative numerosity judgments of the secondary feature-based collections across locations.

### *Predictions of the hierarchical model for children and adults and future directions*

The hierarchical model we offer here makes several predictions. The first concerns the fact that we have not yet determined the upper limit of working memory capacity for fully intersecting structured collections. If the hierarchical model is correct, then there is no a priori reason that performance on our task should fall apart with the addition of a fourth feature dimension. Preliminary results from adults suggest that this prediction will be supported (see footnotes in paper 1), but more data is needed to support this surprising claim. Additionally, if the model is correct, as soon as children develop the ability to represent four collections hierarchically, they should be able to perform at the upper limits of this expansive ability. This prediction has yet to be tested but would offer additional evidence of the continuity of the working memory system over the lifespan.

The hierarchical model also predicts that we might be able manipulate performance on this task by influencing participants to set up the hierarchy with a particular feature dimension at the superordinate level. This prediction is supported by the results of paper 2 of the current dissertation wherein participants were instructed to label each bivalent object placed into a hiding location according to the features along one of two dimensions (e.g., “block, block, ball, block...” OR “small, small, large, large...” ). Paper 2 shows that while older children (seven-year olds) and adults, who performed equivalently, successfully answered questions both congruent and incongruent with the labeled dimension, their performance was significantly better in the congruent condition. Collapsed performance across the incongruent and congruent conditions for adults and older children was 88.3%, similar to the accuracy levels observed in paper 1, consistent with the hierarchical encoding model.



The last prediction of the hierarchical model that I will discuss concerns the mechanisms that would be required to support such a representation in working memory. The kind of working memory involved in our paradigm obviously involves switching among and updating currently held representations because information accumulates over time and in an unpredictable order, but evidence from paper 2 of the current dissertation shows that children as young as three can handle such updates. The observed developmental discontinuity in the ability to organize overlapping collections hierarchically most likely stems from the heavy demands on executive function this kind of information manipulation entails.

Young children share much of the basic architecture of working memory with adults (Cowan, 1997; Feigenson & Carey, 2003, 2005; Feigenson & Halberda, 2004, 2008; Moher, et al., 2012; Oakes & Bauer, 2007; Zosh & Feigenson, 2009), but are limited in the ability to strategically and flexibly establish efficient encoding schemes in working memory. The ability to flexibly attend to multiple features of a stimulus (e.g., color and shape), and the ability to sort items based on abstract categories are both necessary for success in this task, and develop significantly over the preschool years (e.g., Zelazo et al., 1996; Kirkham et al., 2003, and Kharitonova, Chien, Colunga & Munakata, 2009, respectively).

Paper 2 shows that while three- to five-year old children, like adults, can represent three mutually exclusive collections in parallel, the ability to use statistical regularities among intersecting collections to form a hierarchical representation and increase the total amount of remembered information does not appear until about age seven. Though infants demonstrate the ability to form hierarchical representations of chunks in working memory (Feigenson & Halberda, 2008; Rosenberg & Feigenson, 2013), hierarchically organizing heterogeneous collections (e.g. large items that are blocks and balls) in working memory requires the formation

of abstract (i.e. not based solely on perceptual similarity) membership criteria. These findings demonstrate that while the architecture of working memory may remain constant over ontogeny, developmental changes in executive function lead to important discontinuities in the way available mnemonic data structures are employed.

*Ensemble representations of orientation and size rely on qualitatively different computations*

The adult literature on ensemble representation debates whether computations of average size and average orientation rely on similar mechanisms of distributed attention. In paper 3, I provide evidence that these representations rely on qualitatively different processes with distinct developmental trajectories. Specifically, in Experiment 1, 7-month-old infants detected a twofold change in the size of a single element but failed to detect the same change undergone by all elements within a homogeneous array. Contrastingly, infants failed to detect a 9° shift in a single element's orientation, but successfully discriminated homogeneous arrays where all elements shifted 9° in orientation. Infants' pattern of performance in the orientation condition, namely their better acuity for homogeneous arrays than single item arrays, suggests that they are indeed averaging information from multiple items, and benefitting from the reduction of noise in the resulting representation.

That infants are less sensitive to uniform size changes in arrays of homogeneous dots than to size changes in a single dot appears to be in direct contrast with findings that adults' threshold for discriminating size change is the same for individual elements, and mean size judgments of homogeneous and heterogeneous arrays. One interpretation of this difference is that adults' ability to detect changes in average ensemble element size relies on ensemble

working memory representations that are not available at 7-months of age, representing a discontinuity in working memory over development. Alternatively, it is possible that the mechanism underlying adults' representations of average element size relies on cognitive capacities outside of working memory that have yet to develop in infancy, like the ability to effectively direct attention.

When presented with more items than can be maintained in working memory, one strategy is to select some subset of items to attend to and maintain representations of just these items. In fact, evidence suggests that adults' performance on size averaging tasks is consistent with the employment of just such a strategy (e.g. Myczek & Simons, 2008). For infants, however, the ability to direct attention in the context of multiple competing inputs such that individual items can be selected and encoded from multi-item arrays appears to develop significantly between 6- and 10-months (Oakes, Messenger, Ross-Sheehy, & Luck, 2009). If adults average size judgments rely on individual object representations, then the discontinuity we see in infants' ability to detect changes in ensemble element size are not due to a qualitative discontinuity in the type of working memory representations available over ontogeny, but rather a quantitative change in the ability to select individual elements from a multi-element array.

Experiment 2 of paper 3 explores whether infants' failure to detect a size change in the homogeneous arrays was due to an inability to select and encode an individual element. Including a single red element within the otherwise homogeneous arrays used in Experiment 1 did not alter infants' preferential looking patterns to multi-element arrays changing in size or orientation. Given that orientation and color may both be processed pre-attentively (i.e. before selective attention) it is not surprising that this manipulation did not affect infants' ability to discriminate 9° changes in ensemble orientation. The orientation trials in Experiment 2 however,

do offer a replication of our unprecedented initial finding in Experiment 1 that 7-month-olds can detect 9° changes in the orientation of homogeneous arrays.

We suggest that while the pre-attentive cue of discrepant color attracted infants' attention in Experiment 2, it did not hold it long enough to lead infants to encode the pop-out element. Though discrepant color has been shown to attract visual attention at 5.5-months (Catherwood, Skoien, & Holt, 1996), it is possible that infants in Experiment 2 processed the visual array globally despite the pop-out item, or focused their attention on the pop-out item and then disengaged and attempted to encode additional items as individuals, overloading their working memory capacity.

Recent studies of adult visual working memory have suggested that the computation of average ensemble element size may rely on a hybrid of focused attention processes and obligatory pooling of information across multiple elements where no individuals are represented (Allik, et al., 2014). Indeed the question of whether an observer maintains access to representations of individual elements is orthogonal to the question of whether an observer computes average size via focused attention or distributed attention processes. Although adults appear to obligatorily average size information from multiple dots in an array (Allik, et al., 2014), they may do so over a sampled subset of elements. It is possible that infants, like adults, (Allik et al., 2013; 2014), compulsorily pool information across multiple elements in an array even when they need only to attend to a single element; however, unlike adults, 7-month-olds lack the ability to select a subsample of elements within their focused attention and visual working memory capacity limits.

Although the arrays of elements in our study were always homogeneous, our data are consistent with the argument that average element size is not computed under distributed attention. In Experiment 2, however, infants were also unable to use focused attention processes to encode the size of a single pop-out element within homogeneous arrays. Previous evidence suggests that some exogenous cues, like motion, more effectively capture attention and more reliably lead to the encoding of the cued item (Ross-Sheehy, et al., 2011). It remains an open question whether infants could, given the right cue, selectively attend to and encode element size from a homogeneous array. If such a cue is found, future studies varying the heterogeneity of items and the number of cued items in step with age related gains in working memory capacity could test the hypothesis that the ability to represent average element size relies on a hybrid process of obligatory pooling over a subset of individual elements. As in the case of the ability to represent hierarchically organized overlapping sets (papers 1 and 2), the mechanism that supports efficient ensemble processing of average size may be available in infancy, but the ability to represent incoming information in a format optimal for this mechanism may only emerge with age related gains in cognitive control.

Much research in the field of cognitive development concerns the emergence of particular conceptual content such as knowledge of number, objects, space or the social world. In this dissertation I sought to study the development of working memory representations of collections using the same framework, namely by characterizing the initial state and continuities of abilities over ontogeny, characterizing important changes in abilities over development, and characterizing the mechanisms that underlie these discontinuities. In summary, the experiments reported here add to the mounting evidence that large collections can serve as units in working memory, and provide new evidence for additional continuities in working memory over

ontogeny: namely working memory representations of collections that accrue over time, and visual working memory representations of average orientation that rely on distributed attention processes. Additionally experiments in papers 2 and 3 highlight discontinuities in the way information can be efficiently organized to interact optimally with the underlying architecture of working memory. These findings add to the existing literature on cognitive development by extending our knowledge of the types of data structures available in working memory over ontogeny.

## REFERENCES

- Adler, S. A., & Orprecio, J. (2006). The eyes have it: visual pop-out in infants and adults. *Developmental Science*, 9(2), 189-206.
- Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world extracting statistical summary representations over time. *Psychological Science*, 21(4), 560-567.
- Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision research*, 83, 25-39.
- Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research*, 101, 34-40.
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and Visuospatial Short-Term and Working Memory in Children: Are They Separable?. *Child development*, 77(6), 1698-1716.
- Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in early childhood. *Journal of Experimental Child Psychology*, 87, 85-106.
- Alvarez, G.A. (2011) Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences*, 15(3), 122-131.
- 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(2), 106-111.
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, 19(4), 392-398.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12(2), 157-162.
- Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, 56, 149-178.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, 2, 89-195.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, 18(7), 622-628.
- Baddeley, A. D. (2001). Is working memory still working? *American Psychologist*, 56(11), 851.

- Baddeley, A. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society*, 4(05), 523-526.
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, 4(10), 829-839.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, 8, 47-89.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575-589.
- Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long-and short-term memory. *Journal of verbal learning and verbal behavior*, 9(2), 176-189.
- Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental psychology*, 36(2), 190.
- Bailey, C. E. (2007). Cognitive accuracy and intelligent executive function in the brain and in business. *Annals of the New York Academy of Sciences*, 1118(1), 122-141.
- Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, 86(3), 201-221.
- Bertin, E., & Bhatt, R. S. (2001). Dissociations between featural versus conjunction-based texture processing in infancy: Analyses of three potential contributing factors. *Journal of experimental child psychology*, 78(3), 291-311.
- Bhatt, R. S., Bertin, E., & Gilbert, J. (1999). Discrepancy detection and developmental changes in attentional engagement in infancy. *Infant Behavior and Development*, 22(2), 197-219.
- Bornstein, M. H., Krinsky, S. J., & Benasich, A. A. (1986). Fine orientation discrimination and shape constancy in young infants. *Journal of Experimental Child Psychology*, 41(1), 49-60.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory ensemble statistics bias memory for individual items. *Psychological Science*.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, 138(4), 487-502.
- Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, 83(3), 223-240.



- Brannon, E. M., Abbott, S., & Lutz, D. J. (2004). Number bias for the discrimination of large visual sets in infancy. *Cognition*, 93(2), B59-B68.
- Brannon, E. M., Lutz, D., & Cordes, S. (2006). The development of area discrimination and its implications for number representation in infancy. *Developmental science*, 9(6), F59-F64.
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage*, 5(1), 49-62.
- Broadbent, D. E. (1975). The magic number seven after fifteen years. In *Studies in Long Term Memory*. A. Kennedy and A. Wilkes (Eds.), Wiley.
- Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science*, 15(3), 118-121.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106(3), 551.
- Burkhalter, A., Bernardo, K. L., & Charles, V. (1993). Development of local circuits in human visual cortex. *The Journal of Neuroscience*, 13(5), 1916-1931.
- Cabeza, R., & Nyberg, L. (2000). Neural bases of learning and memory: functional neuroimaging evidence. *Current opinion in neurology*, 13(4), 415-421.
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental neuropsychology*, 28(2), 595-616.
- Catherwood, D., Skoien, P., & Holt, C. (1996). Colour pop-out in infant response to visual arrays. *British Journal of Developmental Psychology*, 14(3), 315-326.
- Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences*, 106(14), 5529-5533.
- Chen, L. (2005). The topological approach to perceptual organization. *Visual Cognition*, 12(4), 553-637.
- Chen, L., Zhang, S., & Srinivasan, M.V. (2003). Global perception in small brains: topological pattern recognition in honey bees. *Proceedings of the National Academy of Sciences*, 100(11), 6884-6889.
- Chen, Z. and Cowan, N. (2005). Chunk limits and length limits in immediate recall: A

- reconciliation. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, 31 (6), 1235-1249.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43(4), 393-404.
- Chong, S. C., & Treisman, A. (2005a). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, 45(7), 891-900.
- Chong, S. C., & Treisman, A. (2005b). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, 67(1), 1-13.
- Chong, S. C., Joo, S. J., Emmmanouil, T. A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics*, 70(7), 1327-1334.
- Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science*, 10(5), 408-411.
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, 386, 604-608.
- Cordes, S., & Brannon, E. M. (2008). Quantitative competencies in infancy. *Developmental Science*, 11(6), 803-808.
- Cordes, S., & Brannon, E. M. (2011). Attending to one of many: When infants are surprisingly poor at discriminating an item's size. *Developmental Psychology*, 2, 65.
- Cowan, N. (1997). The development of memory in infancy and childhood. Hove, U.K.: Psychology Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87-114.
- Cowan, N. (2010). The magical mystery four how is working memory capacity limited, and why?. *Current Directions in Psychological Science*, 19(1), 51-57.
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in brain research*, 169, 323-338.
- Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant Capacity in an Immediate Serial-Recall Task A Logical Sequel to Miller (1956). *Psychological Science*, 15(9), 634-640.
- Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology*, 51(1), 42-100.

- Crescioni, A. W., Ehrlinger, J., Alquist, J. L., Conlon, K. E., Baumeister, R. F., Schatschneider, C., & Dutton, G. R. (2011). High trait self-control predicts positive health behaviors and success in weight loss. *Journal of health psychology, 16*(5), 750-759.
- Denson TF, Pederson WC, Friese M, Hahm A, Roberts L. 2011. Understanding impulsive aggression: Angry rumination and reduced self-control capacity are mechanisms underlying the provocation-aggression relationship. *Pers. Soc. Psychol. Bull.* 37:850–62
- D'Esposito, M., Postle, B. R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Experimental Brain Research, 133*(1), 3-11.
- Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision research, 37*(22), 3181-3192.
- Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *JOSA A, 18*(5), 1016-1026.
- de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology, 62*(9), 1716-1722.
- Dehaene, S. (1997). *The Number Sense*. Oxford University Press.
- Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science, 311*(5759), 381-384.
- Diamond, A. (2013). Executive functions. *Annual review of psychology, 64*, 135-168.
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: cognitive functions, anatomy and biochemistry. In *Principles of Frontal Lobe Function*. DT Struss & RT Knight (Eds.), pp. 466-503. London: Oxford Press.
- Diamond, A. (1991). Frontal lobe involvement in cognitive changes during the first year of life. *Brain maturation and cognitive development: Comparative and cross cultural perspectives, 127-180*.
- Diamond, A. (2006). The early development of executive functions. *Lifespan cognition: Mechanisms of change, 70-95*.
- Drachman, D. A., & Arbit, J. (1966). Memory and the hippocampal complex: II. Is memory a multiple process?. *Archives of Neurology, 15*(1), 52-61.
- Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. New York: Columbia University Press.

- Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, 208(4448), 1181-1182
- Estes, W. K. (1973). Phonemic coding and rehearsal in short-term memory for letter strings. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 360-372.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General*, 128(3), 309.
- Feigenson, L. (2011). Objects, sets, and ensembles. In *Space, Time, and Number in the Brain: Searching for the Foundations of Mathematical Thought*. 13-22.
- Feigenson, L. (2008). Parallel non-verbal enumeration is constrained by a set-based limit. *Cognition*, 107, 1-18.
- 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., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive psychology*, 44(1), 33-66.
- Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, 13(2), 150-156.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, 8(7), 307-314.
- Feigenson, L. & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition*, 91, 173-190.
- Feigenson, L., & Yamaguchi, M. (2009). Limits on infants' ability to dynamically update object representations. *Infancy*, 14(2), 244-262.
- Fouriez, G., Rubinfeld, S., & Capstick, G. (2008). Visual statistical decisions. *Perception & psychophysics*, 70(3), 456-464.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological science*, 17(2), 172-179.
- Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at head start. *Developmental Psychology*, 47(2), 404.

- Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *The Journal of Neuroscience*, 29(27), 8726-8733.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology*, 40(2), 177.
- Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, 18(1), 1-16.
- Gentner, D., & Loewenstein, J. (2002). Relational language and relational thought. *Language, literacy, and cognitive development: The development and consequences of symbolic communication*, 87-120.
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 312–7 years old on a stroop-like day-night test. *Cognition*, 53(2), 129-153.
- Garavan, H. (1998). Serial attention within working memory. *Memory & cognition*, 26(2), 263-276.
- Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological bulletin*, 134(1), 31.
- Gilbert, C. D., & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *The Journal of Neuroscience*, 9(7), 2432-2442.
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, 17(17), R751-R753.
- Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time *Journal of vision*, 9(11), 1.
- Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, 17(7), 572-576.
- Halberda, J., Simons, D., & Wetherhold, J. (2006). You can never attend to more than three items at once: Gestalt grouping principles explain changes in capacity. *Manuscript submitted for publication*.
- Hauser, M. D., Carey, S., & Hauser, L. B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1445), 829-833.

- Hollich, G. (2005). Supercoder: A program for coding preferential looking (Version 1.5). *Computer Software]. West Lafayette: Purdue University.*
- Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2009). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology*, 35(1), 20-36.
- Hyde, D. C., & Wood, J. N. (2011). Spatial attention determines the nature of nonverbal number representation. *Journal of cognitive neuroscience*, 23(9), 2336-2351.
- Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, Psychophysics*, 75(2), 278-286.
- James, W. (1890). *The principles of psychology*. New York: H. Holt and Company.
- Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 683.
- Johnson, M. H., & Tucker, L. A. (1996). The development and temporal dynamics of spatial orienting in infants. *Journal of Experimental Child Psychology*, 63(1), 171-188.
- Káldy, Z., & Leslie, A. M. (2003). Identification of objects in 9-month-old infants: integrating ‘what’ and ‘where’ information. *Developmental Science*, 6(3), 360-373.
- Káldy, Z., & Leslie, A. M. (2005). A memory span of one? Object identification in 6.5-month-old infants. *Cognition*, 97(2), 153-177.
- Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130(2), 169.
- Kharitonova, M., Chien, S., Colunga, E., & Munakata, Y. (2009). More than a matter of getting ‘unstuck’: flexible thinkers use more abstract representations than perseverators. *Developmental Science*, 12(4), 662-669.
- Kharitonova, M., & Munakata, Y. (2011). The role of representations in executive function: investigating a developmental link between flexibility and abstraction. *Frontiers in psychology*, 2.
- Kibbe, M.M. & Feigenson, L. (under review). Infants use statistical regularities to chunk objects in working memory.
- Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension switching task. *Developmental Science*, 6(5), 449-467.

- Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, 21(1), 59-80.
- Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental science*, 13(6), 900-906.
- Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense large-number discrimination in human infants. *Psychological Science*, 14(5), 396-401.
- Linden, D. E., Bittner, R. A., Muckli, L., Waltz, J. A., Kriegeskorte, N., Goebel, R., ... & Munk, M. H. (2003). Cortical capacity constraints for visual working memory: dissociation of fMRI load effects in a fronto-parietal network. *Neuroimage*, 20(3), 1518-1530.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279-281.
- Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *JOSA A*, 7(5), 923-932.
- Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta psychologica*, 142(2), 245-250.
- Miller, G.A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development*, 13(3), 395-423.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology*, 41(1), 49-100.
- Moher, M., & Feigenson, L. (2013). Factors influencing infants’ ability to update object representations in memory. *Cognitive development*, 28(3), 272-289.
- Moher, M., Tuerk, A. S., & Feigenson, L. (2012). Seven-month-old infants chunk items in memory. *Journal of experimental child psychology*, 112(4), 361-377.
- Moffitt TE, Arseneault L, Belsky D, Dickson N, Hancox RJ, et al. 2011. A gradient of childhood self-control predicts health, wealth, and public safety. *Proc. Natl. Acad. Sci. USA* 108:2693–98

- Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration, dissociation, and decalage. *Developmental psychobiology*, 40(3), 255-265.
- Munakata, Y. (1998). Infant perseveration and implications for object permanencetheories: A PDP model of the AB task. *Developmental Science*, 1(2), 161-184.
- Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in cognitive sciences*, 15(10), 453-459.
- Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing cognitive control three key transitions. *Current directions in psychological science*, 21(2), 71-77.
- Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, 70(5), 772-788.
- Oakes, L. M. & Bauer, P. J. (Eds.). (2007) *Short- and long-term memory in infancy and early childhood: Taking the first steps toward remembering*. New York: Oxford University Press.
- Oakes, L. M., Messenger, I. M., Ross-Sheehy, S., & Luck, S. J. (2009). New evidence for rapid development of colour–location binding in infants' visual short-term memory. *Visual cognition*, 17(1-2), 67-82.
- Oakes, L. M., Ross-Sheehy, S., & Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychological Science*, 17(9), 781-787.
- Orban, G. A., Vandenbussche, E., & Vogels, R. (1984). Human orientation discrimination tested with long stimuli. *Vision research*, 24(2), 121-128.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature neuroscience*, 4(7), 739-744.
- Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of vision*, 13(8), 6.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, 3(3), 179-197.
- Riggs, N. R., Spruijt-Metz, D., Sakuma, K. L., Chou, C. P., & Pentz, M. A. (2010). Executive cognitive function and food intake in children. *Journal of nutrition education and behavior*, 42(6), 398-403.



- Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of vision*, 11(12), 18.
- Rosenberg, R. D., & Feigenson, L. (2013). Infants hierarchically organize memory representations. *Developmental science*, 16(4), 610-621.
- Ross-sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child development*, 74(6), 1807-1822.
- Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2011). Exogenous attention influences visual short-term memory in infants. *Developmental science*, 14(3), 490-501.
- Rougier, N. P., Noelle, D. C., Braver, T. S., Cohen, J. D., & O'Reilly, R. C. (2005). Prefrontal cortex and flexible cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20), 7338-7343.
- Rovee-Collier, C., Hankins, E., & Bhatt, R. (1992). Textons, visual pop-out effects, and object recognition in infancy. *Journal of Experimental Psychology: General*, 121(4), 435.
- Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80(1), 1-46.
- Scholl, B. J., & Xu, Y. (2001). The magical number 4 in vision. *Behavioral and Brain Sciences*, 24, 145-146.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1), 11.
- Simon, T. J., Hespos, S. J., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive development*, 10(2), 253-269.
- Simons, D. J., & Myczek, K. (2008). Average size perception and the allure of a new mechanism. *Perception & Psychophysics*, 70(7), 1335-1336.
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283(5408), 1657-1661.
- Spelke, E. S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development*, 21(2), 181-200.
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, 10(1), 89-96.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1-29.
- Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Infants Show Ratio-dependent Number

- Discrimination Regardless of Set Size. *Infancy*, 18(6), 927-941.
- Stuss, D. T. (2011). Functions of the frontal lobes: relation to executive functions. *Journal of the international neuropsychological Society*, 17(05), 759-765.
- Talmi, D., Grady, C. L., Goshen-Gottstein, Y., & Moscovitch, M. (2005). Neuroimaging the Serial Position Curve A Test of Single-Store Versus Dual-Store Models. *Psychological Science*, 16(9), 716-723.
- Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, 428(6984), 751-754.
- Todd, J. J., & Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive, Affective, & Behavioral Neuroscience*, 5(2), 144-155.
- Tuerk, A.S., & Carey, S. (2011, March). *Games with sets that match: Hierarchical representations of overlapping sets in working memory*. Poster presented at the meeting of the Society for Research in Child Development, Montreal, Canada.
- Uller, C., Carey, S., Huntley-Fenner, G., & Klatt, L. (1999). What representations might underlie infant numerical knowledge?. *Cognitive Development*, 14(1), 1-36.
- Uller, C., & Lewis, J. (2009). Horses (*Equus caballus*) select the greater of two quantities in small numerical contrasts. *Animal cognition*, 12(5), 733-738.
- Usai, M. C., Viterbori, P., Traverso, L., & De Franchis, V. (2014). Latent structure of executive function in five-and six-year-old children: A longitudinal study. *European Journal of Developmental Psychology*, 11(4), 447-462.
- Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, 411(6840), 953-956.
- Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, 44(2), 575.
- Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of experimental child psychology*, 108(3), 436-452.
- Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2010). The measurement of executive function at age 3 years: psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment*, 22(2), 306.
- Wood, J. N., & Spelke, E. S. (2005). Infants' enumeration of actions: Numerical

- discrimination and its signature limits. *Developmental Science*, 8(2), 173-181.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358(6389), 749-750.
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74(1), B1-B11.
- Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 458.
- Xu, Y., & Chun, M. M. (2005). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440(7080), 91-95.
- Yamaguchi, M., Tuerk, A.S., & Feigenson, L. (May, 2009), *Adults store up to 3 featurally-overlapping sets in working memory*. Talk presented at the meeting of the Vision Sciences Society, Naples, Florida, USA.
- Yerys, B. E., & Munakata, Y. (2006). When Labels Hurt but Novelty Helps: Children's Perseveration and Flexibility in a Card-Sorting Task. *Child Development*, 77(6), 1589-1607.
- Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, 11(1), 37-63.
- Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., ... & Carlson, S. M. (2003). The development of executive function in early childhood. Monographs of the society for research in child development, i-151.
- Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition*, 13(3), 193-201.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233-235.
- Zosh, J.M. & Feigenson, L. (2009). Beyond “What” and “How many”: Capacity, complexity, and resolution of infants’ object representations. In *The Origins of Object Knowledge*. Laurie Santos and Bruce Hood (Eds.), Oxford University Press.
- Zosh, J.M., Halberda, J., & Feigenson, L. (2011). Memory for multiple visual ensembles in infancy. *Journal of Experimental Psychology: General*, 140(2), 141.

APPENDIX 1

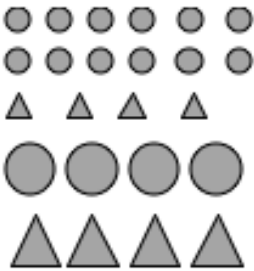
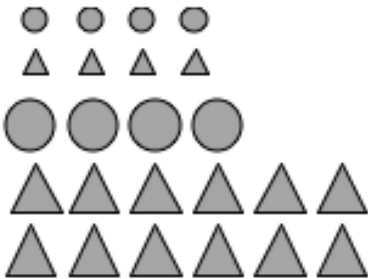
Experiment 1:

Bucket A	Bucket B
----------	----------



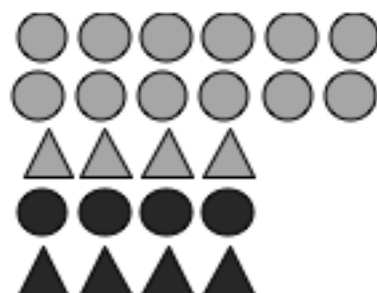
Experiments 2a & 3a:

Bucket A	Bucket B
----------	----------



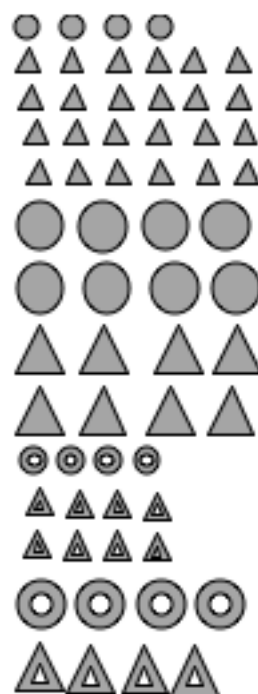
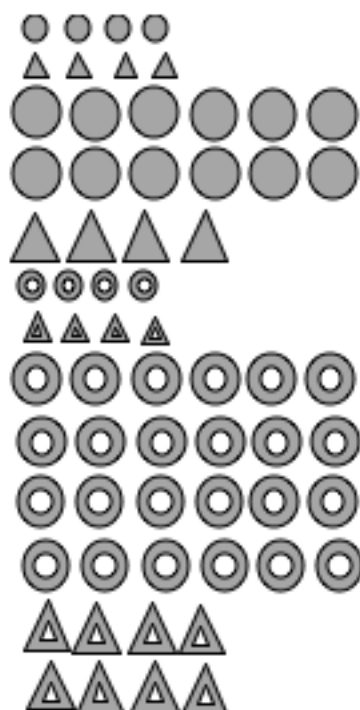
**Experiments 2b & 3b:**

Bucket A	Bucket B
----------	----------



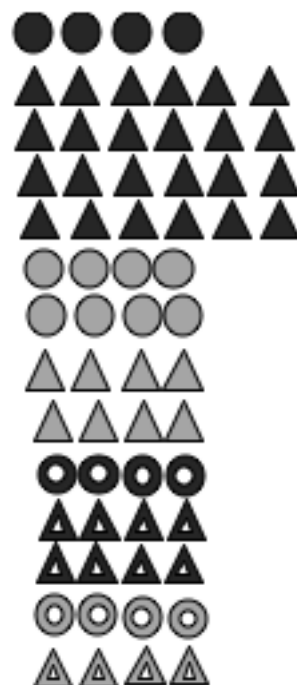
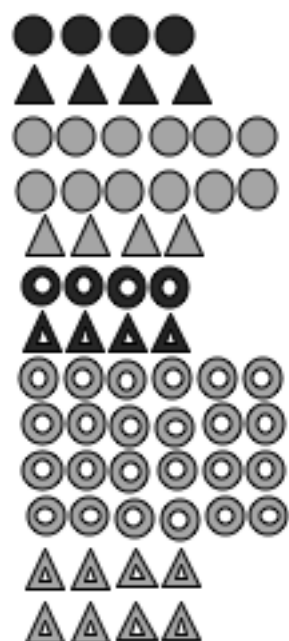
**Experiment 4a:**

Bucket A	Bucket B
----------	----------



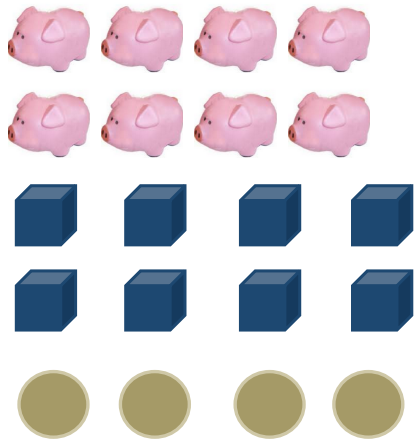
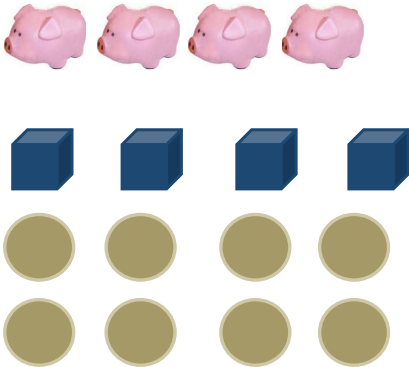
**Experiment 4b:**

Bucket A	Bucket B
----------	----------

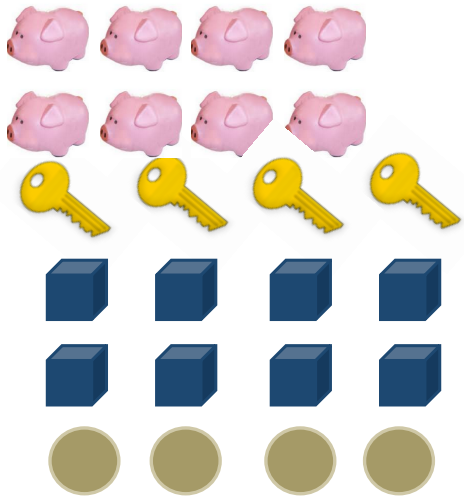
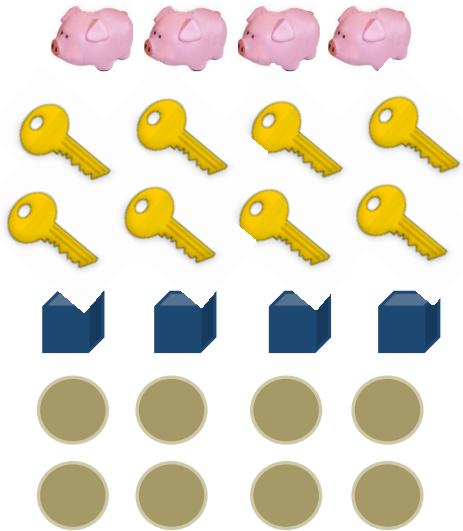


APPENDIX 2

Experiment 1a:

Bucket A	Bucket B
	

Experiment 1b:

Bucket A	Bucket B
	

Experiments 2 and 3:

Bucket A	Bucket B
----------	----------

