CHARTING THE BOUNDARIES OF INFANT WORKING MEMORY

By

MELISSA M. KIBBE

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Psychology

written under the direction of

Alan M. Leslie

and approved by

________________________
________________________
________________________
________________________

New Brunswick, New Jersey

[October 2011]
ABSTRACT OF THE DISSERTATION
Charting the boundaries of infant working memory

By MELISSA M. KIBBE

Dissertation director:
Alan M. Leslie

Visual working memory (WM), used for holding small amounts of information over brief intervals of time, undergoes rapid development in the second half of the first year of life. The research presented in this dissertation examines the nature of this development by testing the limits of infants’ WM in three different age groups, 6-month-olds, 9-month-olds, and 12-month-olds. A violation-of-expectation looking time method was used, in which sets of objects were hidden in multiple locations, and the contents of one of the locations was revealed to have changed, stayed the same, or disappeared completely. Both the number of locations, the number of objects, and the relevance of each object to the task were manipulated. The results show that infants’ WM can be characterized as limited, but flexible. Even in the face of severe limitations on what they can recall, 6-month-olds retain an inkling of an object. By 9 months, infants’ WM for what went where has improved, but breaks down when they have to keep track of more than two locations. By 12-months infants are better still, approaching adult-like levels of performance. Further, what infants recall depends on the relevance of the objects to a social agent. Implications for the structure and development of object WM are discussed.
Acknowledgments

I would like to thank Alan M. Leslie for his tireless support, indispensible guidance, and for his incredibly apt ship voyage metaphor. Special thanks to members of the Cognitive Development Lab, without whom data collection would have been impossible. I would also like to thank Eileen Kowler for her excellent guidance and advice, Randy Gallistel, Rochel Gelman, Renee Baillargeon, Katya Saunders, Deena Weisberg, Lu Wang, and Derek Anderson.

This research was supported by NSF DGE 0549115; NSF BCS 0725169, BCS 092218.
Table of Contents

Abstract of the Dissertation ......................................................... ii

Acknowledgments ................................................................. iii

I. Introduction ............................................................................ 1

II. What do 6-month-olds remember when they forget? .................. 29

III. Tracking what went where across multiple locations ............ 44

IV. Remembering the location of an agent’s goal ....................... 71

V. General discussion ............................................................. 96

VI. References ........................................................................... 108
# List of Illustrations

1. Sample of stimuli used in Sagi & Julesz (1985)…………………………………5
2. Stimuli and results from Luck & Vogel (1998)……………………………………7
3. Results of manual search task by Feigenson & Carey (2003)…………………..20
4. Sample of stimuli and results from Ross-Sheehy, Oakes, & Luck (2003)…….22
5. Method of Kaldy & Leslie (2003)……………………………………………….23
7. Chapter 2 method………………………………………………………………..33
8. Chapter 2 results: Looking time …………………………………………………36
9. Chapter 2 results: Cumulative distribution functions……………………….39
10. Chapter 3, three-location method………………………………………………48
11. Chapter 3, Experiment 1 results (three locations, difficult swap)……………….50
12. Chapter 3, Experiment 1 cumulative distribution functions…………………..51
13. Chapter 3, two-location method…………………………………………………56
14. Chapter 3, Experiment 2 results (two locations, difficult swap)………………..57
15. Chapter 3, Experiment 2 cumulative distribution functions………………….58
16. Chapter 3, three-location, easy swap method…………………………………61
17. Chapter 3, Experiment 1 & 3 results combined…………………………………63
18. Chapter 3, Experiment 1 & 3 cumulative distribution functions……………..64
19. Summary of results from Chapter 3, Experiments 1-3……………………….65
20. Chapter 3, Experiment 4 results (three locations, difficult swap)……………..67
21. Chapter 3, Experiment 4 cumulative distribution functions…………………..68
22. Chapter 4 method……………………………………………………………….86
23. Chapter 4 results: Anticipatory first looks……………………………………..88
24. Chapter 4 results: Looking time…………………………………………………90
25. Chapter 4 results: Cumulative distribution functions…………………………91
26. Hot spot plot showing infants’ gaze pattern during familiarization…………….92
27. Summary of the dissertation results..........................................................99
I. Introduction

Treisman (1996) defined a fundamental problem that the perceptual system faces when attempting to organize sensory information into meaningful, cohesive representations of the world. For example, to recognize your brown Nissan Altima out of several different cars in the lot, you must be able to accurately represent the color, location, size, and other features of your car, and not mistakenly mix those features up with the features of the car next to it. The brain needs to “bind” this information together to accurately represent the individual car. The “binding problem,” that is, the problem of deciding which bit of perceptual information belongs with which object or location, arises because processing of information about objects in the world is distributed across different areas of the brain and different levels of processing (Tootell, Dale, Sereno, & Malach, 1998). The binding problem poses a particular challenge to the human infant, who has limited knowledge of the world and is less able to rely on previous experience to make sense of the world.

Treisman laid out several different types of bindings, but only two of them are relevant to this dissertation. The first is property binding, where individual features, such as color, shape, and texture, must be bound to the object which they define. Property binding allows us to maintain a continuous representation of an object in the world and to identify that object as the same one over periods of time.

The second type of binding relevant to this dissertation is location binding. Here, an object must be bound to its location in space. Location binding allows us to keep track of the same object over time, even if the object is out of view. Again this poses a problem, because information about “what” an object is and “where” that object is in
space is processed in different areas of the brain, the former processed ventrally and the latter processed dorsally (e.g. Ungerleider & Mishkin, 1982; Goodale & Milner, 1992).

In this chapter (Chapter 1), I will discuss research on how property and location bindings are formed and how they are maintained over short intervals in visual working memory (WM), examining both behavioral and neuroimaging evidence. I will then discuss how infants deal with binding problems and review the current literature on infant visual WM.

I will show that the current literature on infant WM leaves a number of questions unanswered. To begin to answer these questions, I have conducted a series of experiments which I will describe in the remaining chapters of this dissertation. In Chapter 2, I ask what happens to the representation of a “forgotten” object in WM by testing 6-month-olds’ expectations about the persistence of an object for which they are unable to recall the specific shape. When infants lose the binding between feature and object, do they also lose the binding between object and location? In Chapter 3, I present four experiments which test 9- and 12-month-old infants’ ability to keep track of object identities across different locations. How do the number of hiding locations and the number of feature values infants must keep track of affect their ability to bind “what” with “where”, and how does this ability develop across the second half of the first year of life? In Chapter 4, I examine whether infants maintain different aspects of the display in WM when they observe an agent acting preferentially on one object versus another. In Chapter 5, I will summarize the findings and discuss their implications for understanding how objects are maintained in infant WM.
**Forming feature bindings**

How do features become bound together and tied to specific locations? Triesman (1996) notes that examining illusory conjunctions can provide some answer. Illusory conjunctions arise when features of one object are erroneously attributed to another object. Looking at the kinds of illusory conjunctions that arise under different conditions can give us some clues about what is required for binding.

Treisman and Schmidt (1982) showed that brief presentation of a target among distractors can result in illusory conjunctions between the features of the items in the display, suggesting that time may be an important factor in feature integration. But time does not appear to be the only factor. Features of objects that are in close proximity or that can be grouped together perceptually (e.g. inside the same contour) or conceptually (e.g. as part of a word) are more likely to be incorrectly bound (Prinzmetal, 1981; Treisman & Paterson; 1984; Prinzmetal & Millis-Wright, 1984). This suggests that spatial configuration provides an important cue to feature binding.

However, other evidence emerged that suggested proximity alone cannot explain illusory conjunctions. Pairs of multi-feature objects are more susceptible to illusory conjunctions if they both lie within the same focus of attention, regardless of how far apart they are (Cohen & Ivry 1989). Outside of the focus of attention, proximity is the factor that can give rise to illusory conjunctions. Thus, spreading attention across multiple locations can result in confusion of the features at those locations. In a similar finding, subjects made more errors when targets were displayed in the periphery where spatial resolution is poor, but made equally as many errors when targets were displayed foveally if attention was diverted by a competing task (Prinzmetal, Henderson, and Ivry,
1995). These results suggest that the veridicality of the representation is limited by the amount of attention allocated to binding.

Shifts in attention can also impact binding. Prinzmetal, Presti, and Posner (1986) tasked subjects with reporting the presence or absence of a target among three distractors. Prior to the onset of the target display, subjects were given a cue directing their attention to a location on the screen where the stimuli would appear (valid trial) or a different location (invalid trial). On target-absent trials, the distractors matched either the color or shape of the target, or a conjunction of both color and shape. Subjects made far more false alarms on target-absent trials following an invalid cue, especially when the stimulus contained a conjunction of the target’s features, providing further evidence for the role of attention in forming bound representations.

Much of the research described above blurs the distinction between property and location binding. Following Treisman and Gelade’s (1980) result that subjects were able to report the color of a simple target among distractors even if they were unable to report its location, several studies sought to separately test feature and location binding by looking at whether unbound features can be represented apart from a location. Using a similar method to Triesman and Gelade (1980), Nissen (1985; Isenberg, Nissen, & Marchak, 1990) also found that features can be represented independently of location, suggesting that binding features to locations does not occur automatically. However, in both Nissen (1985) and Treisman & Gelade (1980), features that were part of conjunctions were not represented independently of location. For example, in Nissen (1985) subjects were unable to report the shape of a colored target if they were unable to
localize the target (but see Johnston & Pashler, 1990, and Bundesen, 1991 for alternative interpretations of the findings).

The authors took these findings to suggest that features can be represented in the absence of location, but accurate representation of conjunctions of features required those features to be bound to locations. However, when detection of a target and identification of that target were explicitly teased apart, a different pattern of results emerges. Sagi and Julesz (1985) asked subjects to find two or more targets among an array of distractors and report whether they had the same orientation or color. They found that time to localize targets was independent of number of distractors, while time to discriminate targets by featural identity increased with number of distractors (see Figure 1). Sagi and Julesz suggest that localization of the targets can be done in parallel while identification of the targets relies on serial processes.

![Figure 1](image1.png)

Figure 1. a. Sample of stimuli used in Sagi & Julesz (1985; Figure 3). B. Detection of targets versus discrimination of targets, as function of number of targets (Sagi & Julesz, 1985; Figure 2). This plot shows the length of the SOA that produces 95% correct performance at detection and discrimination. Detection of the locations of the targets does not change with the length of the SOA, while discrimination, which requires integrating featural information and location information, requires longer SOAs as set size increases.
Indeed, studies designed to test subjects’ ability to track multiple independently moving targets have shown that a limited number of objects (about 4) can be tracked in parallel over space and time (Pylyshyn & Storm, 1988), even when the tracked objects disappear behind occluders and reappear again (Scholl & Pylyshyn, 1999). Since these objects are moving, their features must be bound to them independently of a specific location, because their location is constantly changing. Pylyshyn (1989; 2001) proposed that objects in the world are differentiated by their spatiotemporal and featural properties and are automatically indexed, without necessarily encoding the specific features of the object.

Once an object has been individuated and indexed, the features of the object can then be bound to that index. Kahneman, Treisman, and Gibbs (1992) proposed that object features are stored in an “object file” that is associated with the object’s index. This theory allows for an empty object file; that is, it is possible to index an object without encoding or maintaining its featural identity. These object files can be stored over short periods of time in visual working memory, to be retrieved as necessary.

*Maintaining feature bindings in visual working memory*

Visual working memory (WM) is used to store visual information temporarily so that it can be processed and used in cognitive tasks (e.g. Baddeley & Hitch, 1974). WM is essential for cognitive tasks in which visual data cannot be simultaneously foveated and for which multiple sources of visual information must be used. Yet WM is very limited in capacity (Pashler, 1988; Phillips & Christie, 1977; see Scholl & Xu, 2001 for a
review). A significant body of literature has been devoted to examining the nature of these limits by studying exactly what is being stored in WM.

In a now classic series of experiments, Luck and Vogel (1997) used a change detection task to show that about four objects could be reliably identified following a short interval of time in which the items were out of view. Further, simple objects, defined only by one feature, could be recalled just as accurately as complex objects, defined by a conjunction of four features (Figure 2). Luck and Vogel conclude that bound objects are stored in VSTM and are maintained over time.
Indeed, more recent work indicates fully integrated objects, and not independent features, are stored in VSTM. Gajewski and Brockmole (2006) presented subjects with an array of objects defined by conjunctions of color and shape and asked them to recall the color and shape of one of the items after a delay interval. Between presentation of the stimulus array and test, a valid cue (indicating the location of the object to be recalled) or an invalid cue (indicating a different location) was presented. Subjects performed significantly better on recall during valid cue trials, as was expected. More interesting was the pattern of errors on invalid cue trials. Subjects either recalled the object perfectly (accurate recall of both the color and shape of the item) or were unable to recall any of the features of the object, and only rarely recalled just one of the two features.

Xu (2006) found that the degree to which features were connected and proximal determined the extent to which they were stored as a bound object in WM. Indeed, spatial separation of objects reduces recall errors. For example, more objects can be encoded in WM when the objects are on different 3-D surfaces (Xu & Nakayama, 2007), suggesting that binding errors are reduced when features are grouped into objects that are highly separated from other objects. Jiang, Olson, and Chun (2000) manipulated the configurations of objects relative to each other, and found that the global configuration of the objects affected recall, suggesting that objects in WM are stored in relation to one another on a global level, just as features are stored in relation to one another on the object level.
Limitations on WM

While integrated objects appear to be stored in WM, there are limitations to how accurately objects can be recalled, depending on the properties of the features defining the object. Objects defined by a conjunction of different features, such as color and orientation, are recalled more accurately than objects defined by a conjunction of multiple dimensions of the same feature, such as two different colors, suggesting that only features from different dimensions can be encoded or recalled efficiently (Duncan, 1993; Xu, 2002; see also Magnussen & Greenlee, 1997, and Bilsky & Wolfe, 1995, for evidence of same-feature interference in discrimination and visual search tasks, respectively).

The demands of the task itself can impact which features of the objects are encoded. Droll and Hayhoe (2007) found that change detection for a feature of a multi-feature block was significantly better when that feature was being used at that moment in a block copying task. In a task in which subjects had to search through an array of hidden objects to find three that belong to a category, the cognitive complexity of the category impacted how much information was retained from each object; the greater the difficulty of the category, the more subjects had to search (Kibbe & Kowler, 2011).

Further, the featural properties of the objects in the to-be-remembered array also seem to impact what is recalled. Alvarez and Cavanagh (2004) showed that performance on a change detection task declined as the featural complexity of the stimuli increased. They suggested that WM may be limited to a fixed number of features as well as a fixed number of objects. However, their results may have been confounded by the fact that
increasing complexity in their stimuli also increased the similarity between the items. Awh, Barton, and Vogel (2007) tested WM capacity using a similar change detection paradigm but with stimuli which were designed to increase in complexity without changing in similarity. They found that stimuli with higher sample-test similarity resulted in decreased performance on a change detection task, but change detection was high for even the most complex stimuli, as long as sample-test similarity was low.

Familiarity with stimuli can bolster WM capacity for those items (Eng, Chen, & Jiang, 2006). Curby and Gauthier (2007) measured the effects of perceptual expertise on memory capacity for upright and inverted faces and cars and found that subjects showed significantly higher capacity estimates for upright stimuli versus inverted stimuli. This effect was stronger for upright faces than for upright cars which the authors interpreted as evidence that faces are subject to a more efficient, holistic encoding process. In a related study, Scolari, Vogel, and Awh (2008) found that perceptual expertise is improved by the fidelity of the representations, and not the number of objects that can be held in WM. Resolution can also be improved by longer viewing times, which presumably allows more time for more efficient encoding of items in WM (Chen, Eng, & Jiang, 2006).

**Efficient encoding in WM**

Most studies designed to test the limits of WM have concentrated on using highly discriminable stimuli defined by different features and dimensions. Recently, researchers have begun testing whether the process of encoding items in WM can take advantage of regularities in the stimuli to minimize storage of redundant information and thereby maximize the number of items that can be stored. Brady, Konkle, and Alavarez (2009)
designed a task in which stimuli were drawn from either a uniform distribution or from a distribution in which stimulus values co-varied probabilistically. On each trial, eight colored circles were presented in pairs. In one experiment, the colors were drawn from a uniform distribution, such that no compression was possible. In another experiment, pairs of colors co-occurred with high probability. Because the probability of co-occurrence was controlled by the experimenters, the minimum number of bits required to encode the colors could be computed (Huffman, 1952). Subjects’ task was to report the color of one of the patches indicated after a short retention interval. Based on subjects’ performance, Brady, et al. estimated the capacity of WM both in terms of number of colors and bits of information. They found dramatic increases in estimates of WM capacity when the stimuli exhibited statistical regularity, suggesting human observers were not only sensitive to the statistical properties of the display, but that they could use this information to store items more efficiently in WM. Furthermore, human WM capacity was well predicted by Huffman coding (Huffman, 1952), suggesting that what was being encoded in WM was maximally compressed (see also Mathy & Feldman, under review, for evidence of efficient encoding of number strings in phonological WM).

**Precision of WM representations**

Recall that binding objects to locations and binding features to those objects can be dissociated, with the former occurring automatically and in parallel while the latter requires focused attention on each object serially (e.g. Sagi & Julesz, 1985; Treisman & Gelade, 1982; Kahneman, Treisman, Gibbs, 1992). A similar dissociation can be observed in storage of bound items in WM (Darling, Della Sala, & Logie, 2009; Logie,
Wheeler and Treisman (2002), using the change detection paradigm of Luck and Vogel (1997), found no effect of set size when subjects were asked to detect a change in location, but found performance declined as set size increased when subjects had to detect a change in color and a change in location and color. This pattern of results suggests that locations can be stored in or retrieved from WM independently from identifying featural information.

Indeed, increasing the number of objects in the display can reduce the precision with which the items are recalled. Bays and Husain (2008) devised a task in which subjects had to report the degree to which an object had changed rather than simply detect the change. Subjects viewed 1, 2, 4, or 6 objects followed by a brief retention period. Then, a single display object was shown to have shifted slightly in location or to have changed slightly in orientation. These changes were small by design; traditional change-detection tasks used very large, suprathreshold changes such that detection performance was at ceiling for small numbers of items. As such, tradeoffs between number of items and precision of the WM representations may not be detectable since discrimination of the change would be relatively easy. Subjects’ task was to report the direction of displacement or rotation of the object (outward or inward, clockwise or counterclockwise).

Bays and Husain (2008) found that subjects’ accuracy varied probabilistically as set size increased. Indeed, responses were well-fit by a cumulative Gaussian function, the variance of which increased significantly with set size. For example, at set size 1, accuracy was high, with subjects reporting even relatively small changes in direction of displacement or rotation correctly 100% of the time. By contrast, at set size 6, accuracy
at even the largest stimulus changes never exceeded about 75%. Bays and Husain conclude that WM is a resource that can be allocated to some or all of the display contents (see also Huang, 2010).

In Bays and Husain’s (2008) task, performance did not vary with type of stimulus change (displacement or rotation), which Bays and Husain suggest indicates a common representational structure for these two aspects. Indeed, both position and orientation are location properties of objects rather than identifying properties. Zhang and Luck (2008) measured the precision of WM for color. In their task, subjects viewed 1, 2, 3, or 6 colored squares. After a brief retention period, unfilled squares appeared on the screen in the locations of the colored squares, and a heavy border around one of the locations indicated the probed object. Subjects had to adjust a pointer on a continuous color wheel indicating the color of the probed object.

Contrary to the findings of Bays and Husain (2008), Zhang and Luck (2008) found that performance was similar for small set sizes (1, 2, or 3 items), but dropped off dramatically at set size 6. They concluded that WM could be characterized as a small number of higher-precision “slots” rather than a pool of resources, with each slot possessing a fixed resolution (see Fukuda, Awh, & Vogel, 2010, for a review). However, subsequent studies found evidence that the resolution of these slots varied with set size, but only until the usual WM capacity of 4 was exceeded. When subjects were asked to recall the position of a gap in an object defined by contour, they showed decreased precision as set size increased from 1 to 3, but the precision plateaued as set size increased from 4-6, suggesting that only a subset of objects were encoded but that a limited WM resource had to be distributed across the 3 encoded items (Anderson, Vogel,
& Awh, 2011). When macaque monkeys were trained to saccade to the location of a changed color in a change detection task, performance declined as set size increased, but was always above chance even for larger set sizes (e.g. 4 or 5 objects), suggesting some information was retained from more items in the display than would be expected by chance (Heyselaar, Johnston, & Pare, 2011).

Increasing feature load of each object by presenting conjunctions also results in a decrease in precision of memory, while increasing the number of features in the display (by presenting objects defined by different features) impacts both the probability that an object is in WM as well as the precision with which it is recalled (Fougnie, Asplund, & Marois, 2010). This suggests that the feature load of individual objects does not affect whether they will be encoded, but does affect the precision with which they are encoded.

What drives these observed limitations? One possibility is that attention appears to be required for maintenance of bound objects in WM, but that attentional resources are limited. Johnson, Hollingworth, and Luck (2008) interspersed a demanding visual search task in the retention interval of a change detection task. At test, subjects had to detect either a change to a completely new feature or a change in which features belonged to which object. The attentionally-demanding search task had a detrimental effect on both detection of the featural change as well as detection of the change in feature bindings, suggesting that attention was required to maintain both independent and bound features in WM. Further, maintenance of items in WM can interfere with efficient visual search. When locations of objects must be retained in WM, visual search for a target is slowed, while WM for unbound features does not impede a visual search task (Oh & Kim, 2004; Woodman & Luck, 2004). Taken together, these results suggest that the maintenance in
WM of objects bound to locations and features bound to objects requires attention. Attention also exhibits a similar signature of loss of precision that WM exhibits; in a multiple-object-tracking task, the precision with which the direction of motion of each object was reported declined with the number of objects tracked (Horowitz & Cohen, 2010).

In sum, evidence suggests that WM is limited by both the content of the perceptual array as well as the attentional resources required to maintain bindings. Neural evidence also points to this as well. Xu and Chun (2006) found separate dissociable areas of the brain support WM encoding and maintenance. The inferior intraparietal sulcus (IIPS) represents a fixed number of items regardless of the featural identities of those objects, while the superior intraparietal sulcus (SIPS) and the lateral occipital complex (LOC) support encoding and maintenance of the identifying information and appear to be limited by that information.

**Binding in infancy**

Leslie, Xu, Tremoulet, and Scholl (1998) proposed the first model for binding featural information to objects and objects to locations in infancy. They note that the literature on infants up to that point had largely focused on infants’ expectations about how objects in the world should behave (e.g. Spelke, 1990; Baillargeon, et al., 1985; Kellman & Spelke, 1983; Baillargeon, 1991), while the adult literature had focused on object-based attention (e.g. Kahneman, Treisman, & Gibbs, 1992; Trick & Pylyshyn; 1994; Pylyshyn, 1989).
Crucially, they propose that the “object concept” in infants is supported by a visual indexing system, similar to that studied in adults. Object indexes are automatically initiated as objects come into view, and act as pointers to those objects. These pointers contain no featural information about the objects they index; rather they “stick” with the object as it moves about through space, even if the object goes out of view. Once the index has been established, featural information about the object can be assigned or bound to the index.

Thus, Leslie, et al.’s (1998) model distinguishes between individuation of objects and identification of objects. Objects are individuated when separate indexes are assigned to those objects. Objects can be individuated by spatiotemporal properties (e.g. location) or by featural properties (e.g. color). Successful individuation does not necessitate binding featural information to indexes. Once the indexes are assigned, the featural information used to initiate those indexes need not be retained. However, successful identification requires binding those features to the individuals (which infants as young as 1 month can do; Taga, Ikejiri, Tachibana, Shimojo, Soeda, Takeuchi, & Konishi, 2002), and maintaining those feature bindings over time. To illustrate, imagine you are watching a red car and a blue truck drive into a tunnel and emerge at the other end. If you have individuated the two vehicles, you should expect that two vehicles will drive out of the tunnel (and not 1 or 3 cars, for example). If you have identified the two vehicles, you should expect a red car and a blue truck to drive out of the tunnel (and not a blue car and red truck, or a motorcycle and a tractor-trailer truck).

This distinction between individuation and identification proves useful when studying the development of WM in infants. For example, Xu and Carey (1996) found
that 12-month-olds, but not 10-month-olds, could individuate by identity alone two objects that appeared alternately from either side of a screen on the same spatiotemporal trajectory (both age groups could individuate the objects given both spatiotemporal and featural cues). Xu and Carey suggested that infants needed to have conceptual knowledge of the objects in order to individuate them by identity alone. However, an alternative possibility is that once a feature is used to individuate an object, it does not necessarily have to be bound to that object. Thus, when a featurally distinct object appears to move on the same spatiotemporal trajectory, it will not be individuated as a separate object. Tremoulet, Leslie, and Hall (2000) showed that the features used to initiate an object representation are not necessarily bound to that object using a memory task. Features used to individuate are not necessarily bound to objects and stored in WM. It is this very important feature of WM that the research presented in this dissertation aims to explore.

Much of the work that will be discussed below requires infants to detect changes in objects that had been hidden from view. To detect a change to a physical object, infants have to not only encode location and identity, but also reason about the kinds of changes objects can undergo. Wang and Baillargeon (2007; Baillargeon, 2008) propose a three-system account of infants’ detection of impossible events. They suggest, as do Leslie, et al. (1998), that infants track objects via spatiotemporal information. Once the object is indexed, infants begin to form an object representation, in which features are bound to the object. These features can be basic (such as physical existence) or variable (such as height or color). When infants watch an event unfold, they use a physical reasoning system to monitor the event, using information about the object from their
object-indexing and object-representational systems, as well as core knowledge about objects, such as permanence. When an object is occluded, infants use their physical knowledge of “occlusion” and the relevant information from the object-index and object-representational systems to build a physical-causal representation of the event. Infants can thus reason about how events should unfold, and detect violations if events do not unfold as predicted. One of the essential aspects of this model that is most crucial to this dissertation is that, depending on the type of event, different information from the object representation will be used to reason about physical violations. For example, information about height may be used in reasoning about events in which an object is covered with a bowl, but not with a tube. Variable identifying information, such as color or shape, may not be used in reasoning about an occlusion event, but basic information such as the object’s existence might (Wang, Baillargeon, & Paterson, 2005; Chapter 2).

**Maintaining individuals in infant WM**

Evidence from multiple methodologies testing infants’ ability to track individuals and identities suggests that infants, like adults, are limited in the number of items they can encode or maintain in WM. Recall that, in adults, WM is limited to about 3 or 4 individual objects, with the amount of identifying information that is stored with each object varying depending on the perceptual properties of the stimuli or the demands of the task. Similar signature limitations can be found in infants. Let us first look at studies examining infants’ ability to track individuals, regardless of identity.

Infants can maintain a representation of what is hidden at a single location, and can update that representation. In Wynn’s (1992) study, infants saw a Minnie Mouse doll
occluded by a screen, then watched as another Minnie Mouse doll was placed behind the screen. Infants looked longer at the unexpected outcome of 1 or 3 dolls versus the expected outcome of 2 dolls, suggesting they could update their WM to accommodate the new object.

Infants are able to keep track of homogeneous sets across multiple locations as well. In Feigenson, Carey, and Hauser’s (2002) foraging task, 10- and 12-month-old infants watched as sets of graham crackers were placed in each of two different opaque containers. The sizes of each set varied, but one location always had more than the other location. They found that infants preferentially crawl to the container with more crackers, suggesting that they were able to represent the number of items at each location and compare those two representations. However, Feigenson, Carey, and Hauser found a distinct limit to infants’ performance. Infants only crawled to the container with the greater number of crackers when the number of crackers in either of the containers did not exceed 3. For example, infants preferentially chose 3 crackers to 1, but were at chance when the choice was between 2 and 4 crackers.

This limit of about 3 items also appears in other tasks. Feigenson and Carey (2003) obtained similar results with a manual search task. Twelve- and 14-month-old infants watched as an experimenter placed either 1, 2, 3, or 4 identical table tennis balls inside of a box whose contents were hidden from the infant by a felt flap. The experimenter removed one or more of the balls and showed them to the infant, at which point the infant was allowed to search in the box to retrieve any balls that were left (unbeknownst to the infant, the experimenter held back any of the remaining balls so that the infant could not reach them). Feigenson and Carey found that infants searched longer
when there were balls remaining in the box, but only up to set size 3. At set size 4, infants did not search longer than when the box was empty (Figure 3).

These studies suggest that infants have a WM limit of about three items, and that this limit does not change across the second half of the first year of life. However, these studies looked at infants’ ability to track homogenous sets. Recall that, in adults, individuation of items occurs in parallel while binding features to those individuals requires serial attention. Do infants show the same signature WM limits when asked to recall identities as when they are asked to recall individuals?

**Maintaining identities in infant WM**

Ross-Sheehy, Oakes, and Luck (2003), adapted the change detection task of Luck and Vogel (1997) for young infants. The display consisted of two screens. On each
screen was 1, 2, or 3 colored squares which appeared for a short interval, disappeared for a short interval and then reappeared. On one screen, the stimuli always remained the same. On the other screen, the colors changed. If infants were able to retain the color information across the short retention interval, they should prefer to look at the changing, and thus more interesting, screen. Ross-Sheehy, et al. found that infants 6.5 months of age preferred the changing stream at set size 1 only; by 10 months, infants preferred the changing stream even at the largest set size (see Figure 4). In a follow up study, Oakes, Ross-Sheehy, and Luck (2006) asked whether infants could bind features to location. In that study, set size was held constant at 3, and colored squares in the changing stream changed location rather than color. Here, 6.5-month-olds did not prefer the changing stream, while 7.5-month-olds did (see Oakes, Hurley, Ross-Sheehy, & Luck, 2010, for similar findings for location changes rather than color changes). However, because all of the items change on each trial, it is difficult to tell whether infants 7.5-months and up are succeeding because they are attending to all the items in the display, or just a subset of those items. This “sampling problem” makes it difficult to estimate infants’ WM capacity.
To address the sampling problem, Kaldy and Leslie (2003; 2005) asked whether infants could keep track of which object was hidden at separate locations. In their task, infants were familiarized with either two shapes (a triangle and a disk) or two colors (red and green). On each familiarization trial, the order of presentation and the positions of the objects were counterbalanced so that no long-term bindings between object and location could be formed. During test trials, the objects were hidden one at a time, and then one of the objects was revealed to have changed (unexpected outcome) or stayed the same (expected outcome) (Figure 5). Thus, infants’ memory could be tested for each item separately, either by revealing the most recently hidden item (easier to remember because it was hidden last) or the first-hidden item (more difficult to remember because after it was hidden infants’ attention was drawn to the intervening object). Using this methodology reveals that 6.5-month-old infants are only able to recall the shape of the
easier-to-recall object, while 9-month-olds can recall the shape of both objects (but not their colors).

Kaldy and Leslie (2003; 2005; Leslie & Kaldy, 2007) conclude that 6.5-month-old infants have a WM span of about 1 identity, but by 9 months of age, this span increases to about 2 identities. These estimates are similar to those of Ross-Sheehy, et al. (2003), though they disagree with the results of Oakes, et al. (2006). It seems that there is
mounting evidence that infants’ WM for *identities* develops across the second half of the first year, while infants’ WM for *individuals* remains constant at about 3. However, since this evidence is spread out across many different methodologies and studies, it is difficult to draw concrete conclusions about the development of infants’ WM abilities. It is unclear whether perceptual or attentional demands impact performance on the tasks as well. In tests of iconic memory, infants’ capacity estimates are similar to adults (5 items vs. 5.7 items) but their performance, unlike adults, is significantly below ceiling (Blaser & Kaldy, 2010), suggesting that their performance may be impacted by task artifacts such that it is difficult to compare performance across tasks.

*Efficient encoding in infant WM*

Recall that adults’ WM ability can be impacted by other factors, such as the discriminability of the stimuli (Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007), familiarity with the stimuli (Curby, et al., 2009), and statistical regularity of the stimuli (Brady, Konkle, & Alvarez, 2009). Recent work has begun to test whether infants’ WM capacity is sensitive to these factors.

Feigenson and Carey (2005) explored whether infants retained any information about the graham crackers in the buckets when their WM limitation was exceeded (Feigenson, Carey, & Hauser, 2002). They found that, where infants failed to discriminate a bucket containing 1 cracker and a bucket containing 4 crackers, they succeeded at discriminating 0 versus 4 crackers, as well as 1 small cracker versus 4 large crackers. Feigenson and Carey (2005) concluded that infants are retaining some information about the crackers, even if they fail to represent the exact number of crackers
when their apparent limit of 3 objects is exceeded. Indeed, changing the characterization of the choice (e.g. between “big” and “small” or between “some” and “none”) seemed to impact infants’ performance.

Controlling the size of the change in these displays proves to be important. Kaldy and Blaser (2009) used a forced-choice preferential looking paradigm (Teller, 1978) to determine the salience of color, luminance, or shape changes versus a baseline stimulus. They found that different features had different salience profiles, and that these profiles differed between 6- and 9-month-olds. For example, 9-month-olds found smaller changes in color more salient versus the baseline than 6-month-olds, who needed much larger changes. These results suggest the possibility that, in previous studies where 9-month-olds recall an identity that 6-month-olds do not, the change in the stimulus may not be salient enough for 6-month-olds. This might reflect a perceptual or attentional limit, rather than a WM limit. Or, it could reflect a resolution limit similar to that found in adults (see section, above). This could mean that infants’ WM capacity for identities is more severely limited in resolution than adult WM, but would represent as many identities as adults under the right conditions.

Infants, like adults, also seem able to use regularities in the stimuli to help them encode more items (like adults; Brady et al., 2009; Curby, et al., 2009). Feigenson and Halberda (2008) found that 14-month-old infants could use “chunking” cues to help them remember more items in a manual search task. These cues were conceptual (e.g. two cars and two cats), linguistic (e.g. referring to sets of identical objects using different labels), or spatial (e.g. separating sets of identical objects). All of the cues resulted in infants
succeeding with larger set sizes, in contrast to previous studies showing that infants fail to discriminate set sizes greater than 3 (Feigenson & Carey, 2003).

Further, infants appear to selectively encode only some of the features of an object, depending on the task (like adults; Droll & Hayhoe, 2007; Kibbe & Kowler, 2011). Yoon, Johnson, and Csibra (2008) asked whether 9-month-old infants will selectively recall information about an object depending on the kind of communication cues they get from an adult agent (Figure 6). Infants watched an agent make one of two motions to an object, either a non-communicative reaching action or a pointing action typical of a communicative gesture. A curtain was then lowered over the agent and two opaque screens were placed on either side of the stage, one occluding the object. When the screens were removed, infants either saw the object change identity (from one type of toy to another) or change location (from one side of the stage to the other). Infants looked significantly longer at the identity change condition than the location change condition when the agent had pointed to the object. Conversely, infants were more surprised by a location change than an identity change when a non-communicative reaching motion had been made toward the object. Yoon et al. concluded that communicative gestures caused infants to attend more to identity, where non-communicative gestures caused infants to attend more to location.
Figure 6. Stimuli used in Yoon, Johnson, & Csibra (2008; Figure 1). Infants who observed the communicative gesture looked longer at the identity change but not at the location change versus a control. Infants who observed the non-communicative reaching gesture had the opposite looking pattern.
Overview of dissertation research

The research presented in this dissertation examines the development of infants’ ability to maintain individuals and identities in WM.

The experiments described in Chapters 2 and 3 of this dissertation offer a systematic developmental approach to the study of infants’ developing ability to keep track of what went where. The experiment in Chapter 2 sheds light on the nature of what is encoded in WM by asking whether infants recall the individual even if they are unable to recall the identity of an object. Then, in Chapter 3, 9- and 12-month-olds WM capacities are tested under conditions where sampling confounds are controlled and both set size and number of locations infants are required to keep track of are varied. The goal of these studies is to systematically tease apart WM for individuals and identities and to examine the constraints on WM capacity as it develops.

In Chapter 4, I ask whether infants can use their understanding of agents to help them recall a shape they have been previously shown to forget (Chapter 2; Kaldy & Leslie, 2005). If the to-be-remembered object is the goal of an agents’ action, infants may be more likely to encode its shape. However, it is possible that they then may not recall its location. The studies in Chapter 4 test these questions by using the same experimental set-up as in Chapters 2 and 3, but with an agent present.

In Chapter 5, I will summarize the findings and discuss them in the context of current adult and infant WM literature, then suggest some directions for future research.
II. What do 6-month-old infants remember when they forget?

Introduction

Infants have a limited ability to represent and recall objects in detail. Current evidence suggests that during the first year infants can track up to about three hidden objects, close to the adult limit of about four (e.g., Scholl & Pylyshyn, 1999), and that this limit does not change between 5 and 12 months of age (Cheries, Wynn & Scholl, 2006; Feigenson, Carey, & Hauser, 2002; Fiegenson & Carey, 2003; for reviews, see Chen & Leslie, 2009 and Feigenson, 2007).

Another body of literature, following Xu and Carey (1996), suggests that infants’ memory for objects might be much more limited if we examine what infants can recall about the objects they are tracking. Young infants are often only able to recall the identifying properties of a subset of hidden objects in a display. For example, Káldy and Leslie (2003; 2005) asked infants to keep track of objects hidden sequentially in two different locations. They found that 6.5-month-old infants could recall the shape of only the most recently hidden of the two objects (2005), while 9-month-olds could recall the shape of the first-hidden (harder-to-recall) shape but not its color (2003) (for sets of visible objects recalled over very brief intervals, see Ross-Sheehy, Oakes, & Luck, 2003).

What happened to the forgotten second shape in Káldy and Leslie’s (2005) study? One clue may be that the spatiotemporal individuation of objects appears to take priority over identifying features, such as color and shape, (Leslie, Xu, Tremoulet & Scholl, 1998; Spelke, Breinlinger, Macomber, & Jacobson, 1992; Xu & Carey, 1996). Mareschal and Johnson (2003) found that 4-month-old infants had difficulty recalling identifying features.

1 Chapter 2 is based on a manuscript by M.M. Kibbe and A.M. Leslie, submitted for publication.
information other than location, unless the featural information pertained to an action relevant to the object. Perhaps the integration of brain systems underlying “what” with “where/how” information proceeds relatively slowly in infants (Káldy & Sigala, 2004).

Tremoulet, Leslie, and Hall (2000) found that 12-month-olds used both shape and color information to individuate a pair of objects that were revealed singly and sequentially. However, infants retained only shape, but not color, information so that later they were blind to an unexpected color change. These authors concluded that information used to initiate an object representation is not necessarily bound to that representation.

When identifying information is not incorporated into the object representation, young infants may track the ‘where’ of multiple objects but not be able to recall the ‘what’ for those objects (for example, what their colors are).

As noted, Káldy and Leslie (2005) found that six-month-olds can recall only one of two featurally specified objects. If this is so, what happens when infants retain the identity of only one of the objects in a pair they are tracking and not the identity of the other? Is the forgotten object entirely forgotten thereby, or does some inkling of its presence remain in memory? There are at least two broad possibilities. The first is that presence of the individual is forgotten along with its identity. On this account, six-month-olds will not be surprised if, when the screen is removed, the unremembered object has completely disappeared. The second is that six-month-old memory span is actually larger than one. Six-month-olds may retain the identity of only one hidden object per scene, but they maintain a representation of the second individual without storing its uniquely identifying features. On this account, despite not remembering its shape (or color),
infants will nevertheless be surprised if the unremembered object has completely disappeared.

We asked whether six-month-olds remember the existence of a hidden object even if they cannot remember its specific identity. We used the looking time task devised by Káldy and Leslie (2003, 2005), which requires infants to keep track of both spatiotemporal and featural properties of objects, that is, not just where, but what. This means that infants not only have to individuate the objects (which they can do by location or by feature), but bind the specific identity of each object to its location in space and maintain those bindings over time. Since we are primarily interested in how and whether these bindings are maintained, we chose to continue the use of perceptual simple visual stimuli, after Káldy and Leslie (2005). This allowed us to reduce the influence of higher-level cognitive processes that may help infants to encode or maintain object identities in other ways (e.g. chunking; see Feigenson & Halberda, 2008). Objects are hidden one at a time in distinct locations, each behind its own screen. Then, one of the screens is raised to reveal either the object that had been hidden there originally, the object hidden in the other location, or no object at all. The key advantage to this method is that it allows us to test infants’ memory for any object in a multiple object array. Specifically, we can test infants’ memory for the object that was hidden last (which is easier to recall) or the object that was hidden first (which is harder to recall) (see Leslie & Káldy, 2001, 2007, for detailed discussion of this methodology and reasons for supposing that it tests infant working memory). This methodology is especially powerful in studying how identities are bound to locations, because we can test each location without interference from the
other. Because six-month-olds forget the identity of the first-hidden object, this is the screen we will remove.

**Method**

**Subjects**

Subjects were 36 healthy full-term infants (20 females) between 21.6 and 31.4 weeks of age (mean = 26 weeks, SD = 3 weeks). An additional 6 infants were excluded due to fussiness (3), experimenter error (2), and parental interference (1). Subjects were recruited from central New Jersey through phoning lists and advertisements, and received a reimbursement and a small gift. Infants were divided evenly into three groups corresponding to conditions ($n = 12$ per group).

**Design**

The experiment followed the two-screen alternating-side violation-of-expectation method of Káldy and Leslie (2003, 2005). Infants were familiarized with two objects, a disk and a triangle, placed on an empty stage. The side of placement of the two objects was alternated from trial to trial, so that each object appeared equally often in both locations. Side of first placement was counterbalanced across subjects. Following familiarization, the experimenter (E) placed two screens, one on each side of the empty stage. Continuing the alternating-side placement, E individually placed the two objects on stage, one in front of each screen, and then placed them one at a time behind their respective screens.
Once hidden, the screen in front of the *first* hidden object (harder-to-remember) was removed revealing one of three possible outcomes: the object that originally had been hidden there (*control* condition); the object that had been hidden behind the other screen (*swap* condition); no object (*vanish* condition), see Figure 7.

![Figure 7](image)

*Figure 7.* Objects were hidden one at a time. One of the screens was removed to reveal the expected object that was hidden *first* (harder-to-recall object) (*control* condition); the object that had been hidden behind the other screen (*shape swap* condition); or no object (*vanish* condition). All groups viewed the same familiarization trials followed by one of the three outcome conditions for the test trials.

*Apparatus*

Infants sat on their parent’s lap at a distance of 91.5 cm from a 95 x 48 x 56 cm white foam-core stage. Stimuli consisted of two wooden shapes, a disk (diameter = 10.15 cm) and a triangle (base = 10.15 cm, height = 11.4 cm), both painted red. The shapes were placed on the front of the stage at a distance of about 99 cm from the infant (6 deg visual angle), then moved to the back of the stage at a distance of about 138 cm from the infant (4 deg visual angle). During test trials, shapes were hidden behind two dark gray
foam-core screens (17.75 x 17.75 cm) placed about 133 cm from the infant. Between trials, E raised a yellow curtain to cover the stage.

Procedure

E first drew the infant’s attention to the front and back corners and middle of the stage by jingling bells she wore around her wrist. A hidden observer, watching the infant’s face on a monitor, used this to get a sense of each infant’s eye positioning relative to the stage.

The experiment then proceeded with familiarization and test phases. Throughout, E timed her movements to the beat of a metronome ticking every second. In four familiarization trials the two objects were placed one at a time on the front of the stage; after 4 seconds, the objects were moved individually to the back of the stage in the order in which they appeared initially. After 8 seconds, E raised a curtain covering the viewing area and ending the trial.

Before the test phase, E asked the parent to close his or her eyes. There were four test trials; each began with two screens being placed toward the back of the stage. The two objects were then placed one at a time on the front of the stage in alternate positions to the previous trial. After 4 seconds, E hid the objects one at a time behind their respective screens. E then drew the infants’ attention to the screen which occluded the object that was hidden first (the harder-to-remember location) by jingling bells around her wrist. She then raised that screen to reveal one of the three outcomes (Figure 7).

Infants were videotaped with a camera giving a head-and-shoulders view; looking time was measured by an observer blind to condition. When E raised the screen, she
signaled to the observer to begin timing by saying “now”. The observer then timed how long infants looked at the stage by holding a button down whenever the infant looked at the stage area and releasing the button whenever the infant looked away. A computer running custom software recorded all timings. When the infant looked away for 2 seconds, the stage lights went off automatically and E raised the curtain ending the trial. Two additional blind observers later verified looking times by rescoring from videotape. Reliability, calculated as ±1 second per trial, was greater than 90% across trials.

**Results**

One test trial was excluded due to experimenter error, thus analyses were based on 143 trials total. Raw data were right skewed and non-normal and were log transformed for all analyses. We analyzed log looking times in a Condition (3) x Trial (4) repeated measures ANOVA. There was no effect of Trial (F 3, 93 <1.0) and no Trial x Condition interaction (F 6, 93 = 1.24, p = .29, η² = .074). A significant effect of Condition was found (F 1, 31 = 3.54, p = .041, η² = .186). Further analyses therefore dropped Trials as a factor. Figure 8 shows raw and log transformed looking times averaged over test trials (left top and bottom panels, respectively) and first test trial (right top and bottom panels, respectively) for each condition.
Averaged log looking times were examined by comparing Vanish and Swap conditions against Control using Dunnett’s t for multiple comparisons against a single control: Vanish log looking times were significantly longer (p = .021, two-tailed), Swap were not (p = .39, two-tailed), suggesting that infants expected an object to be revealed but were unable to recall the shape of that object.

Typically for infant looking time data, the largest effects occurred on the first test trial (Kaldy & Leslie, 2005). ANOVA on log first test trial looking times showed a significant effect of Conditions ($F_{2,33} = 4.076$, $p = .026$, $\eta^2 = .198$). Dunnett’s t comparing Swap against the Control group showed no significant log looking time difference ($p = .67$, two-tailed), replicating Káldy and Leslie (2005). Planned comparison of Vanish versus Swap showed longer log looking times for the vanish outcome ($t_{22} =$
2.256, \( p = .034 \), two-tailed). These results were confirmed non-parametrically (Vanish vs Swap: Mann-Whitney \( U = 36.5 \), \( z = -2.05 \), \( p = 0.04 \), two-tailed; Swap vs Control (\( U = 58 \), \( z = -0.808 \), \( p = \text{n.s.} \)).

Bayes Factor analyses.

Figure 9 visualizes our looking time data. The plots show the theoretical distributions of the data obtained by estimating the maximally likely mean and standard deviation given the data (Matlab function \textit{mle}). These distributions are plotted as cumulative Gaussian distributions in Figure 9, along with the data itself.

We used these theoretical distributions to quantify the likelihood that the same or different processes generated the data. Standard significance-testing statistics allows us only to reject the null hypothesis at a given confidence level; but it does not allow us to infer how likely it is that the null is actually true (e.g., Hays, 1994). However, we are just as interested in whether infants actually forgot the shape of the object as we are in whether they nevertheless remembered its existence. Showing forgetting requires us to “prove the null” of no difference in looking between the Swap and Control groups. Further, it is useful to show that the difference between the Vanish group and the Swap and Control groups is robust. Recent developments in Bayesian decision science make it easy to assess the extent to which available data favors either the null or the experimental hypothesis (Gallistel, 2009; Rouder, Speckman, Sun, Morey & Iverson, 2009). Gallistel provides an online application to compute Bayes Factor by comparing the theoretical distribution of control group data to that of an experimental group and obtaining the odds
that the same process generated both sets of data
(http://cognitivegenetic.rutgers.edu/ptn/ptn_online/bf2.aspx).

Using this method to compare log average looking times in the Control and Swap
groups yields odds 2.44:1 in favor of the null hypothesis. Because Káldy and Leslie
(2005) reported the same comparison, we combined the data from the two studies and
recomputed Bayes Factor. This analysis yielded support for the null hypothesis with
odds 5.45:1 in favor of the null. In everyday terms, currently available data makes it over
5 times more likely that the infants forgot the shape of the first-hidden object than the
alternative hypothesis that they noticed it had changed. By contrast, collapsing Swap and
Control groups in the present study and comparing them with Vanish finds odds 7.4:1 in
favor of the alternative hypothesis (H₁), namely, that infants in the Vanish group noticed
that the object had disappeared. Combining Swap + Control data from the current study
with that from Kaldy and Leslie (2005) and comparing this to the Vanish condition
provides substantial evidence, with odds of 31.2:1, in favor of H₁.
Figure 9. Each plot shows both cumulative distribution functions for both the data itself (crosses) and the theoretical distribution of the data obtained by estimating the maximally likely parameters given the data. Panel a shows the Control and Swap conditions. In panel b, data from Control and Swap conditions were collapsed. In panel c, Control and Swap data from the current study was combined with data from Kaldy & Leslie (2005).
Discussion

What do infants remember when they “forget” an object? We found six-month-olds were unable to recall the shape of the harder-to-remember of two objects, but were surprised when that object disappeared completely. The results of the current study converge with previous results showing that infants can keep track of multiple objects in homogenous sets (see Chen & Leslie, 2009; Feigenson, 2007) but can keep track of only one object identity in a heterogeneous set of two objects (Káldy & Leslie, 2005). Further, the results suggest that infants are able to maintain representations of at least one individual at each location, but that the more recently formed representation contains information about the identity of the object, whereas the "older" representation does not. Why do we see this difference between the two object representations?

One possibility is that there is rapid decay of information held in WM. Because the objects are hidden one at a time, a period of about 4 seconds elapses between when infants see the first object hidden and when their memory for that location is tested. If the information held in WM decays rapidly, it is possible that this interval is too long for infants to maintain an informative representation of the identity of the object. However, Káldy and Leslie (2005) found that infants could maintain WM representations of the easier-to-remember object over longer delays matching the length of time between hiding and revealing the harder-to-remember object. Other studies have demonstrated young infants’ capacity to remember objects over even longer delays (e.g. Baillargeon, DeVos, & Graber, 1989; Baillargeon & Graber, 1988; Luo, Baillargeon, Brueckner, & Munakata, 2003; Rose, Feldman, & Jankowski, 2001).
Another possibility is that infants’ working memory has limited resolution, such that only the object that was seen most recently is recalled with enough fidelity to notice a shape change (Eng, Chen, & Jiang, 2005; Awh, Barton, & Vogel, 2007; Zhang & Luck, 2008; see Xu & Chun, 2006 for evidence from fMRI). Under this possibility, both objects are represented in WM, but only one is represented with sufficient detail. If infants have only a fuzzy representation of the harder-to-recall object, then the shape change might not be salient enough for them to discriminate the shape-memory signal from the revealed shape.

However, shape has been shown to be a highly salient object feature. For example, Wilcox (1999) showed that 4.5-month-old infants use shape, but not color or pattern, to individuate objects, while older infants are able to use both these features (see Wilcox, Haslup, & Boas, 2010, for neuroimaging evidence using NIRS). Wilcox suggested that the use of shape to reason about object identity occurs much earlier in development than the use of other identifying features, and indicates a bias to selectively attend to this feature. In the current study, infants had a further incentive to attend to this feature because the objects were uniquely identified only by shape.

Further, infants detect even subtle changes in shape. Káldy and Blaser (2009) used a preferential first-looking paradigm with 6.5-month-olds to assess the relative visual salience of shape-, color-, and luminance-defined stimuli. The value of a feature (e.g. hue saturation) is systematically varied and presented on the same screen as an unchanging baseline stimulus. Adapting the method of Teller (1979; see Káldy, Blaser & Leslie, 2006), the stimulus that attracts infants’ first look is determined. Infants require a much smaller change in shape to produce preferential first looking compared to color-
and luminance-defined stimuli. Even small changes in shape (such as adding a notch in the contour of a polygon) are, for young infants, highly noticeable. If the large shape change in the current study (disk to triangle) was insufficiently salient, it suggests such a severe limitation to the resolution of WM for the harder-to-recall object as to be essentially nonfunctional.

Still another possibility is that infants are using object indexes to keep track of individual objects as they move in and out of occlusion, but that their ability to bind identities to those indexes is still immature (Leslie, Xu, Tremoulet, & Scholl, 1998). During the interval between the first and the last object hiding, infants’ attention was drawn away to a different location/hiding. The ability to deploy and sustain attention, and to inhibit shifts in attention, develops over the first year of life (Colombo, 2003). Maintaining feature bindings across multiple occluded locations requires representations sustained by attention, and the immaturity of underlying neural systems, probably in the ventral stream (Káldy and Sigala, 2004; Kaufman, Csibra & Johnson, 2003), means that successive hidings produce loss of existing bindings but do not actually overwrite existing storage, as suggested by Káldy and Leslie (2005).

Future work should attempt to distinguish these possibilities. Although shape could be the most basic feature used to identify objects, it is possible that there is an even more basic property. The stimuli used in the current study were distinct in shape, but they were indistinguishable in topology. Indeed, adults who were shown two shapes over extremely brief intervals (5 ms) failed to discriminate triangles and circles (which share topology), but succeeded at discriminating circles and donuts (which differ in topology) (Chen, 1982). If the resolution of WM limits infants’ performance in our task, a
topological change might be discriminable enough to evade these limitations. On the other hand, if infants fail to detect even a topological change, then the evidence for a featureless object representation will be compelling.

In the next chapter, the nature of WM limitations and development is explored in a series of studies with 9- and 12-month-olds.
III. **Tracking what went where across multiple locations**

**Introduction**

In Chapter 2, I showed that 6-month-old infants cannot recall the shape of the harder-to-recall of two objects, confirming the results of Kaldy and Leslie (2005). I also extended those results by showing that, even though infants fail to recall the *identity* of the object, they do remember the *individual*. These results converge nicely with findings that show that 6-month-olds have a limited memory span of about one object (Kaldy & Leslie, 2005; Ross-Sheehy, Oakes, & Luck, 2003), but are able to keep track of multiple objects that are out of view, even adding to and subtracting from their memory for the contents of a location (e.g. Wynn, 1992). Puzzlingly, infants’ WM for individuated objects does not develop between 6 and 12 months, while their WM for identified objects appears to develop from one at 6 months, to two at 9 months, to three at 12 months (Leslie & Kaldy, 2007). The results presented in Chapter 2 suggest distinct but interacting limits on 6-month-olds WM in which individuated objects can be tracked relatively cheaply while maintaining identifying information is costlier. In this chapter, I explore these limits in 9- and 12-month-old infants.

By 9 months of age, infants have been shown to recall the identities of two objects (Kaldy & Leslie, 2003; Ross-Sheehy, Oakes, & Luck, 2003). In Kaldy and Leslie’s (2003) study, infants’ WM was tested for the harder-to-recall shape by showing them either the expected outcome (the original, harder-to-recall shape) or the unexpected outcome of the shape that was hidden last, and was thus easier to recall. This aspect of the experiment introduces a confound that makes it difficult to conclude definitively that
9-month-olds are able to recall the shapes of both objects. It is possible that infants’ longer looking at the unexpected outcome is driven not by an identity change to the object at the harder-to-recall location, but by a location change of the object at the easier-to-recall location. Thus, infants do not necessarily have to remember the specific shape of the harder-to-recall object. All they have to do is individuate the two objects, and successfully bind shape to location of only the last object they saw. In order to gain a robust demonstration of how bindings are maintained in infant WM, this possibility needs to be investigated.

In the current chapter, I will present a series of experiments that seek to systematically examine the capacity and constraints on 9- and 12-month-old infants’ WM for features and locations. In Experiment 1, to investigate whether 9-month-old infants can recall the specific identity of the harder-to-recall object, we designed a modification of the violation-of-expectation (VOE) method which controls for the confound described above. A third shape and location for that shape are added, and infants’ memory is tested for by swapping the object of interest for the object that was hidden first rather than the object that was hidden last. Both Experiment 1 and Kaldy and Leslie (2003) test infants’ memory for the penultimate shape, but in Experiment 1 infants can no longer rely on their memory representation for the easier-to-recall last-seen object in order to succeed at the task. As a preview, infants no longer show increased looking to the shape change under these new conditions. Experiments 2 and 3 then seek to explore what limits 9-

---

N.B. None of the experiments described in this chapter explicitly test infants’ memory for the easier-to-recall, last-seen object. However, it is reasonable to assume, based on previous work, that if infants are able to recall the shape of an object hidden first (or second), they are able to also recall the shape of the object hidden last because it is easier to remember (Kaldy & Leslie, 2003; 2007, Leslie & Chen, 2007). Therefore, discussions of the results of the experiments in this chapter will make this assumption whenever applicable.
month-olds’ ability to keep track of what went where. Experiment 4 then tests whether 12-month-olds can succeed where 9-month-olds fail.

**Experiment 1: Three locations, difficult swap**

**Method**

**Subjects**

Subjects were 23 healthy full-term infants (13 females) between 34.2 and 43.5 weeks of age (mean = 38.5 weeks, SD = 3 weeks). An additional 3 infants were excluded due to fussiness (2) or experimenter error (1). Subjects were recruited from local towns around Rutgers University through phoning lists and advertisements, and received a reimbursement and a small gift. Infants were divided into two groups, where each group saw one of two outcomes detailed below (between-subjects design).

**Design**

The design was similar to that of the experiment described in Chapter 2, except a third shape and a third screen (hiding location) for that shape were added. Infants were familiarized with three objects, a disk, a triangle, and a square, placed on an empty stage. Just as in the experiment described in Chapter 2, the location of placement of the three objects was alternated from trial to trial, so that each object appeared equally often in each location. Objects were presented either right-to-left or left-to-right, alternating on each trial. Following familiarization, three screens were placed on the empty stage.
Continuing the alternating-side placement, the three objects were placed on the stage, one in front of each screen, then placed one at a time behind their respective screens.

Once hidden, the screen in front of the second hidden object was removed to reveal one of two possible outcomes: the object that had been hidden there originally (control condition); or the object that had been hidden first (difficult swap condition), see Figure 7.

Apparatus

Stimuli and setting were the same as the experiment described in Chapter 2, except a third shape (a red wooden square, 9.5 x 9.5 cm) and a third location for the shape (an additional dark gray foam-core screen, 17.75 x 17.75 cm) were added.

Procedure

The experiment proceeded similarly to the experiment described in Chapter 2. In four familiarization trials, the three shapes were placed one at a time on the front of the stage; after 4 seconds, the shapes were moved one at a time to the back of the stage in the order in which they appeared initially. Shapes always appeared either right-to-left or left-to-right, and the location of each shape was alternated across trials so that each shape appeared equally often in each position. Infants were allowed to view the shapes for 8 seconds, after which the experimenter raised a curtain covering the viewing area and ended the trial.

Each of the four total test trials began with three screens being placed toward the back of the empty stage, one directly in the center and the other two on either side. The
three shapes were then placed one at a time on the front of the stage in alternate positions to the previous trial. After 4 seconds the shapes were hidden one at a time each behind their respective screens, either right-to-left or left-to-right. The experimenter then drew the infants’ attention to the center screen by jingling bells around her wrist. She then raised that screen to reveal one of the two outcomes (Figure 10).

Looking time was measured by an observer blind to condition. These times were later verified by two additional blind observers who rescored the infants’ looking time after the experiment (reliability, calculated as ±1 second per trial, was above 90%).

**Figure 10.** Trial sequence for Experiments 1, 3, & 4, and their respective outcomes. Shapes were always hidden right-to-left or left-to-right, and infants’ memory for the middle location was probed.

**Results**
Analyses were conducted on 92 total test trials. Since the raw looking time scores were right-skewed, these scores were log-scaled to correct for the skew. All analyses presented were conducted on the log-transformed data (see Figure 11 for raw and log-scaled mean looking times).

A Trial (4) X Condition (2) repeated measures ANOVA showed no within-subjects effect of Trial ($F_{3,63} < 1, p = 0.85, \eta^2 = 0.014$) and no Trial X Condition interaction ($F_{3,63} < 1, p = 0.43, \eta^2 = 0.042$). There was also no main effect of Condition ($F_{1,21} < 1, p = 0.74, \eta^2 = 0.005$). Comparisons of looking times on the first trial only, where effects are often the strongest (see Chapter 2), also revealed no significant effect of condition; infants who saw the difficult swap outcome did not look longer than infants who saw the control outcome ($t_{21} = 0.44, p = 0.33$, one-tailed). Figure 12 shows the theoretical cumulative distribution functions (CDFs) of the data (see Chapter 2).
Figure 11. Nine-month-old infants’ raw looking time scores (top panel) and log scaled scores (bottom panel) averaged across trials for Experiment 1. Error bars show +/- 1 SEM.
Experiment 1, it would be useful to also find support for the null hypothesis, rather than to simply fail to reject the null. In Chapter 2, we outlined the usefulness of Bayes Factor analysis in providing evidence for or against the null hypothesis. Bayes Factor analysis performed on the log transformed data from Experiment 1 revealed odds that favor the null hypothesis over the alternative (BF = 3.57).

Figure 12. Cumulative distribution functions (CDFs) for Experiment 1 for both the data itself (crosses) and the theoretical distribution of the data obtained by estimating the maximally likely parameters given the data.
Discussion

When sampling confounds were controlled, that is, when infants could not use their memory for the easier-to-recall object to notice a change in the object of interest, they no longer noticed the shape change. Why do infants fail to recall the shape of the penultimate object in the current study, while succeeding on the first of two objects in the previous study by Kaldy and Leslie (2003)?

One possibility is that infants are not actually able to recall the shape of the first of two objects. Infants in Kaldy and Leslie’s (2003) study may have simply noticed a location change of the shape they saw last, and not a shape change of the specific object of interest. Indeed, previous research has shown that infants can individuate objects by feature without necessarily binding that featural information to location (Chapter 2; Tremoulet, Leslie, & Hall, 2000). Under this possibility, 9-month-olds are indeed outperforming 6-month-olds, in that they detect a change that 6-month-olds do not (Chapter 2). But, 9-month-olds WM capacity is even more limited than was previously thought; they may be able to keep track of two objects in two locations, but they are only able to recall the specific featural identity of one of those objects. Under this possibility, there would actually be little development between 6 and 9 months.

A more optimistic possibility is that the demands of the three-location-difficult-swap VOE task may have been too high for 9-month-olds. In designing Experiment 1, two changes were made from the original Kaldy and Leslie (2003) design. Both an additional location and an additional shape were added. It is possible that infants in Experiment 1 failed to notice a shape change because keeping track of shapes across three locations infringed on their limited attentional resources, preventing them from
maintaining feature bindings across locations. It is crucial to note that infants have no way of knowing which location they will be tested on, and thus have no incentive to allocate resources effectively to only a subset of items in the array (see Chapter 4 for a novel method for “instructing” infants). Increasing the number of locations (and indeed, the number of objects) infants have to track may exceed their attentional or WM limits to such an extent that they are unable to recall items in the array that they previously had no trouble with. Similar performance was seen with 10-12-month-old infants on manual search and foraging tasks where infants had to track homogenous sets; if the size of the set exceeded about 3 items, infants completely failed to search for hidden items (Feigenson, Carey, & Hauser, 2002; Feigenson & Carey, 2003; 2005; see Zosh & Feigenson, 2009 for a review). Under this possibility, 9-month-olds may have the capacity to recall two shapes, but not when the attentional demands of the task are too high.

In Experiments 2 and 3, I test these possibilities by manipulating the demands of the task. In Experiment 2, I reduce the number of screened locations infants have to track to two, but use three objects presented individually but two at a time to test infants’ memory for the penultimate shape with a difficult swap. If 9-month-old infants are able to recall the specific shape of the object, but were hampered by the increased number of locations in Experiment 1, they should once again look longer at the shape change even when they cannot simply rely on their memory for the easier-to-recall, last-hidden shape.

**Experiment 2: Two locations, difficult swap**

*Subjects*
Subjects were 19 healthy full-term infants (8 females) between 34.5 and 42.7 weeks of age (mean = 38.2 weeks, SD = 2.4 weeks). An additional 2 infants were excluded due to fussiness (1), and equipment malfunction (1). Infants were divided into two groups, where each group saw one of two outcomes detailed below.

**Design**

Infants were familiarized to all three shapes (a disk, a triangle, and a square), but only two at a time. During familiarization trials, infants watched two shapes from the set of three total shapes placed sequentially on the front of the stage (e.g., disk and triangle, triangle and square, disk and square). The location of placement of the two shapes was alternated from trial to trial, so that each shape appeared equally often in each location.

Following familiarization, two screens were placed on the empty stage. Continuing the alternating-side placement, two shapes were placed on the stage, one in front of each screen, then placed one at a time behind their respective screens. Once hidden, the screen in front of the *first* hidden object was removed to reveal one of two possible outcomes: the shape that had been hidden there originally (*control* condition); or the shape that had not appeared on that trial, but to which infants had been familiarized (*difficult swap* condition), see Figure 13.

**Apparatus**

Stimuli and setting were the same as Experiment 1, except only two screens were used.

**Procedure**
The experiment proceeded similarly to Experiment 1. In six familiarization trials, the two shapes were placed one at a time on the front of the stage. Infants viewed the shapes for 4 seconds, after which E moved the shapes to the back of the stage in the order they were initially presented. Infants viewed the shapes for an additional 6 seconds, after which the curtain was raised over the display. Trials were counterbalanced so that each shape appeared equally often in each location and were presented both first and last in presentation order.

Each of the three total test trials began with two screens being placed toward the back of the empty stage. Two objects were then placed one at a time on the front of the stage in alternate positions to the previous trial. After 4 seconds the objects were hidden one at a time each behind their respective screens. E then drew the infants’ attention to the location of the shape that was hidden first, by jingling bells around her wrist. She then raised that screen to reveal either the shape that had been hidden there originally (control outcome) or the shape that had not appeared on that particular trial (difficult swap outcome) (Figure 13).

Looking time was measured by an observer blind to condition. These times were later verified by two additional blind observers who rescored the infants’ looking time after the experiment (reliability was above 90%).
Figure 13. Top panel shows examples of familiarization trials for Experiment 2. Bottom panel shows test trial sequence and experimental outcomes. During test, the probed object was swapped for the one shape from the set of three that was not present on that trial.
Results

Analyses were based on 55 total test trials. Two additional trials were excluded due to experimenter error. Right skew in the raw scores was corrected by log transform for all analyses (Figure 14). Log looking times were analyzed in a Condition (2) x Trial (3) repeated measures ANOVA. There was no effect Trial ($F_{2, 30} = 1.13, p = 0.34, \eta^2 = 0.067$) or and no Trial x Condition interaction ($F_{4, 30} = 1.44, p = 0.27, \eta^2 = 0.094$).

Repeated measures ANOVA revealed only a marginal effect of Condition ($F_{1, 15} = 3.13, p = 0.09, \eta^2 = 0.173$). However, planned comparisons using Student’s t showed infants in the difficult swap condition looked significantly longer than infants in the control condition ($t_{17} = 2.395, p = 0.016$, one-tailed). Although there was no main effect of Trial, as is often the case in violation-of-expectation experiments, the largest effect occurred on the first trial ($t_{16} = 2.737, p = 0.007$, one-tailed). Figure 15 shows the best-fitting Gaussian CDFs for both the averaged data and the first trial only.
Figure 14. Raw mean looking time (top panels) and log-scaled mean looking times (bottom panels) for Experiment 2. The left panels show data averaged across all four trials. The right panels show data from the first trial only. Infants looked significantly longer at the difficult swap outcome than the control outcome; this effect was strongest on the first trial. (Error bars show SEM.)

Figure 15. CDFs for Experiment 2. Panel a shows distributions for data averaged across all three test trials. Panel b shows distributions for data from the first test trial only.
Further support for the alternative hypothesis comes from Bayes Factor Analysis. The odds are in favor of the alternative for the data averaged across trials (BF = 4.95), and for the data from Trial 1 only the BF odds provide substantial evidence for the alternative (BF = 13.62).

**Discussion**

When infants only had to keep track of two locations, they could detect a change in the shape of the harder-to-recall object. Experiment 2 controlled for the confound in Kaldy and Leslie’s (2003) original design, because the change was purely an identity change to the object of interest, rather than a location change of the easier-to-recall object. The results of Experiment 2 suggest that 9-month-old infants can indeed keep track of two shapes, and can do so robustly, when they are only asked to keep track of two locations. This suggests that the demands of the three-location task are not driven by the difficult swap, but by the number of locations.

In Experiment 3, we further manipulated the demands of the three-location task by making it a bit easier. Infants’ memory for the penultimate object was once again probed, but the penultimate shape was swapped for the shape that was hidden last. In this way, Experiment 3 is nearly identical to the experiment of Kaldy and Leslie (2003) apart from the third object and screen. Infants could detect the change by noticing that the location of the easier-to-recall object had changed, or by detecting an identity change in the object of interest. By giving infants two ways to detect the change, the demands of the task are somewhat reduced. If infants detect the change, we can conclude that they can recall one shape and its location, but not two shapes. If infants do not detect the
change, this suggests that having to keep track of three locations *severely* limits their ability to track what went where.

**Experiment 3: Three locations, easy swap**

*Subjects*

Subjects were 10 healthy full-term infants (5 females) between 34.8 and 44.4 weeks of age (mean = 39.8 weeks, SD = 2.8 weeks). These infants were compared to the control group from Experiment 1. An additional 3 infants were excluded due to fussiness (1), experimenter error (1), or parental interference (1).

*Design, Apparatus, and Procedure*

The design of the current study was identical to Experiment 1, except that the test outcome was different: when the experiment removed the screen that occluded the second-hidden object, she revealed the object that had been hidden last (*easy swap* condition), see Figure 16.
Figure 16. Trial sequence for Experiment 3. The outcome for Experiment 3 is shown in the bottom right panel. Shapes were always hidden right-to-left or left-to-right, and infants’ memory for the middle location was probed.

Results

Since everything about the procedure leading up to the reveal outcome was identical to Experiment 1, infants’ looking times in Experiment 3 were compared to the looking times of the control group in Experiment 1 (for ease of comparison, data from both conditions of Experiment 1 are plotted with the data from Experiment 3 in Figure 17). Theoretical CDFs from Experiments 1 and 3 are plotted in Figure 18.

Analyses were based on 83 total test trials. One additional trial was not included in analysis because of infant fussiness on the last of the four test trials. Results of a Trial (4) x Condition (2) repeated measures ANOVA showed no effect of Trial ($F_{3,54} < 1, p = \ldots$)
0.79, $\eta^2 = 0.02$) and no Trial x Condition interaction ($F_{3,54} < 1, p = 0.7, \eta^2 = 0.013$).

There was also no main effect of Condition ($F_{1,18} < 1, p = 0.91, \eta^2 = 0.001$), suggesting infants did not look longer at the shape change outcome than the control outcome. An additional comparison of looking times on the first trial only also revealed no significant effect of condition ($t_{19} = 0.19, p = 0.42$, one-tailed).

Support for the null hypothesis comes from Bayes Factor Analysis, with odds of 3.01:1 in favor of the null hypothesis. As in Experiment 1, these odds provide some evidence to support the null, although sample sizes are somewhat low. We also collapsed across both the difficult and easy swap conditions and compared them to the control outcome, since we were interested in whether infants can detect any change. Bayes Factor Analysis yielded odds in favor of the null, 3.82:1.

Since infants in Experiments 1 and 3 could not detect a change in the probed shape when tracking objects across three locations, but could do so in Experiment 2 when tracking objects across two locations, we used the data from Experiments 1 and 3 as a control group and compared it to the swap outcome looking times from Experiment 2. We can accept the alternative hypothesis that infants detect a shape change under two locations, but not under three locations, with odds of 4.98:1.
Figure 17. Combined raw looking time (top panel) and log-scaled looking time (bottom panel) from Experiments 1 and 3.
Discussion

Infants in Experiment 3 failed to detect that the probed shape had changed, even though they were given multiple means to detect the change (either by detecting a location change of the easier-to-recall shape or by detecting a shape change to the probed object). This result is interesting, because it represents a failure to replicate the results of Kaldy and Leslie (2003) simply because a third location and third shape were present in the display.

The results of Experiments 1-3 are summarized in Figure 19. These results suggest that 9-month-olds do indeed have the WM capacity to recall two shapes (in line with previous work by Kaldy & Leslie, 2003; Ross-Sheehy, Oakes, & Luck, 2003), but
only when they have only two locations to keep track of. When infants are asked to track three locations, their capacity to recall which object went where is severely limited.

Figure 19. Schematic summary of the results of Kaldy & Leslie (2003) and Experiments 1-3. Nine-month-olds can recall two shapes when they have to track them across two locations, but not when they have to track three locations.

Is the three-location task simply too difficult, or might infants develop the capacity to keep track of shapes across three locations? Experiment 4 asks whether 12-month-old infants can detect the difficult swap in the three-location task. If 12-month-
olds succeed, it would suggest that the attention or WM resources required for the task undergo rapid development between 9 and 12 months of age.

**Experiment 4: Three locations, difficult swap (12-month-olds)**

*Subjects*

Subjects were 24 healthy full-term infants (14 females) between 47.1 and 55.6 weeks of age (mean = 51.8 weeks, SD = 3.2 weeks). An additional 6 infants were excluded due to fussiness (3), experimenter error (1), and parental interference (2). Infants were divided into two groups, where each group saw one of two possible outcomes.

*Design & Procedure*

Experiment design and procedure were identical to the difficult swap condition in Experiment 1 (Figure 10).

*Results*

Analyses were based on 94 total test trials. Two additional trials were excluded due to infant fussiness toward the end of the experiment. Right skew in the raw scores was corrected by log transform for all analyses (Figure 20). Figure 21 shows the theoretical CDFs for data averaged across all four trials, and data averaged across the first two trials only.

Log looking times were analyzed in a Condition (2) x Trial (4) repeated measures ANOVA. There was no effect Trial ($F_{3, 63} = 2.171, p = 0.125, \eta^2=0.076$) or and no Trial
x Condition interaction ($F_{6, 63} = 1.62, p = 0.218, \eta^2 = 0.091$). There was a significant main effect of Condition ($F_{1, 21} = 6.926, p = 0.02, \eta^2 = 0.248$). Unlike the 9-month-olds in Experiment 1, 12-month-olds looked significantly longer at the difficult swap outcome than the control outcome ($t_{22} = 2.53, p = 0.01$, one-tailed). This result is largely driven by infants’ looking on both the first trial ($t_{22} = 2.36, p = 0.016$, one-tailed) and the second trial ($t_{22} = 2.64, p = 0.007$, one-tailed).

Bayes factor analysis shows support for the alternative hypothesis on the averaged data (BF = 6.93), with greater support for the alternative on the first two trials combined (BF = 12.75).

Figure 20. Raw looking times (top panels) and log-scaled looking times (bottom panels) for 12-month-olds in Experiment 4. Plots on the left show data averaged across all 4 trials. Plots on the right show data averaged across the first two trials only.
Figure 21. Cumulative distribution functions for Experiment 4. Panel a shows CDFs for data averaged across all four test trials. Panel b shows CDFs for the first two trials only.
General Discussion

Two interesting patterns emerge from the results of the studies reported in this chapter. First, as the number of locations or hiding events infants have to track increases, their ability to keep track of what item went where decreases. Second, the ability to keep track of what went where increases with development. At 9 months, infants can keep track of which of two shapes went in which of two locations, but performance breaks down with three locations. By 12 months, infants are able to successfully keep track of (at least) two shapes across three locations. Previously, in Chapter 2, it was shown that 6-month-olds can only recall one object identity when asked to track two objects, suggesting a developmental trajectory in which WM capacity develops rapidly in the span of six months. General implications for the structure of and limits on WM will be discussed in Chapter 5.

Some questions still remain about the nature of the limits on infants’ WM. In the three-location task, the number of locations and number of objects are confounded on any given trial. Two additional studies are currently underway to examine whether the task demands that severely limit infants’ WM capacity are driven by the number of locations infants attend to, or by the number of objects that they need to track. In one study, two shapes are hidden in two of three total locations. Infants’ attention is drawn to the third location by the wave of a hand, but no object is hidden in that location. If infants fail to detect a shape change in this experiment, we can conclude that it is due to the demands of attending to more than two locations, even though there is nothing in that location in that test trial. In another study, three shapes will be hidden in two locations with two objects hiding in one of the locations and the third hiding in the second location. If infants fail to
detect a shape change in this study, we can conclude that it is due to set size limitations and not to spreading attention across locations. If infants fail to detect a change in both experiments, it suggests that both set size and number of locations work to limit the capacity of infant WM.

Another possibility is that the number of occlusion events infants see is working to limit their ability to keep track of what went where (Baillargeon, 2008). If this is the case, infants’ performance may be better if they are asked to track three objects hidden in different locations behind one long screen. Or, different types of hidings (such as a covering event and an occlusion event) might produce different performance. Future work would explore these possibilities as well.

In Chapters 2 and 3, we used a method in which we controlled for sampling confounds in infants’ response by revealing only one shape at a time. In this way, we can be assured that infants’ responses are based solely on the aspect of the display that we are interested in. In Chapter 4, we ask whether it is possible to cue infants to encode specific aspects of the display prior to testing.
IV. Remembering the location of an agents’ goal

Introduction

The experiments described thus far, and indeed the vast majority of experiments designed to study infants’ object concept, use tightly controlled stimuli wherein objects are presented by themselves or with other objects. These experiments have effectively mapped infants’ rather robust expectations about the physical properties of objects (Baillargeon et al., 1985; Spelke, 1990; Kellman & Spelke, 1983; Johnson & Aslin, 1995; Johnson, Cohen, Marks, & Johnson, 2003; Streri, Gentaz, Spelke, & Van de Walle, 2004; Baillargeon, et al., 1985; Baillargeon, 1991; Sitskoorn & Smitsman, 1995) and their extremely limited ability to keep track of objects once they are out of view (Chapters 2 and 3, this dissertation; Feigenson et al., 2002; Feigenson & Carey, 2003; Kaldy & Leslie, 2003; 2005; Leslie & Chen, 2007; Ross-Sheehy, Oakes, & Luck, 2003; Oakes, Ross-Sheehy, & Luck, 2006).

However, infants’ day to day life is populated not just by objects, but by people. Moreover, people interact with objects constantly. Indeed, outside of the laboratory, objects rarely lack a purpose or use. The infant herself reaches for a ball because she wants to play with the ball. The properties of that physical object (roundness, blueness, ballness) become important to her only in the context of her desire for the object. Thus, the intentions of human agents can highlight perceptual features that may otherwise have seemed irrelevant or not worth remembering. And, since infants are severely limited in their capacity to represent and recall objects and their features (Chapters 2 and 3), this
property of agent-object interactions could play a critical role in what information gets encoded in WM.

**Agents and objects**

Recent work has shed light on what infants understand about how agents and objects interact. Infants, like adults, tend to focus on those aspects of an agent’s actions that can explain the agent’s intentions. In one of the first studies of its kind, Woodward (1998) tested infants’ ability to understand goal-directed behavior using a habituation paradigm. During a habituation phase, 9-month-old infants watched as a hand reached for and grasped one of two toys, always preferring one to the other. During the test phase, the toys were swapped and the agent either reached for the same toy in the new location (goal-consistent, path-inconsistent outcome) or to the different toy in the old location (goal-inconsistent, path-consistent outcome). A different group of infants saw similar actions performed by a rod with a hand-like protrusion on the end in both habituation and test phases, except that the rod simply touched the toy rather than grasping it.

Woodward found that, when a hand was performing the action, infants were surprised when the hand reached for a new toy; when a rod was performing the action, infants looked equally long at both conditions. Five-month-olds exhibited similar tendencies, though not to the same extent that the 9-month-olds did; while 5-month-old infants encoded hand activity and rod activity differently, they were slower to recover from habituation and their looking behavior during test trials was only marginally reliable. By 6 months of age, infants in Woodward’s studies reliably attended to the toy when a hand was doing the grasping, but not when an inanimate claw was doing the
grasping, showing that it is not the act of grasping, but the fact that an agent is doing the grasping, that prompts infants to selectively encode specific aspects of the scene that might explain the human agent’s actions. Both 5- and 9-month-old infants attributed a goal to an agent only when the agent’s actions were purposeful (grasping an object) rather than accidental (letting the hand fall on an object with no intentions toward that object) (Woodward, 1999). These results suggest that infants as young as five months can recognize an agent and attribute a goal to an agent based on the physical properties of the objects on which the agent acts.

Infants can also use their understanding of an agent’s knowledge about the world to reason about the agent’s desires. Tomasello and Haberl (2003) demonstrated this in a cooperative playing scenario. Infants played with two adults and two toys, at which point one of the adults left the room and a new toy was introduced. When the adult came back into the room, she expressed excitement when looking at all three toys and asked the infant, “Can you give it to me?” Both 12- and 18-month-old infants correctly gave her the new toy, demonstrating not only that they could identify which toy was novel to the adult even though they, the infants, were already familiar with the toy, but also that they reasoned that her excitement and desire for a toy was indicative of her desire for the novel object.

Infants in the second year of life have also been shown to attribute false beliefs to agents. Onishi and Baillargeon (2005) adapted the classic Sally-Anne task (Baron-Cohen, Leslie, & Frith, 1985) for 15-month-old infants. Infants were familiarized with an agent placing a toy into a box. They then saw one of two possible actions: the agent looked away and the toy moved from one box to another, or the agent watched the toy
move to another box. Infants then saw the agent either reach for the toy where she
*thought* it was (acting on her false belief of the state of the world), or reach for the toy
where it actually was (acting on the state of the world). Infants were surprised when the
agent reached into the box where the toy actually was, suggesting that infants expected
the agent to act on her (false) representation of the world.

Even younger infants are able to keep track of an agent’s representations of
objects in a scene and take that into account when interpreting the agent’s actions. Luo
and Baillargeon (2007) tested whether 12.5-month-old infants would attribute to an agent
the goal of acting on a particular object when the agent could not see the object. During a
Woodward-style familiarization period, infants saw an agent repeatedly reach for a
cylinder as opposed to block. In one condition, the block was placed in front of a
transparent screen such that it was visible to the agent; in another, the block was hidden
from the agent behind an opaque screen but was visible to the infant. During a test
period, the locations of the objects were reversed and the screens were removed. Infants
saw the agent reach for either the cylinder (goal-consistent, motion-path-inconsistent
action) or the block (goal-inconsistent, motion-path-consistent action). Luo and
Baillargeon found that infants were surprised when the agent reached for the block *only*
when the block had been visible to the agent during the familiarization phase of the
experiment. When the agent could not see the block, infants did not interpret the agent’s
reaching for the cylinder during the familiarization trials as a preference for the cylinder.
This suggests that infants interpret the agents’ reaching and grasping behavior as
signifying a preference for a particular object over another only when the agent can make
a choice, based on the identifying properties of the objects, of which object to grasp
based on her representation of the scene, not simply when two objects are present in the scene.

In a follow-up experiment, infants watched the agent move the block behind an opaque screen, such that it was hidden from the agent but visible to the infant. The agent then proceeded to reach for and grasp the cylinder, as in the experiment described above, with the block hidden from her view. During the test trials, the screens were removed and the location of the objects reversed. Infants looked reliably longer when the agent reached for the block as opposed to the cylinder, suggesting they were surprised when the agent acted on a new goal object. Recall that when the block was hidden from the agent’s view, and the agent did not have any knowledge of the block’s presence, infants did not attribute a preference for a particular object to the agent. However, infants reliably attribute that preference not only when the block is perceptually available to the agent throughout the trials, but also when the agent knows the object is there. This is interesting, because it more strongly suggests that infants are indeed basing goal attribution judgments on the contents of the agent’s mental representations and not simply on what is perceptually available to the agent.

Luo and Baillargeon’s findings show that from at least 12.5 months, infants can not only keep track of the contents of a person’s mental representation of a scene but also use the person’s mental representation to explain their actions. Similar results were found with 13-month-old infants when non-human agents were used. In a similar paradigm, Surian, Caldi, and Sperber (2007) familiarized infants with a caterpillar who watched two food items hidden one at a time, each behind one of two barriers. The food items were placed in the same location in each familiarization trial, and in each trial the
caterpillar then crawled behind the barrier that occluded the food item it “preferred”.

During test trials, before the caterpillar re-entered the scene, the food items were placed either behind barriers that were lying flat (so that the food was visible) or behind barriers which occluded them from the caterpillar, and their locations were swapped. When the caterpillar reentered, it either crawled to where its preferred food actually was or to where it was during the familiarization trials.

In order to properly predict where the caterpillar should go, infants needed to keep track of the caterpillar’s goal as well as its belief about the state of the world. In the case where the caterpillar can see the food, the action of crawling to where the food previously was would be a goal-inconsistent action. However, in the case where both items are occluded, crawling to the old location would be consistent with the caterpillar’s knowledge of the situation and its beliefs about where the food is located. Infants were surprised when the caterpillar acted in a way that was inconsistent with what its knowledge about the world should be given what it could see, not when it acted in a way that was inconsistent with the actual state of the world. Further, when the caterpillar was present when the objects were hidden, infants were surprised when the caterpillar crawled to the wrong location given its goal; they expected it to act on its knowledge of the locations of the two items. Surian, et al. note that these results show that infants expect agents to act on the information they have about the world, and not the information that the infant has about the world.

In each of the studies described above, infants are using their knowledge of the agent’s representation of the world in order to predict the agent’s action. Woodward argues that the tendency to want to explain an agent’s actions results in infants selectively
attending to aspects of the situation that will be the most informative for that goal. This idea of “selectively attending” to only the task-relevant aspects of the scenario is implied in studies of early language learning. Woodward notes that in Tomasello and Barton (1994), 2-3-year olds attributed the meaning of a novel verb to a purposeful action rather than an accidental one, suggesting that they are disregarding information that is unlikely to provide any insights about the novel verb. Thus, explaining action in terms of agency can constrain the amount of information an infant must take in and encode to only that which is most relevant for learning and for future use. Agency communicates information.

Indeed, infants are sensitive to what agentive communication can tell them about objects in the world. By at least 8 months of age, infants can infer the presence of a hidden object when they observe an actor directing her gaze toward the location of the object (Csibra & Volein, 2008; Luo, 2010). However, it is not until between 9 and 12 months of age when infants infer a relationship between the object of an agent’s gaze and the agent’s preference for that object. Woodward (2003) showed that 6- and 8-month-old infants understood an agent’s preference for an object toward which she directed her gaze only when she also grasped the object; by 12 months, infants needed only directed gaze to infer a preference.

Yoon, Johnson, and Csibra (2008) asked whether 9-month-old babies selectively recall information about an object depending on the kind of communication cues they get from an adult agent. For example, adults often point at objects and name them, cueing the infant to the object’s identity, which is essential for word learning. Yoon et al. reasoned that pointing to an object provides a cue to pay attention to what the object is,
such that the *where* of an object is less important than its identity (e.g. it would be a *doll* even if it were not sitting on the chair). In their study, an object was placed on either side of a stage with an agent present. Infants watched the agent make one of two motions to an object, either a non-communicative reaching action or a pointing action typical of a communicative gesture. A curtain was then lowered over the agent and two opaque screens were placed on either side of the stage, one occluding the object. When the screens were removed, infants either saw the object change identity (from one type of toy to another) or change location (from one side of the stage to the other). Infants looked significantly longer at the identity change condition than the location change condition when the agent had *pointed* to the object. Conversely, infants were more surprised by a location change than an identity change when a non-communicative reaching motion had been made toward the object.

This result is surprising since 9-month-old infants have shown that they are able to integrate identity and location information for two objects (Chapter 3, Kaldy & Leslie, 2003). Yoon, Johnson, and Csibra (2008) conclude that this binding is tenuous, such that WM for *where* an object is and WM for *what* it is are easily decoupled (see also Chapter 2). But a different interpretation is that WM is elastic and that different properties of the object are encoded depending on the situation. Recognizing and attending to social cues could allow infants to use their limited WM in a more efficient way. Interestingly, in Yoon, et al.’s experiment, pointing cued infants to remember the objects identity while reaching cued infants to remember the object’s location. Yoon et al. sought to contrast actions that are purposefully communicative with actions that are not. In Woodward’s (1998) study, as in many other similar studies cited in the current paper, a goal-directed
reaching motion cued infants to selectively attend to the goal object and not to its location. However, none of the studies looking purely at infants’ ability to attribute goals to agents require infants to make use of WM. Experiments designed to test infants’ attribution of false beliefs (e.g. Onishi & Baillargeon, 2005; Luo & Baillargeon, 2007; Surien, Caldi, & Sperber, 2007) do require infants to rely on WM representations, but in these studies infants have privileged knowledge of the changes that are made to the display whereas the agent does not. In Yoon et al., changes are made to the objects without the infants’ knowledge, just as they are in classic studies of infant object cognition.

The current study

The current study asks whether an agent’s goal-directed behavior toward objects in the visual array will affect what information infants encode about the array. To answer this question, the experimental method of Chapters 2 and 3 (see also Kaldy & Leslie, 2003; 2005) was combined with a Woodward (1998)-style goal attribution method. Familiarization trials proceeded similarly to all the WM experiments described in this dissertation, with one important difference: an actor was present watching as the shapes were placed one at a time on the stage, and she always reached for one shape versus the other. During test, the objects were hidden from the infants’ view, but visible to the actor. This contrasts with previous work described above where the infant had privileged access to the visual array. Once hidden, the actor retrieved her goal object either from the location where it was hidden (actor’s goal- and infant’s memory-consistent outcome), or the opposite location (actor’s goal-consistent, infant’s memory-inconsistent outcome).
The actor always retrieved her shape from the *first-hidden location* (the harder-to-recall location, Chapters 2 & 3), regardless of whether it was hidden there or not.

There were two primary measures of interest in the current study. The first is whether infants can *anticipate* where an agent will search for her goal object, even though the infant cannot see the object. Anticipation is measured by whether the infant looks ahead to the goal location before the agent reaches (anticipatory eye gaze). If infants anticipate where the actor will search, it would suggest that they understood her goal and were able to predict her action without access to the perceptual properties of the objects themselves.

The second measure of interest is whether infants will look longer when the goal shape is retrieved from the wrong location (violation of expectation). Recall that 6-month-old infants were unable to recall the shape of the first-hidden object in the display (Chapter 2), but could recall the shape of the last-hidden object (Kaldy & Leslie, 2005), while 9-month-olds infants could recall both shapes (Chapter 3, Kaldy & Leslie, 2003). In the current study, both 5-7-month-old infants and 9-12-month-old infants were tested, the former having previously failed the pure object-WM version of the task, the latter having previously passed it.

**Method**

*Subjects*

Subjects were 13 healthy, full-term 5-7-month old infants (mean age 6 months 0 days, SD = 19 days) and 19 healthy, full-term 9-12-month-olds (mean age 10 months 10 days, SD = 38 days). An additional 2 infants were excluded due to fussiness. Subjects were recruited from local towns around Rutgers University through phoning lists and
advertisements, and received a monetary reimbursement and a small gift. Infants were divided into two groups, where each group saw one of two outcomes (control or swap).

**Apparatus**

Stimuli were displayed on a Tobii T-60XL 24” video eye tracker (60 Hz sampling rate). The Tobii T-60XL system does not require the subjects’ head to be stabilized, allowing for a 40x20x27 cm range of head movements. Eye movements were measured binocularly.

Tobii Studio software was used to display of the stimuli and collect eye movement data. We used the software to specify regions of interest (ROIs) in the display, which allowed us to measure infants’ gaze to specific areas of the display.

**Stimuli**

Infants, seated on a parents’ lap at a distance of 60 cm from the screen, were first shown a calibration screen in which a moving cartoon lobster, accompanied by a ringing bell to capture infants’ attention, appeared in five different points in the display.

Following successful calibration, infants were shown a video of a gloved hand placing real objects on a stage in front of an actor. The actor was wearing a light blue t-shirt and a beige visor which covered her eyes. Objects were a red wooden triangle (6.67° x 7.26° visual angle at the front of the stage; 5.71° X 6.06° at the back of the stage) and a red wooden disk (7.26° visual angle at the front of the stage; 6.06° at the back of the stage). During test, two foam-core screens (12° x 8.5° visual angle) covered in light green paper were placed on either side of the stage. Video was displayed at 800x600 pixels resolution.
Three ROIs were defined for the stimuli during the test phase: one for each screen, and one for the actor’s face.

*Design*

Infants were familiarized with two objects, a disk and a triangle, placed on an empty stage by a gloved hand. The side of placement of the two objects was alternated from trial to trial, so that each object appeared equally often in both locations. Side of first placement was counterbalanced across subjects. On each familiarization trial, an actor watched the hand place the objects on the stage, then move them toward the actor at the back of the stage. The actor then reached for and grasped either one shape or the other and moved it to the front of the stage. On *disk-goal* trials, she always reached for the disk, no matter where the disk appeared or which order the objects appeared in. On *triangle-goal* trials, she always reached for the triangle. The purpose of these trials was not just to familiarize infants with the stimuli, but to teach them that the actor always preferentially takes an action on one of the two shapes.

Following familiarization, two screens were placed on each side of the empty stage. Continuing the alternating-side placement, the two objects were then placed on the stage, one in front of each screen, then placed one at a time behind their respective screens while the actor watched the action. Once the objects are hidden from view, the scene was paused. During this time, we measured whether infants would anticipate where the actor would search for her goal shape prior to the actor taking an action. This was done by measuring infants’ first saccade to one of the ROIs.
After the anticipation interval, the actor then retrieved the goal object\textsuperscript{3} from either the goal location (\textit{goal- and memory-consistent outcome}) or the opposite location (\textit{goal-consistent, memory-inconsistent outcome}).

Procedure

Prior to the start of the experiment, the Tobii Studio 5-point infant calibration program was run. Calibration stimuli consisted of an animated image of a toy lobster combined with a sound designed to gain and hold infants’ attention. The lobster appeared at five points on the screen. If a signal from the eyes was not obtained at one or more of the points, the calibration was re-run.

The experiment then proceeded with familiarization and test phases. At the beginning of each trial, infants heard a tone and saw an actor sitting behind an empty stage with her arms folded on the back of the stage, gazing at a neutral position in the center of the stage. The tone signaled the beginning of each trial and helped orient infants’ attention back to the video after each trial ended. The actor’s eyes were hidden from the infant by a visor, so that the eyes would not capture infants’ attention. Over the course of both familiarization and test trials, the actor always oriented toward any action happening on the stage.

During the four familiarization trials (Figure 22a), a gloved hand appeared from the right side of the screen and placed two objects (disk and triangle) one at a time on the front of the stage. A “voila!”-type tinkling sound accompanied each placement of the shape. This was followed by a 4 second interval in which the actor looked at a neutral

\textsuperscript{3} Within the framework of typical WM experiments (Chapters 2 & 3), in the current task, the hand retrieved the objects from the “harder-to-recall” location. This characterization works well in purely object-based scenarios; however, in the current study, where an agent’s preference plays a role, the locations are referred to as “goal location” and “opposite location” to avoid confusion.
central point on the stage and infants were allowed to view the shapes. The objects were then moved one at a time to the back of the stage in the order in which they appeared initially. This movement was accompanied by a whooshing sound to keep infants’ attention. Then, following a phone-ringing tone, the actor reached for one of the shapes (always the same shape on each trial), grasped it, and moved it to the front of the stage. In every trial, the phone-ringing tone was always followed immediately by the actor grasping and moving the goal object. Thus, the tone can be thought of as an “anticipation prompt”, which serves to signal the infant that the actor is about to act.

Once the actor moved the goal shape to the front of the stage, the scene paused and the infants were allowed to view the scene for an additional 8 seconds. Following the 8 second interval, the screen faded to black and the next trial faded in immediately thereafter.

Two test trials immediately followed the four familiarization trials (Figure 22b). At the beginning of each of two test trials, the gloved hand placed two screens one at a time on either side of the stage. Then, just as in the familiarization trials, the hand placed the two shapes one at a time at the front of the stage in alternate positions to the previous trial, accompanied by the “voila!” tinkling sound. After a 4 second interval where the actor was in a neutral position, the hand placed the objects one at a time (accompanied by the whooshing sound) each behind their respective screens. Once the objects were hidden, the actor returned to a neutral position and the anticipation prompt (1 second in duration) was played on the video’s soundtrack. Following the prompt, there was an additional 3 second interval, in which the actor took no action. During this 4 second anticipation period, infants’ eye movements were measured. Then the actor reached
toward either the “expected” goal location, or to the “unexpected” opposite location, and removed her goal object, placing it on the stage in front of the screen. After 15 seconds, the screen faded to black.

Eye movements were recorded by Tobii Studio software during the entire experiment. In addition to the data gathered by Tobii, infants were videotaped with a camera giving a head-and-shoulders view. Looking time was also recorded offline by a human observer who was blind to the condition, using a computer running custom software. Looking times were later verified by two additional blind observers who rescored the infants’ looking time after the experiment (reliability was above 90%).
Infants saw four familiarization and two test trials, with order and placement of shapes counterbalanced.

**Results**

*Anticipatory looking*

The target of infants’ first saccade immediately following the anticipation tone was measured. A first saccade was considered “anticipatory” if it landed on one of the two screens. Infants did not make anticipatory first looks on every trial. Five-7-month olds made anticipatory first looks on 16/28 trials (57%). On the remaining 12 trials, infants’ gaze stayed directed at the visored face (10) or the first saccade was directed to
somewhere else in the scene (2). Nine-12-month-olds made anticipatory first looks on 22/36 trials (61%). On the remaining 14 trials, infants’ gaze stayed on the face (8) or was directed elsewhere (6) (see Figure 23a).

Infants’ looking patterns were analyzed for both test trials. Looks to the face or to elsewhere in the display were not considered anticipatory looks and were not used in calculating infants’ success or failure at anticipating the agent’s goal. If infants’ first look was to the opposite location on one or both test trials, they were considered to have failed to anticipate the agent’s action. If infants’ first look was to the goal location on both trials, or to the goal location on one trial and not to the opposite location on the other trial, they were considered to have successfully anticipated the agent’s action.

Infants in the 5-7-month age range showed correct anticipation of the agent’s actions. Nine out of the 13 infants anticipated correctly on one trial and looked to the face on the other trial. One infant anticipated correctly on one trial and looked elsewhere in the scene on the other trial. These infants were considered to have successfully anticipated the agent’s action. The remaining three infants looked to the same location on both trials, resulting in one “correct” and one “incorrect” response, and thus were considered to have failed to anticipate correctly. These data amount to 10/13 infants successfully anticipating the agent’s action, binomial test p= 0.046 (one-tailed; Figure 23b).

Performance of infants in the 9-12-month age range, on the other hand, was much poorer. One infant out of 19 anticipated correctly on both trials. Two anticipated correctly on one trial and looked at the face on the other. Seven infants anticipated correctly on one trial, and looked elsewhere in the scene on the other trial. These infants
were considered to have successfully anticipated the agent’s action. Two infants looked at the face on both trials. The remaining infants failed to anticipate correctly on one (n = 6) or both (n = 1) trials. These data amount to 10/17 infants successfully anticipating the agent’s action, binomial test p = 0.629 (Figure 23b).

![All Looks](image_url1)

![Anticipatory First Looks](image_url2)

**Figure 23.** a) All looks, including looks to face and looks to elsewhere in the display. b) Proportion of infants’ anticipatory first looks to the goal location. Dashed line represents chance performance.

*Looking Time*
Looking time was measured for infants in both age groups (Younger Infants (Y, 5-7 months old) and Older Infants (O, 9-12 months old)). Looking times were right-skewed, so analyses were performed on log-transformed data. Mean looking times and log transformed looking times are shown in Figure 24. For each age group, a 2 (Trial) X 2 (Condition: swap or control) X 2 (Goal Shape: disk or triangle) Repeated Measures ANOVA was conducted. For young infants, there was no effect of Trial ($F_{(1,9)} = 0.498, p = 0.5, \eta^2 = 0.05$), no Condition X Trial interaction ($F_{(1,9)} = 0.02, p = 0.8, \eta^2 = 0.002$) and no effect of Goal Shape (that is, whether the agent’s goal was the disk or the triangle, $F_{(1,9)} = 0.289, p = 0.6, \eta^2 = 0.03$) was found. For older infants, a significant main effect of Trial was found ($F_{(1,15)} = 10.469, p = 0.006, \eta^2 = 0.411$), but there was no Trial x Condition interaction ($F_{(1,15)} = 0.204, p = 0.7, \eta^2 = 0.013$) and no main effect of Condition ($F_{(1,15)} = 0.088, p = 0.7, \eta^2 = 0.006$) or Goal Shape ($F_{(1,15)} = 2.67, p = 0.12, \eta^2 = 0.151$), suggesting infants looked longer on the first trial than on the second trial, regardless of what condition they were in.

Planned comparisons using Student’s t, carried out on log looking times averaged across trials, revealed no effect of Condition (Y: $t_{11} = 0.665, p = 0.52$; O: $t_{17} = 0.623, p = 0.54$). These results show that neither age group looked significantly longer at the swap outcome versus the control outcome.

A further analysis isolated looking times on trials where a correct anticipatory first look was made. There was still no effect of condition (Y: $t_{10} = 0.95, p = 0.36$; O: $t_{12} = 0.004, p = 0.99$). Even though infants anticipated the agent’s action correctly, they did not look longer when she reached to the opposite location and removed her goal shape.
Bayes Factor analysis showed relatively weak support for the null hypothesis for both 5-7-month-olds (BF = 1.93) and 9-12-month-olds (BF = 2.2), likely due to the small number of subjects in each group. Figure 25 shows the theoretical CDFs of the data for each age group.

<table>
<thead>
<tr>
<th>5-7-month-olds</th>
<th>9-12-month-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
</tbody>
</table>

Figure 24. Infants’ looking times (in seconds) to a swap or control outcome. Top panels show raw scores, while bottom panels show the log-scaled scores. Error bars indicate +/- 1 SEM.
Discussion

Infants’ pattern of looking reveals an interesting picture of how they parse and represent occlusion events when an agent is present. First, a significant portion of infants’ fixations were to the agent’s face. In fact, infants spent a great deal of time
looking at the agent throughout the course of the experiment. Figure 26 shows a heat map visualization of infants’ looking patterns to the entire display during a familiarization trial. Although the agent herself takes up a large portion of the space in the scene, infants’ fixations are clustered around her face (even though only the lower portion of the face is visible). Further, they make more fixations to the face than to the shapes in the scene, suggesting that infants are attempting to track the agent’s intention as well as the location of the shapes.

Figure 26. Heat map visualization of infants’ fixations during the fourth familiarization trial of the disk-goal condition. Darker red regions indicate more fixations.

Five-7-month-old infants correctly anticipated where the agent would search for her goal object, while 9-12-month-olds did not. However, this did not give younger infants a memory advantage over older infants. Both groups of infants did not look longer when the agent reached to the opposite location and pulled out her goal object. This result is especially surprising, since by 9 months of age, infants have been shown to
encode both shapes in the WM task with no agent present. Although they do not anticipate the agent’s action, her presence at the very least interferes with their ability to encode what went where.

There are several possible explanations of infants’ behavior on the current task. One possibility is that infants may have encoded the shape, but not the location, of the goal object. Recall that Yoon, Johnson, and Csibra (2008) found that 9-month-olds encoded the identity, but not the location, of an object when an agent pointed to the object in a communicative gesture, but showed the opposite pattern when an agent grasped the object. In the current study, the agent’s preference for one shape versus the other may have cued infants to attend to the shape of the objects as the most relevant aspect of the display, just as the communicative gesture did in Yoon, et al.’s study. The anticipation behavior of 5-7-month olds could indicate a prediction of the expected action of the agent without necessarily binding shape information to the location. That is, they might form a representation of the agent’s action while the shape is still in view or while it is being hidden, basing their anticipatory behavior on her expected action, and not necessarily because the goal shape is hidden in that specific location. Thus, when the agent removes the goal shape from the wrong location, infants fail to be surprised because they see the identity they expect.

Another possibility is that infants only encoded the shape at the goal location, but not at the opposite location. The agent’s preference for one shape versus the other may have acted to highlight one location. Since only one object is relevant to the agent’s action, only the contents of the goal location need be encoded and recalled. Since infants have limited WM, allocating those limited resources only to the goal-relevant location
may be the better strategy, since the contents at the opposite location are irrelevant to the agent’s goal.

Still another possibility is that the differences between the two shapes was not enough to induce infants to attribute a preference to the agent. Spaepen and Spelke (2007) found that 12-month-old infants attributed a preference for a kind of object, rather than a specific individual object, to an agent. Since the shapes used in the current study (chosen for continuity with the previous chapters) were of the same kind, infants may not have attributed a goal to the agent.

Further studies need to be conducted to disambiguate these possibilities. The current study is limited, in that infants were not tested with a swap outcome at the goal location (that is, the agent reaching to the goal location and pulling out the non-goal object). If infants are encoding the shape of the goal object, but not the location, they should be surprised when the agent pulls the non-goal shape out of either location. If infants are only encoding the shape at the goal location, they should be surprised when the agent pulls the non-goal shape out of the goal location, but they should not look longer when either shape is removed from the non-goal location.

A crucial difference between the current study and the other experiments presented in this dissertation is that, rather than a disembodied hand removing the occluder to reveal the shape, the agent herself takes an action on that shape, reaching behind the occluder and moving the shape from behind it to in front of it. Since the agent is taking an action, this may be a cue to infants to attend to and track what she is doing. It is also possible that representing and tracking an agent’s mental state taxes infants’ working memory. Under this possibility, infants are using their limited resources to keep
track of the actions that the agent is taking rather than to track and retain the identities and locations of the objects in the display. Tracking the agent takes precedence over tracking the object. Alternatively, it is also possible that “all bets are off” when an agent is involved. Infants have been shown to accept without surprise when a box exhibiting self-propelling qualities does impossible things, such as floating without support (Luo, Kaufman, & Baillargeon, 2009). Infants in may recall which shape went where, but they may not mind that an agent pulls out an unexpected shape.

A further study, in which the occluders are removed by a force other than the agent, will shed light on these possibilities.
V. General Discussion

In Chapter 1, I outlined a fundamental problem that faces the perceptual and cognitive systems of both infants and adults: how do we decide which object is in which location and which feature goes with which object, and how do we keep track of those bindings over short intervals of time in WM? The studies presented in this dissertation shed light on the structure of WM for objects by testing the limits on infants’ WM across the second half of the first year of life.

In Chapter 2, I asked what 6-month-olds remember when they forget the specific identity of an object. In a two-screen violation-of-expectation task, two shapes were hidden from infants’ view, and their memory for the harder-to-recall (first-hidden) of the two shapes was probed. Infants were not surprised when the object was revealed to have changed shape, but were surprised when the object disappeared completely (suggesting they expected it to persist, but did not detect a shape violation, Baillargeon, 2008). The results suggest that infants can represent an individual in WM without necessarily representing any identifying information about that individual. At 6 months, infants can recall only one shape (Kaldy & Leslie, 2005), but at least two individuals, when they see two hiding events side by side.

In Chapter 3, I manipulated the number of locations infants had to keep track of and measured the effects on what infants can recall about each location. I showed that 9-month-olds can recall two shapes in two locations, and can do so robustly, even when sampling confounds are more rigorously controlled. Thus, 9-month-olds’ performance is a marked improvement over 6-month-olds’, who can only recall one shape when they see two objects hidden in two locations. However, when 9-month-olds are asked to track
three locations, they can no longer recall two shapes, even when they are given multiple
different ways to succeed (either by recalling the specific shape of the object of interest,
or by recalling the location of the last-hidden shape). By 12 months, infants can recall
two shapes when tasked with keeping track of three locations, suggesting a development
across the three month span between the age groups tested.

In Chapter 4, I asked whether infants could be instructed to attend to the featural
differences between the objects to improve their ability to remember the harder-to-recall
object. Infants were familiarized with an agent who always chose one shape over the
other. When the objects were then hidden from view, we measured whether infants
would anticipate where the agent would search for her goal object, and whether they
would be surprised if the goal object was retrieved from the wrong location. Five to
seven-month-olds, but not nine to 12-month-olds, anticipated correctly, but neither group
looked longer at the unexpected outcome. In the discussion section of Chapter 4, several
possible explanations were considered for infants’ failure to be surprised when the goal
shape was retrieved from the unexpected location. Whatever the explanation, the results
show that infants encode different aspects of the display in the presence of an agent with
a goal.

The results of the experiments in Chapters 2-4 are summarized in Figure 1.
Looking at Figure 1, it is easy to see that a pattern emerges. Areas highlighted in pink
show where performance improves across age groups. At 6 months, infants cannot recall
the shape of the harder-to-recall object, but are surprised when it disappears. By 9
months, infants can recall more information about the harder-to-recall shape; they can
recall its specific identity (and by implication its existence, although that was not
explicitly tested). When 9-month-olds have to keep track of three locations, they perform more like the 6-month-olds in the two-screen task; they are not surprised when the probed object is swapped for the first-seen object, or even for the last-seen object. By 12 months, infants have no trouble recalling the shape of the probed object in the face of three locations.

Performance on the experiment in Chapter 4 is highlighted in purple. When an agent is involved in the scene, her preference for one object over the other results in infants across age groups unable to detect a change in the easier-to-recall object. However, younger infants are able to anticipate where the agent will search for her goal object, suggesting some prediction of how the event should unfold given the agent’s preference and the location of her goal shape.
Figure 1. Summary of the results of the experiments from Chapters 2-4 (the “easy swap” two-screen experiment with 9-month-olds shown in this chart is from Kaldy & Leslie, 2003). Infants can recall more shapes as age increases, and as the number of locations they have to track decreases. When an agent is present, none of the age groups tested looked longer at the shape change outcome, though 5-7-month-olds anticipated correctly where the agent would search for her goal shape.

The experiments from Chapters 2-4 show that infants’ WM cannot be characterized as having a fixed capacity that increases with development. While the number of shapes infants can accurately recall does increase, the evidence presented in this dissertation suggests that this is not due to a simple increase in the number of items infants can recall. Rather, it is the amount of identifying information that infants can recall about each item that increases across the second half of the first year. Further, the amount of identifying information that can be recalled about each object is dependent
upon the number of locations or objects in the scene. What does this tell us about the structure of infants’ WM?

The resolution-limited hypothesis

One possibility is that infants start out with severe limitations on the resolution of WM, which become less severe as infants approach the second year of life (and perhaps beyond). Zhang and Luck (2008) propose that adult WM is limited to a fixed number of high resolution slots with which to store the contents of a visual array. Infants also appear to have a fixed number of slots, close to the adult limit of about 4 (Feigenson & Carey, 2003; 2005; Feigenson, Carey, & Hauser, 2002). However, the results of the experiments in Chapters 2-4 show that, contrary to Zhang and Luck’s adult model, not all of these slots are created equal. Even when infants’ set size limit is not exceeded, infants can only recall a subset of objects in detail, while retaining an inkling of the objects for which identifying information has been lost. If there is indeed a resolution limit on WM, the infant evidence suggests that some slots have a higher resolution than others.

Although Zhang and Luck (2008) conclude that the fixed-resolution slot model provides the best fit for their data, they also suggest an alternative model that may better account for infants’ pattern of performance. The alternative model also posits a fixed number of slots, but the precision of each representation is dictated by a limited resource that must be shared across the slots. Zhang and Luck use a metaphor in which the WM resource is represented by a bottle of juice and each slot is represented by a cup. It is possible to pour equal amounts of the juice into each cup, so the resource is shared equally across slots. Or, it is possible to fill up one of the cups, and only put a small
amount of juice in each of the other cups, resulting in a high-resolution representation in one slot, and lower-resolution representations in the other slot.

How might the “slots + resources” model account for infants’ performance? Recall that in the multiple-location violation-of-expectation method used in Chapters 2-4 (and in Kaldy & Leslie, 2003; 2005), objects are hidden one at a time, and infants’ memory is probed for objects hidden at different points in the sequence. Since we cannot instruct infants about which location we are going to reveal, we assume that the last object they see is the easiest to remember (since it is presented alone, with no subsequent distracters), the penultimate object is slightly harder to remember, and so on, and indeed this assumption is supported by the behavioral evidence, some of which I have presented herein. As infants watch the sequence of events unfold, with objects being hidden one at a time, they might adjust the way they allocate WM resources as each new item is hidden. If each new hiding event is seen as the most relevant thing to pay attention to at the time, more resources may be allocated to encoding that object in detail, resulting in the highest resolution representation being of the last-hidden object.

Infants’ performance in the experiments reported in Chapters 2-4 provides some support for this hypothesis. At 6 months, infants can recall the shape of the last-seen object (Kaldy & Leslie, 2003), but cannot recall the shape of the first-seen object, although they are surprised when it disappears completely (Chapter 2). Under the slots + resources model, 6-month-olds’ performance would be due to allocating more resources to the most recently hidden object, and fewer resources to the first-hidden object, such that they have a high-resolution representation of one object, and a low-resolution representation of the other. Infants were unable to detect when the first-hidden object
changed shape because their memory representation of the shape at that location was fuzzy. And although this fuzzy representation does not contain enough information to identify the shape, infants would nevertheless remember that some object was hidden in that location.

By 9 months of age, infants have more resources at their disposal (more “juice in the bottle”) to allocate to encoding objects in the array; 9-month-olds can recall two out of two shapes where 6-month-olds can recall only one shape out of two. But when there are three objects in the to-be-remembered array, 9-month-olds perform more like 6-month-olds. If 9-month-olds have a fixed amount of resources to share across slots, their highest-resolution representation may be the most recently viewed object, while the rest of the array is subject to lower-resolution encoding. When there are only two shapes in the array, resolution of the harder-to-remember object is good enough to detect a change in shape; but with three shapes in the array, the fidelity with which each object is represented is no longer sufficient. By 12 months, infants are able to represent at least two of the objects in a 3-location sequence with enough fidelity to detect a shape change.

Further evidence comes from infants’ pattern of failure in Chapter 4. When the probed object was goal-relevant, infants in all age groups showed looking patterns that indicated they did not recall the shape of the other object. Since both 9- and 12-month-old infants have no trouble recalling two shapes without an agent present, this would suggest that infants are shifting resources toward the task-relevant object and away from the most recently seen object, resulting in a higher-resolution representation of the first-hidden object and a lower-resolution representation of the last-hidden object. The slots + resources model would predict that 9 to 12-month-old infants might be able to detect
subtler changes in the goal object, since they appear to be allocating more WM resources to encoding the goal object than they normally would to a non-goal object in that same location. For example, perhaps 9-month-olds will remember the color of the harder-to-remember first-hidden object, something they ‘normally’ do not do (Káldy & Leslie, 2003).

While the slots + resources model looks promising, it has limitations in its explanatory power. The model was constructed based on adult change-detection studies in which the entire contents of the visual array disappear instantaneously and simultaneously (thus in a way that specifies they no longer exist), and representations are retained over brief intervals before they are compared to a new array in which one or more of the items might have changed. Since the objects disappear simultaneously, decisions about where to allocate limited resources can be made all at once.

By contrast, for infants in the experiments described in Chapters 2-4 (and indeed, for both infants and adults in most real-world situations), decisions about resource allocation had to be made dynamically as events proceed in a 3-D space containing persisting objects. Because there is no way to instruct infants on task requirements, at the start of a test trial, infants had no way of knowing that the entire contents of the display would eventually be hidden from view, nor which location would be probed. Once the first object was hidden from view, it is likely that infants would dedicate all or most of their limited WM resources to encoding that object, equivalent to filling up the metaphorical cup with all the juice from the bottle. Once the second object is hidden, what happens to the representation of the first-hidden object? Do resources get re-allocated, such that each representation ends up with some share of the resources? What
controls the decision to dedicate the most resources to the last-seen object? Under this scenario, items that are already stored in WM suffer a decrease in resolution as new items are encoded, which suggests that resource management limits both maintenance and encoding of items in WM. While the current data does not rule out this possibility, Zhang and Luck (2008) do not propose a mechanism by which this dynamic, online reallocation of resources would be possible.

The binding-limited hypothesis

Another possibility is that infants are limited in the number of bindings between feature and location they can maintain in WM. When infants view the sequence of events in a multiple-location violation-of-expectation task, they have to keep track of what shape went where in order to respond correctly when their memory for each location is probed. This requires successful binding of feature to object and object to location.

If infants are limited in the number of feature bindings they can maintain, we would expect them to be able to recall the shape of only a subset of objects, but still be able to recall that an individual was hidden in each location (Leslie, et al., 1998). Under this hypothesis, keeping track of objects in multiple locations can be done relatively cheaply and in parallel (Scholl & Pylyshyn, 1999; Pylyshyn & Storm, 1988; Sagi & Julesz, 1985) while binding identifying features to those objects must be done serially and requires attention, while maintaining those bindings also requires attention (Wheeler & Triesman, 2002; Cohen & Ivry, 1989; Prinzmetal, Presti, & Posner, 1986). Thus infants, whose ability to allocate attention endogenously develops across the first year (Colombo, 2001), would show an increase in the number of identities they can recall,
while the number of individuals they can track would stay about the same across development.

The binding-limited hypothesis suggests that limited attentional resources must be allocated to maintaining the contents of locations during multiple-location hiding events such as the ones infants saw in Chapters 2-4. Since infants have no way of knowing which objects will be hidden and which locations will be probed, they must attend to each hiding event in turn and update their representation of the scene as best they can. And since objects are hidden sequentially, each subsequent hiding event can be thought of as a distractor, drawing infants’ attention toward the object that is being hidden. The last object hidden would be the most recent focus of attention, and would be the most likely to have identifying information bound to it. The presence or absence of identifying information in the representation of older items in the sequence would be subject to limitations on infants’ ability to maintain feature bindings over time and in the presence of distractors. When infants are cued in to the relevance of a shape at a particular location, as in Chapter 4, they may simply keep track of only that object, even if an intervening distractor object was hidden after the relevant object.

Thus, when 6-month-olds are surprised that an object had vanished, but not that it had changed shape, it suggests that they are able to track two individuals across two locations, but can only maintain featural information about one of these items. Nine-month-olds can track and recall shapes in two locations, but have trouble with three locations, suggesting that the bindings between feature and object are fragile and subject to the demands of the task. By 12-months, infants can recall at least two shapes when tracking three locations. But when infants’ attention is drawn by a highly salient agent,
infants will no longer encode two shapes in two locations, perhaps because of the resource draw imposed by the agent.

Conclusion and general future directions

The experiments presented in this dissertation show that infants’ WM can be characterized as limited but flexible. Even in the face of severe limitations on what they can recall, 6-month-olds retain an inkling of an object. By 9 months, infants’ WM for what went where has improved, but breaks down when they have to keep track of more than two locations or hiding events. By 12-months infants are better still, approaching adult-like levels of performance. But infants also demonstrate that they can use their limited WM strategically; when only one of two locations is task-relevant, infants don’t bother encoding what was hidden in the irrelevant location.

Future studies would seek to test the extent of the limitations and flexibility of infants’ WM. Various studies examining limitations are currently underway and were described in the discussion sections of each of the chapters. These studies would further examine the structure of infants’ WM, and would attempt to tease apart or reconcile the resource-limited and binding-limited hypotheses. However, these studies are limited in that we are measuring behavioral reactions to changes in the display. An important supplementary line of work would develop probabilistic models of infants’ WM to gain insight into just how much memory is required to perform the way infants do on these tasks.

In addition, studies exploring flexibility would shed light on how infants use WM. In Chapter 4, the goal-relevance of one of the items in the array limited what infants
recalled about each object. Can infants also be induced to encode more objects in a scene? Adults can use mnemonic techniques to remember more items, and take advantage of regularities in arrays of visual objects (Brady, Konkle, & Alvarez, 2008; 2009) or lists of numbers of words (Miller, 1951; Mathy & Feldman, under review) to increase the number of items they can store in WM. Feigenson and Halberda (2008) showed that 14-month-old infants can also be induced to recall more object identities when they are given spatial, conceptual, and linguistic cues that suggest objects can be sorted into chunks. What about young infants, whose WM is severely limited and who have limited knowledge of the world? Further, are there limits to the number of chunks infants can recall? Answering these questions would shed further light on the structure of WM, including how items are encoded, how they are maintained, and what the ties are between WM and long-term memory.
VI. References


Luo, Y. & Baillargeon, R. 2007. Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition* 105(3), 489-512.


