WORKING MEMORY CAPACITY IN CONTEXT: MODELING DYNAMIC PROCESSES OF BEHAVIOR, MEMORY, AND DEVELOPMENT

CONTENTS

I. WORKING MEMORY CAPACITY IN CONTEXT: MODELING DYNAMIC PROCESSES OF BEHAVIOR, MEMORY, AND DEVELOPMENT 7

II. THE COGNITIVE DYNAMICS THEORY OF VISUAL WORKING MEMORY 25

III. EMPIRICAL TESTS OF PREDICTIONS COMPARING CAPACITY ESTIMATES ACROSS TASKS AND DEVELOPMENT 56

IV. MODEL SIMULATIONS TESTING THE REAL-TIME STABILITY HYPOTHESIS OF DEVELOPMENTAL CHANGES IN VISUAL WORKING MEMORY 82

V. NEW QUESTIONS AND REMAINING CHALLENGES TO ACCOUNT FOR DEVELOPMENTAL IMPROVEMENTS IN VISUAL WORKING MEMORY 109

REFERENCES 128

APPENDIX 139

ACKNOWLEDGMENTS 148

COMMENTARY

EXPLORING THE POSSIBLE AND NECESSARY IN WORKING MEMORY DEVELOPMENT 149

Nelson Cowan

CONTRIBUTORS 159

STATEMENT OF EDITORIAL POLICY 160

SUBJECT INDEX 162
Working memory is a vital cognitive skill that underlies a broad range of behaviors. Higher cognitive functions are reliably predicted by working memory measures from two domains: children’s performance on complex span tasks, and infants’ performance in looking paradigms. Despite the similar predictive power across these research areas, theories of working memory development have not connected these different task types and developmental periods. The current project takes a first step toward bridging this gap by presenting a process-oriented theory, focusing on two tasks designed to assess visual working memory capacity in infants (the change-preference task) versus children and adults (the change detection task). Previous studies have shown inconsistent results, with capacity estimates increasing from one to four items during infancy, but only two to three items during early childhood. A probable source of this discrepancy is the different task structures used with each age group, but prior theories were not sufficiently specific to explain how performance relates across tasks.

The current theory focuses on cognitive dynamics, that is, how memory representations are formed, maintained, and used within specific task contexts over development. This theory was formalized in a computational model to generate three predictions: 1) capacity estimates in the change-preference task should continue to increase beyond infancy; 2) capacity estimates should be higher in the change-preference versus change detection task when tested within individuals; and 3) performance should correlate across tasks because both rely on the same underlying memory system. I also tested a fourth prediction, that development across tasks could be explained through increasing real-time stability, realized computationally as strengthening...
connectivity within the model. Results confirmed these predictions, supporting the cognitive dynamics account of performance and developmental changes in real-time stability. The monograph concludes with implications for understanding memory, behavior, and development in a broader range of cognitive development.

Working memory is a ubiquitous cognitive function that underlies a broad range of behaviors over development, including everyday activities such as following multi step instructions, keeping track of a referent to understand a pronoun, or learning to follow the rules of a game. Extensive evidence has shown that working memory is important for academic performance and everyday functioning; however, the basic cognitive processes subserving working memory through development are still poorly understood. In particular, the theories proposed to explain how working memory functions and develops are disconnected across research areas focusing on infancy, childhood, and adulthood. The aim of this research is to develop a process-oriented theory of working memory and how it develops that can span multiple tasks and developmental periods.

In order to arrive at a general theory of working memory development, it is necessary to understand how behavioral results relate across tasks designed to assess working memory in different age groups. Longitudinal studies have shown that infants’ performance on looking tasks predicts later performance on standardized tests of cognitive abilities at 2–3 years and 11 years (Rose & Feldman, 1997; Rose, Feldman, Jankowski, & Van Rossem, 2005). Performance on laboratory measures of working memory correlates with concurrent measures of children’s academic performance (e.g., Raghubar, Barnes, & Hecht, 2010) and general intelligence in both children and adults (e.g., Conway, Kane, & Engle, 2003; Cowan, Fristoe, Elliott, Brunner, & Saults, 2006). However, little is known about how infants’ performance on looking tasks relate to children’s performance on working memory tasks, much less how these lab-based measures relate to higher-level cognitive skills. Are there common processes across tasks and development, or do these results reflect separable contributions to higher level cognitive skills at different points in development?

Inconsistent results from training studies provide further evidence that current theories have not achieved a full understanding of how working memory relates to other cognitive functions. Recent reviews of working memory training studies show mixed results (Melby-Lervåg & Hulme, 2013; Shipstead, Hicks, & Engle, 2012; Shipstead, Redick, & Engle, 2010, 2012; Wass, Scerif, & Johnson, 2012). For example, Shipstead, Hicks, and Engle (2012) reviewed studies intended to improve attention through visuospatial working memory training (e.g., Klingberg, Forssberg, & Westerberg, 2002). Across these studies, Shipstead and colleagues found fairly limited transfer not only to “far” tasks (purported to measure constructs like fluid
intelligence) but also to some “near” tasks (other measures of visuospatial working memory), as well as inconsistent replication. They argued that one of the central obstacles to the progress of research in this domain is a lack of theories specifying the cognitive processes involved across tasks (see also Gibson, Gondoli, Johnson, Steeger, & Morrissey, 2012; Hulme & Melby-Lervåg, 2012; Logie, 2012; Melby-Lervåg & Hulme, 2013).

As a first step toward developing a theory that specifies the common processes across tasks, the current monograph builds from two largely disconnected research areas: working memory in children and adults, and visual memory development during infancy. Both of these areas have amassed a sizable knowledge base, but theories have thus far not been integrated to account for how infant memory relates to later assessments of working memory. Indeed, some theorists question whether infant looking behavior reflects working memory at all, or if the cognitive processes underlying correlations between working memory and more general cognitive skills only emerge later in development (see Cowan, 2007, for discussion).

In the sections that follow, I first review research and theories on working memory in children and adults, then describe infant research and how it may relate to later studies of working memory. Finally, I present a new theory designed to take a first step toward bridging these disconnected research areas by developing a specific account of developmental changes in visual working memory capacity. To preview, the theory I propose here extends an existing computational framework to model visuospatial cognition, the Dynamic Field Theory, to capture performance across tasks and development. Model simulations and behavioral results show that the lab tasks used with infants and adults can yield inconsistent capacity estimates even though they rely on the same underlying memory system. Furthermore, developmental changes in these tasks can be explained by a single mechanism: the strengthening of excitatory and inhibitory connections over development. This theory emphasizes the dynamics of cognition, that is, how cognitive processes combine in the moment of a particular task to guide behavior (Simmering & Perone, 2013) and presents a new perspective on development, reconceptualizing cognition in terms of stability. This approach to understanding cognitive development can be extended beyond the types of tasks captured in the model presented here, with implications for understanding working memory and development more generally.

WORKING MEMORY RESEARCH IN CHILDREN AND ADULTS

The psychological concept of “working memory” emerged with Baddeley and Hitch’s (1974) proposal of a model combining short-term stores of domain-specific information (i.e., verbal versus visual) with a combined
“central executive” that controlled and manipulated information within these stores. From this perspective, the central executive performed the “work” of working memory (i.e., control of the “slave” sub-systems) and the domain-specific stores were considered short-term memory. Despite Baddeley and Hitch’s clear distinction between short-term and working memory, the terms have been used inconsistently throughout the cognitive and developmental science literatures (see Simmering & Perone, 2013, for discussion). In general, simple laboratory tasks that require only maintenance of information for recall or recognition (such as repeating a list of digits or pattern recognition) have been considered measures of short-term memory. By contrast, complex tasks requiring manipulation of information (e.g., repeating a series of items backward) or switching between storage and processing tasks (e.g., counting sets of dots and storing the results from the series for later recall) have been attributed to working memory. Theorists have taken different analytical approaches to identify processes in common versus separable across these types of tasks, as reviewed in the next section.

Differentiating Short-Term and Working Memory in Children and Adults

One motivation for distinguishing short-term versus working memory is based on their relative predictive power of more general cognitive skills, such as reasoning and general fluid intelligence. In adults, these types of higher cognitive skills correlate more strongly and reliably with performance on complex tasks than simple tasks (see Unsworth & Engle, 2007, for review). This suggests that the processing involved in the complex tasks (i.e., the “work” of working memory) makes a more substantial contribution to general reasoning and intelligence than simple storage.

A second motivation for distinguishing short-term and working memory comes from latent variable analyses showing that simple and complex tasks load on different factors. For example, Kane and colleagues (2004) compared performance across multiple assessments of verbal and visuospatial memory using simple and complex tasks. Their analyses showed that simple tasks loaded to domain-specific storage factors (verbal versus visuospatial short-term memory) while the complex tasks loaded to a domain-general processing factor (which they termed “executive attention”). Analyses of children’s performance have indicated comparable separation of processes between the ages of 4 and 15 years (although some complex tasks were too difficult for young children to perform; Gathercole, Pickering, Ambridge, & Wearing, 2004; see also, Alloway, Gathercole, & Pickering, 2006; Alloway, Gathercole, Willis, & Adams, 2004).

These results differentiating short-term and working memory are compelling, but an important question remains: are the same processes
operating in similar tasks at different points in development? The similarity in factor loading is consistent with comparable processes, but it is not conclusive: differentiation between simple and complex tasks in all age groups does not necessarily establish the same source of this contrast within each age group. For example, adults may rely heavily on executive attention across complex tasks (as proposed by Kane et al., 2004); for children, however, inhibition (or other cognitive processes) may play a more central role on these tasks, with attention contributing equivalently to simple and complex tasks. It is also possible that the same processes operate across all tasks, but the demands are minimized in simple tasks, reducing the ability to detect correlations. Evidence for this type of limitation comes from studies showing that higher cognitive functions correlate with simple tasks when those tasks were made more difficult. For instance, Unsworth and Engle (2007) found that performance on simple span tasks related more strongly to general intelligence if trials were scaled to be beyond the individual’s capacity limit. Similarly, Cowan (2007) reviewed studies showing that children’s intelligence correlated with simple digit span performance if children’s ability to rehearse was suppressed. Thus, the processes underlying simple span tasks are related to higher cognitive skills, but these relations can be masked when task demands are low.

Another limitation of the latent variable approach to studying working memory is that it is difficult to apply earlier in development. First, it requires participants to complete a large number of tasks to increase the validity of the analysis; this could limit child samples to only those willing and able to perform many lab tasks. Second, the complex tasks that show strong predictive power in older children and adults are difficult for younger children (and impossible for infants). Alloway and colleagues have had some success adapting such tasks for children as young as 4 years (Alloway et al., 2004, 2006). However, their analyses also showed less separability of factors in this youngest age group (i.e., 4–5 years; Alloway et al., 2006), suggesting developmental changes in how cognitive processes combine to support behavior in these tasks. Even studies with adults have shown that the contrast between simple and complex visuospatial tasks depends on which tasks are chosen. For example, Miyake and colleagues found that performance on both simple and complex visuospatial tasks related equally strongly to measures of executive function, leading them to conclude that short-term memory may not be as dissociable from working memory in the visuospatial domain as in the verbal domain (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). Thus, relying on the latent variables identified from different sets of tasks may be missing some important cognitive processes as they change over development.

As an alternative approach, the current research uses a process-based computational model to explore how underlying processes combine in
specific task contexts. The utility of this approach within the developmental literature is well illustrated by the work of Munakata and colleagues. Their modeling results have shown how active versus latent memory traces compete to drive performance in two tasks in which children show perseverative responding at different ages: the A-not-B task during infancy (e.g., Munakata, 1998) and the dimensional change card sort task during early childhood (e.g., Morton & Munakata, 2002a). This modeling approach can explain why children perform in seemingly inconsistent ways across tasks that purportedly measure the same construct, for example, why infants perseverate in a reaching version of the A-not-B task but not in a looking version of the task (Munakata, McClelland, Johnson, & Siegler, 1997).

One advantage of using process-based computational models is that they can be applied across task types that may be suitable only during limited periods of development. Drawing on the examples modeled by Munakata and colleagues, young infants can be tested in variations of the A-not-B task (see Marcovitch & Zelazo, 1999, for review), but not the card sort task (which can be used with children 2.5 years or older; Zelazo, 2006); A-not-B-type tasks can be adapted for use beyond infancy (e.g., Schutte & Spencer, 2002; Spencer & Schutte, 2004), when children are able to complete the card sort task. A formalized modeling framework can be used to test whether the same processes are sufficient to explain performance in one type of task over development (e.g., comparing A-not-B tasks between infants and preschoolers) as well as across tasks during one developmental period (e.g., comparing preschoolers’ performance across A-not-B and card sort tasks). Furthermore, successful implementation of these processes within a formalized model can lead to novel predictions regarding performance in other tasks (e.g., extending insights from the A-not-B and card sort tasks to a speech interpretation task, Morton & Munakata, 2002b) or age groups (cf. effects of long-term memory on working memory in early childhood, Schutte & Spencer, 2002, and adulthood, Lipinski, Simmering, Johnson, & Spencer, 2010). Thus, process-based computational modeling can shed light onto the cognitive processes developing over a wide range of ages to support behavior across multiple laboratory tasks.

Theories of Working Memory and Development

There are three general theoretical approaches to describing how working memory, as measured in complex tasks, functions and changes over development. First, as described briefly in the previous section, is Baddeley’s multi-component model (Baddeley & Hitch, 1974). This model includes a central executive as a supervisory system that controls subsystems that originally included verbal and visuospatial storage systems, but was later updated to include an episodic buffer that integrated information across
modalities (Baddeley, 2000); each of these components has a limited capacity. The majority of evidence supporting this model comes from studies using complex paradigms with adults, as well as studies of adults with cognitive dysfunction (e.g., dementia; see Baddeley, 2001, for review). The multi-component model was not intended to address developmental change; as described above, evidence for separation of verbal and visuospatial subsystems and a domain-general central executive process has been found in children as young as 4 years (Alloway et al., 2006). Theories focusing on verbal skills suggest that developmental increases in rehearsal speed could support improvements (e.g., Hulme & Tordoff, 1989; but see Cowan, Elliott, et al., 2006).

A second approach is Cowan’s (2005) attentional model, which characterizes memory as an embedded process system in which a subset of long-term memory is activated as short-term memory, with a portion of this activated memory held within the focus of attention. Maintaining information in this state of heightened activation, accessible to conscious processes of manipulation, is considered working memory and has a limited capacity of only three to five “chunks” of information. Cowan, Fristoe, et al. (2006) differentiated the scope of attention—how much information could be stored—and control of attention—how information is processed—as partially overlapping but partially independent contributions to performance in complex tasks. In this model, development is driven by increases in both the scope of attention (i.e., capacity) and attentional control (see also, Cowan, Hismjatullina, et al., 2010; Cowan, Morey, AuBuchon, Zwilling, & Gilchrist, 2010). These authors note that developmental improvements in other cognitive processes, such as encoding strategies or rehearsal, also support performance on lab tasks, but argue that such processes cannot fully explain development (see also Cowan, 2013; Dempster, 1981). Critically, Cowan (2013) concludes that, while there is strong evidence for increases in capacity, the mechanism(s) underlying such increases are currently unknown.

A third class of theories describe working memory as an undifferentiated resource that is flexibly allocated to processing and storage, as opposed to the partial independence posited by Cowan and colleagues or the full independence in the multi-component model. In these undifferentiated models, memory in early development is limited by less efficient processing. Some theories focus exclusively on either time-based decay (e.g., Towe & Hitch, 2007) or resource-sharing interference (e.g., Case, 1995), but more recent models combine time- and resource-based factors (e.g., Oberauer & Lewandowsky, 2011), including consideration for processing speed supporting information refreshing in memory (e.g., Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009). Within these models, developmental change is attributed to the factor(s) considered most important for explaining performance limitations: reduced decay or refreshing speed is most relevant.
for time-based explanations, and resistance to interference is central to resource-based explanations, with both time and resources contributing to development in combined models.

In summary, theories explaining performance on complex working memory tasks differ in the proposed relation between processing and storage, as well as whether developmental improvements arise through changes in processing, storage, or both. Baddeley’s (2000) multi-component model proposed total separation of storage and processing, with no specific account for how each improves over development. Cowan’s (2013 see also, Cowan et al., 2005; Cowan, Fristoe, et al., 2006) embedded process model suggested partial separation between the scope and control of attention, with an emphasis on developmental increases in scope/capacity. The time- and resource-based accounts (e.g., Barrouillet et al., 2009; Case, 1995; Towse & Hitch, 2007) included undifferentiated processing and storage, but differentially attributed developmental change to resistance to interference, decay, and/or refreshing.

As this section demonstrates, these theories of working memory development were designed to account for performance on complex tasks that require demanding processing with concurrent storage, with much developmental change arising through improvements in balancing processing and storage. These theories can be difficult to apply to simpler tasks that minimize processing demands and emphasize storage, but still show marked developmental improvements (Simmering & Perone, 2013). Furthermore, it is unclear how or whether these theories could account for earlier periods of development, as complex tasks are rarely used with children younger than 6 years of age. This limited application to early development reasserts the question of whether the working memory processes described in children and adults can be measured in infants. The following section reviews infant tasks that correlate with later abilities and theories to explain these developmental improvements.

WORKING MEMORY RESEARCH IN INFANCY

Complex working memory tasks clearly cannot be used for testing infants’ memory, which raises the question of how (and whether) working memory can be assessed during infancy. Some theorists have questioned whether infants possess working memory (as opposed to only short-term memory), as infants cannot complete the kind of manipulations that are standard in working memory research with older children and adults (see Cowan, 2007, for discussion). Reznick (2007) argued that it is not possible to empirically differentiate short-term versus working memory, as it is conceptualized in adults, during infancy; he endorsed the term working memory to describe the
processes underlying the types of tasks discussed in this section. Cowan (2007) agreed with this description, noting that young children’s performance on simple tasks, like digit span, correlates significantly with intelligence; he concluded that infants and young children use working memory to perform simple tasks that later can be attributed to short-term memory. As such, the current monograph uses the term working memory in the context of infant tasks.

Infant visual working memory is primarily assessed through reaching or looking behaviors. The standard reaching paradigm is the delayed response task in which an object is hidden among, for example, three boxes and the infant is allowed to search for the object following a short delay (see Pelphrey & Reznick, 2003; Reznick, 2007, 2009, for reviews as well as discussion of the related A-not-B paradigm). Capacity in this task is conceptualized not only as the number of objects that can be remembered but also the length of time information can be maintained (also referred to as the durability of memory; Pelphrey & Reznick, 2003). The delayed response task can be used with infants as young as 5–6 months of age, as soon as reaching behavior becomes reliable. Results across a host of studies suggest a roughly linear increase in memory durability from 6 to 12 months, with a few studies demonstrating continued increases in durability through the second year (Reznick, 2009).

Feigenson and colleagues developed a search paradigm to assess how much information can be held in memory during infancy by hiding multiple objects in one container, then allowing infants to search after one or more of the objects has been retrieved by the experimenter (Feigenson & Carey, 2003, 2005; Feigenson & Halberda, 2008; Zosh & Feigenson, 2012). Infants’ search times after they saw all the objects removed from the container were compared to search times after they saw all but one of the objects removed (note that no objects actually remained in the box, regardless of how many were removed within the infants’ view, to ensure that search times were not affected by success at finding objects). Memory capacity of the total number of objects was inferred from longer search times on trials when infants saw all but one object removed. Using this method, infant’s memory capacity has been estimated to reach three objects at 12–14 months of age (Feigenson & Carey, 2003). Studies using this paradigm have not compared performance across age groups, and therefore cannot provide information on how performance changes over development.

Results from reaching tasks show that infants can remember the location of a single object by about 5.5 months and over development this information can be retained over increasingly long delays. By the end of the first year, infants have also gained the ability to retain multiple object representations in memory at once. Less is known about improvements in memory between 1 and 3 years of age due to the challenges inherent in testing these age groups with continuity to either younger or older age groups (Reznick, 2007);
reaching tasks are rarely used as an index of capacity beyond 18 months of age. Furthermore, performance on reaching tasks has generally not been compared to other tasks (either concurrently or longitudinally), so it is unknown whether or how it relates to performance on complex tasks or general cognitive skills.

A second common approach to assessing infant visual memory is through looking paradigms. These tasks generally take two forms: habituation tasks in which a single stimulus is presented until the infant’s attention decreases to a preset criterion, then a novel stimulus is presented (see Colombo & Mitchell, 2009, for review); and visual paired comparison tasks in which two stimuli are presented together, usually following familiarization to one of the stimuli (e.g., Fagan, 1970). These tasks rely on infants’ tendency to look toward a novel stimulus as an index of their memory for the familiarized/habituated stimulus, and both have been shown to predict cognitive skills later in childhood (e.g., Kavšek, 2004; McCall & Carriger, 1993; Rose, Feldman, & Jankowski, 2012).

Early research with the visual paired comparison task suggested that preferences changed developmentally, with young infants (2.5–3.5 months) preferring familiar objects, no reliable preference around 4.5 months, and novelty preferences thereafter (Rose, Feldman, & Jankowski, 2007). However, further studies showed that this sequence of preferences (familiar to null to novel) occurred within a single age group following increasing lengths of familiarization (see Rose et al., 2007, for review). These results indicated that amount of familiarity with a stimulus required to exhibit a novelty preference decreases over development, that is, the speed of encoding increases developmentally. Memory can also be retained for longer periods with age, in parallel to the findings of increased memory durability in the delayed response task described above. In addition, during the first year of life, infants are able to remember more complex visual stimuli and can recognize objects that require some abstraction (e.g., a slightly rotated view or two-dimensional version of three-dimensional object; Rose et al., 2007).

Infant performance in habituation paradigms shows similar developmental changes in speed of encoding, with older infants habituating more quickly than younger infants (Colombo & Mitchell, 2009). In both visual paired comparison and habituation tasks, individual differences within an age group correspond to developmental differences: an infant who shows faster encoding relative to other same-aged infants typically shows a stronger novelty preference (Rose et al., 2007). Furthermore, individual differences in looking tasks have long-term predictive power for higher cognitive functions. Rose and colleagues have shown that infants’ performance in paired comparison tasks predicts general cognitive functioning at 2–3 years (e.g., Rose et al., 2005), as well as working memory (along with other executive functions; Rose et al., 2012) and general intelligence at 11 years (see Rose et al., 2007, for
In particular, infants’ processing speed (as indexed by both reaction times and number of trials to reach criterion for a novelty preference) and novelty preferences in paired comparison tasks at 7 and 12 months had significant paths to working memory at 11 years (Rose et al., 2012).

In a meta-analysis, Kavšek (2004) concluded that measures of habituation rate (i.e., number of looks during familiarization) and dishabituation (i.e., novelty preferences following habituation) had a combined predictive validity of 0.34 for later intelligence assessed between 1.5 and 11 years of age. In an earlier meta-analysis, McCall and Carriger (1993) found similar predictive correlations for habituation or recognition memory (0.41 and 0.39, respectively), suggesting similar processes across looking tasks. Taken together, results from these meta-analyses and Rose and colleagues’ studies indicate that infants’ looking performance predicts a range of later cognitive functions.

How do results like these relate to the findings described above, that performance on working memory tasks relates to measures of intelligence in children and adults (Cowan, Fristoe, et al., 2006)? Three possibilities exist: the variance explained by infants’ looking behavior and older children’s working memory performance could (1) overlap entirely, (2) overlap partially, or (3) not overlap at all. Although much research has been dedicated to understanding how intelligence is predicted by different laboratory tasks (e.g., looking tasks in infancy, working memory tasks during childhood), little is known about how these various laboratory tasks relate to each other over development. A few studies have shown correlations between looking behavior during infancy and working memory later in childhood (e.g., Marchman & Fernald, 2008; Rose et al., 2012), but the specific tasks used to assess memory vary both within and across age groups, making it difficult to synthesize results across studies.

Theories of infant memory development differ based on the task used to develop the theory, but may provide an avenue for extension to later development. Developmental improvements on the delayed response task have generally been attributed to brain maturation, citing improved functioning within the prefrontal cortex during this age range (see Pelphrey & Reznick, 2003; Reznick, 2007, 2009, for reviews). These explanations generally suggest that developmental improvements in performance arise through more robust working memory representations and/or processes. As discussed further in Chapters 2 and 5, this notion of the robustness of memory has potential to explain a broader range of developmental changes. This requires a deeper understanding of the link between memory and behavior across tasks, which is a central goal of the current monograph.

Similar explanations can be found across theories of infant looking performance. Most theories are grounded in Sokolov’s (1963) comparator model, in which memory representations form over time with exposure to the stimulus. As infants encounter a new stimulus, they try to match it to a memory...
representation; if a match is found, attention is inhibited, but if no match is found, attention remains engaged until the stimulus is sufficiently encoded to no longer appear novel. Rose and colleagues attributed both individual and developmental differences to similar processes (Rose et al., 2007). In particular, they identified processing/encoding speed and attention (as indexed by look duration and rates of switching between displays) as separable contributions to recognition memory in the visual paired comparison task. Similarly, encoding speed in habituation is an important individual and developmental difference (Colombo & Mitchell, 2009). Although encoding speed was the focus of most theories in infant visual memory, McCall (1994) proposed that inhibition might be an equal (and not necessarily independent) contribution to infants’ looking behavior.

To formalize these processes of attention, encoding, and inhibition underlying looking behavior, Schönéer and Thelen (2006) developed a simple model of performance in habituation tasks; in this model, two nodes built excitation and inhibition, respectively, with exposure to a stimulus. This model showed how individual and developmental differences could arise through the same mechanism by illustrating how performance differed between “fast” and “slow” habituators when tested in one type of habituation task, a violation of expectancy paradigm (e.g., Baillargeon, 1987). Perone and colleagues expanded this type of architecture to account for a broader range of infants’ looking performance, including the shift from familiarity to novelty preferences in habituation (Perone & Spencer, 2013b) and visual paired comparison (Perone & Spencer, 2014), as well as capacity estimates from a variant of visual paired comparison, the change-preference task (described further below and in Chapter 2; Perone, Simmering, & Spencer, 2011).

These model implementations, along with similar connectionist accounts (e.g., Munakata, 1998; Westermann & Mareschal, 2004), point to the same brain maturation implicated in delayed response performance as the likely foundation for improving working memory processes underlying looking behavior. Both the Schönéer and Thelen (2006) and Perone (Perone et al., 2011; Perone & Spencer, 2013b, 2014) models provide formalizations to show how changes in connectivity within such models can give rise to increases in encoding speed, attention, and inhibition—processes central to theories presented by Colombo and Mitchell (2009), McCall (1994), and Rose et al. (2007).

The previous section argued that working memory theories were difficult to “scale down” to earlier development because they are tied closely to the complex tasks they were designed to explain. A more promising approach might be to “scale up” theories of infant memory to account for later working memory development. A step in this direction is illustrated by the model Perone and colleagues have used to address changes in infant looking behavior. In addition to the increases in encoding speed, attention, and inhibition mentioned above, the strengthening connections within the
model lead to memory representations that are more robust. Conceptual theories of development in reaching tasks posited more robust representations leading to improved performance; these parallels across explanations of development suggests that a single underlying developmental mechanism may be sufficient to explain a variety of behavioral changes (discussed further in Chapter 5; see Simmering, Schutte, & Spencer, 2008, for one example). Expanding this type of model could incorporate many of the conceptual explanations for working memory improvements within a single theoretical framework. This monograph aims to do this by starting with simpler tasks, as described in the following section.

VISUAL WORKING MEMORY CAPACITY AND DEVELOPMENT

The current monograph focuses on two tasks designed to measure visual working memory capacity in different age groups. Visual working memory provides a good starting point for expanding theories of development because it can be assessed beginning in infancy, unlike verbal working memory (which requires adequately developed language skills) or performance in complex tasks (which are difficult to use with children younger than 6 years; Simmering & Perone, 2013). Tasks assessing capacity estimates are the focus of this monograph because they are constrained to limit the use of other processes, as described further below.

The first task, change detection, was constructed to minimize verbal recoding or “chunking” items together, therefore providing a more reliable assessment of capacity limits in visual working memory (see Vogel, Woodman, & Luck, 2001, for a description of how the task was developed). This task can be used with children as young as 3 years of age with only minor modifications (Simmering, 2012). Figure 1 shows a standard change detection trial (e.g., Vogel et al., 2001). On each trial, a memory array is presented briefly followed by a short delay. Then, a test array appears in which either all of the objects match the memory array (no-change trials), or the feature(s) of one object changed to a new stimulus value (change trials). The participant reports whether the memory and test arrays were “same” or “different.” Performance is typically near ceiling for arrays with small numbers of items (referred to as set size), and performance declines as set size increases. Capacity is estimated from this task by combining performance across change and no-change trials, with corrections for guessing, across set sizes (see formula in Chapter 3). Adults’ capacity is typically estimated to be approximately three to five simple items (e.g., colored squares or oriented bars) in the change detection task (see Cowan, 2010, for review). Although the change detection task may seem more like a simple task than the complex tasks described above, Cowan, Fristoe, et al. (2006) and Rose et al. (2012) have shown correlations between
this task and higher cognition, suggesting that it could tap the processes of interest when trying to explain fundamental cognitive skills.

To assess developmental changes in visual working memory capacity, the change detection task has been modified for use with children between the ages of 3 and 12 years by reducing the number of trials and extending the duration of the memory array (Cowan et al., 2005; Cowan, Fristoe, et al., 2006; Riggs, McTaggart, Simpson, & Freeman, 2006; Simmering, 2012). Figure 2 shows results across these studies, with estimates increasing from about two to five items during childhood. As this figure shows, capacity increases steadily across this age range (see Simmering, 2012, for discussion of differing estimates in 5-year-olds). Of the theories described in the working memory section above, only Cowan’s embedded process model has been applied specifically to change detection performance. According to that perspective, the controlled nature of the change detection task leaves only increases in capacity (i.e., the scope of attention) to account for developmental improvements (Cowan et al., 2005; see also, Cowan, AuBuchon, Gilchrist, Ricker, & Sauls, 2011). However, as noted above, this model does not include a specific mechanism by which capacity increases (Cowan, 2013). Riggs, Simpson, and Potts (2006) proposed that developmental changes in neural synchrony could account for this capacity increase (discussed further in Chapter 2).

![Figure 1](image1.png)

**Figure 1.**—Sample trial of the change detection task in set size three (not drawn to scale). Different patterns represent different colors.

![Figure 2](image2.png)

**Figure 2.**—Capacity estimates from the change detection task with children. References: 1 = Cowan et al. (2005), these estimates were approximated from figures, precise values were not reported in text; 2 = Cowan, Fristoe, et al. (2006); 3 = Riggs et al. (2006); 4 = Simmering (2012) “card” task; 5 = Simmering (2012) standard task.
To explore capacity changes during infancy, Ross-Sheehy, Oakes, and Luck (2003) combined features of the change detection task with the visual paired comparison task to develop the change-preference task. In this paradigm, arrays of colored squares, like those used in change detection tasks, were presented on two side-by-side displays and blinked on and off over the course of the trial. A schematic of this task is shown in Figure 3; in one display (determined randomly across trials), a single color within the array changed following each delay, approximating the change trials from the change detection task. The other display held the same set of colors following each delay, as in no-change trials from change detection.

Rather than requiring explicit “same”/“different” judgments as in the change detection task, the change-preference task relied on infants’ general novelty bias, under the assumption that a novelty preference indicates memory for the items within the displays. By extension, if the number of items was beyond the infant’s capacity, the infant would not be able to detect the difference across displays and should show no preference. As such, the amount of time the infant looked to each display was tabulated on each trial and change-preference scores were calculated as the proportion of time the infant looked at the change display. Mean change-preference scores were compared to chance (i.e., 0.5), and capacity was estimated as the highest set size in which mean change preference scores were above chance. Using this paradigm, Ross-Sheehy et al. (2003) found that 6-month-old infants only showed reliable change preferences for displays with one item each. By the age of 10 months, however, infants showed preferences when the set size (i.e.,

![Figure 3](image-url)
number of objects in each array) was as high as four items. These findings suggest that between the ages of 6 and 10 months, infants’ visual working memory capacity increases rapidly to adult-like levels.

It is important to note that Ross-Sheehy et al. (2003) did not definitively argue that infants’ preferences reflected changes in capacity as measured in the change detection task. They note that infants could use a variety of processes to support performance in this task, although they made efforts in the task design to isolate short-term memory (the term they preferred). Ross-Sheehy et al. suggested that even if the task does not measure “absolute” capacity, it shows increases in “relative” capacity, that is, in the context of the looking task. Although these authors were careful to avoid strong claims about exactly how capacity compared across tasks, this subtle distinction has not been preserved in the literature more generally (e.g., Blaser & Kaldy, 2010; Cowan, 2007; Cowan, Fristoe, et al., 2006; Feigenson, 2007; Oakes, Messenger, Ross-Sheehy, & Luck, 2009; Riggs et al., 2006). For the purposes of this monograph, the characterization of the change-preference and change detection tasks as both measuring capacity follows the general literature and should not be considered a representation of the view presented in the original study by Ross-Sheehy et al. Furthermore, even if the change-preference task is considered to measure “relative” capacity or more general properties of visual working memory, it is worth investigating how the mechanism underlying developmental improvements in this task related to performance and development in the change detection task.

Comparing studies with children (shown in Figure 2) to results from the infant task is not straightforward: why is there an apparent discrepancy in capacity estimates over development, with 10-month-olds’ capacity estimated at four items in the change-preference task, but 3-year-olds showing capacity estimates of less than two items in change detection? Does this suggest the tasks are tapping different underlying cognitive systems and/or types of capacity, or could this pattern of results across tasks and development emerge through common processes? Addressing these questions requires understanding both the real-time processes involved in performing these tasks, and the developmental processes that could explain capacity increases. To this end, Chapter 2 reviews these tasks and related theories in depth, then presents a new theoretical account of performance across tasks and development. The following section provides a brief overview of this new theory.

PROPOSING A NEW THEORY OF VISUAL WORKING MEMORY CAPACITY DEVELOPMENT

The theory put forth in this monograph builds on components of previous theories to strive toward a comprehensive explanation of
performance across tasks and development. A central goal within this theoretical perspective is to understand the link between representations and behavior—what are the processes that operate in different lab tasks, and how do these relate to behavior beyond the lab? By establishing a specific account of the real-time processes by which visual working memory supports behavior, this theory can provide leverage on the questions of how visual working memory develops as well as how it relates to other cognitive skills. This monograph illustrates this theoretical account by bringing together the change-preference and change detection tasks within a single framework and showing how the same cognitive processes, as implemented in a processed-based model of visual working memory, can give rise to diverging capacity estimates across tasks.

As reviewed in the previous sections, a number of different explanations have been proposed to account separately for performance in complex tasks, infant looking paradigms, and assessments of visual working memory capacity. In particular, processing speed has been identified in both complex task performance and infant looking performance, although the sense of processing differs (i.e., performing operations vs. encoding, respectively). Increasing capacity was included in theories of both complex tasks and visual working memory capacity tasks. Within each of these domains, some theorists proposed that attention contributes to individual and/or developmental differences. Using a framework similar to Perone and colleagues’ (Perone et al., 2011; Perone & Spencer, 2013b, 2014; Simmering & Perone, 2013), my theoretical perspective is designed to integrate these types of changes into a single account. In particular, within this computational framework, changes in the strength of connections between nodes lead to emergent consequences for encoding speed, looking behavior, capacity, and comparison between memory representations and perceptual inputs.

SUMMARY

Working memory is a vital cognitive skill that underlies performance in a broad range of everyday tasks. Working memory abilities correlate reliably with higher cognitive functions such as reasoning and intelligence, both concurrently and longitudinally. Understanding how working memory develops and relates to other cognitive skills requires bringing together research and theories across sub-domains that are currently evolving separately. Complex tasks that require both processing and storage are the standard assessment of working memory in older children and adults, and theories of developmental changes in working memory have been tailored to explaining these tasks. Infant studies cannot employ such complex tasks, resulting in a line of studies and theories that do not connect infants’
developmental improvements in visual memory with the construct of working memory in later childhood and adults. As a bridge between these disconnected research areas, this monograph focuses on the change-preference and change detection tasks to develop a more general theory, with a goal of identifying common processes across tasks and development.

The next chapter presents the cognitive dynamics theory, implemented in a dynamic neural field model, to explain previous findings using these tasks and generate novel predictions for how these tasks relate within the same participants over development. Simulations illustrate how the model’s performance can explain not only why capacity estimates differ, but also how performance relates across tasks and what developmental changes could explain improvements in both tasks. These model simulations led to a set of three specific behavioral predictions, tested empirically in Chapter 3 by comparing performance of the same participants across the change-preference and change detection tasks. In Chapter 4, I test a fourth prediction—that this modeling framework can quantitatively capture these differences in capacity estimates across tasks and development using an existing developmental mechanism, the strengthening of excitatory and inhibitory connections in the model. Chapter 5 outlines remaining challenges and questions raised by these results, and concludes by discussing these findings within the context of the broader literature on working memory development.
II. THE COGNITIVE DYNAMICS THEORY OF VISUAL WORKING MEMORY
Vanessa R. Simmering

The goal of this monograph is to advance a theory of working memory development that brings together concepts from prior theories to provide a more comprehensive account of behavior across tasks and age groups. As reviewed in Chapter 1, research on working memory in children and adults is largely disconnected from research on working memory during infancy. The current monograph focuses on two tasks—the change detection and change-preference tasks—as a way to bridge these literatures. This chapter begins by describing the theoretical accounts of development in these two tasks, then presents a new theory that focuses on cognitive dynamics as a way to understand capacity limits and development. This chapter describes the theory in detail as it is implemented in a formal model.

Computational models provide an important tool in evaluating possible cognitive processes in common across tasks and developmental periods (Simmering et al., 2008). Implementing multiple tasks within a single computational framework imposes a level of specificity that can reveal gaps in conceptual theories of behavior and development. The current model is designed to test whether the different cognitive mechanisms that have been proposed to account for developmental changes in working memory across
domains might be explained through a common underlying source. Furthermore, the specific formalization presented here is based on models of neural population dynamics, which may provide a more direct link between behavioral, cognitive, and neural development in the future.

THEORIES OF DEVELOPMENT IN THE CHANGE DETECTION AND CHANGE-PREFERENCE TASKS

Chapter 1 introduced two methods for assessing visual working memory capacity: the change detection task, which has been used with children as young as 3 years through adults, and the change-preference task, which has been used with infants between 4 and 12 months of age. Studies using the change detection task have shown a gradual increase in capacity estimates from about two items at 3 years, up to four to five items by 12 years (Cowan et al., 2005; Cowan, Fristoe et al., 2006; Riggs et al., 2006; Simmering, 2012). As reviewed briefly in Chapter 1, most theories attribute this developmental improvement to an increase in the number of items that can be held in visual working memory at one time (Cowan et al., 2005, 2011; Riggs et al., 2006). However, Cowan (2013) noted that the mechanism by which capacity increases is still unknown.

Riggs, Simpson, and Potts (2011) proposed that capacity may increase through improvements in neural synchrony. In particular, they cite Vogel et al.'s (2001) description of the Raffone and Wolters (2001) model that attributes the limited number of items that can be held in memory to a limitation in the temporal resolution of synchronous neural firing. This model has been proposed to account for an object-based, rather than feature-based, capacity limit (e.g., adults can remember four multi-feature objects as robustly as they remember four single-feature objects, rather than features adding up across objects to exceed capacity; Luck & Vogel, 1997), but the same rationale can be applied to remembering single-feature objects. The speed of firing could increase developmentally, allowing more items to be represented simultaneously.

Although both of these developmental explanations focus on the number of items held in memory, theoretical accounts on the nature of capacity limits in the adult literature vary in the degree to which they emphasize the number of items (typically referred to as “slot” models) or the resolution of items in memory. From a resource perspective, Bays and Husain (2008) argued that adults can hold an unlimited number of items in memory, but as the number of representations increases their resolution decreases proportionally. As such, performance on high set size trials is limited due to the poor resolution of items, not a failure to encode some items. Two studies have shown that resolution improves developmentally during the same period.
that capacity increases (Burnett Heyes, Zokaei, van der Staaij, Bays, & Husain, 2012; Simmering & Patterson, 2012). Burnett Heyes et al. (2012) attribute this increase in resolution to “sharpening” neural representation; this is similar to the mechanism implemented in the current model (discussed further with respect to resolution in Chapter 5; see also Simmering & Miller, 2016; Simmering & Patterson, 2012).

The primary task used to assess capacity limits during infancy, the change-preference task, presents a seemingly inconsistent developmental trajectory. Results from this task suggest that capacity increases rapidly from one to four items between 6 and 10 months of age. In particular, capacity in 6-month-old infants was estimated to be only a single item, with no preference when tested in set size two or three (Oakes et al., 2009; Oakes, Ross-Sheehy, & Luck, 2006; Ross-Sheehy et al., 2003). When tested on variants of this task (described further below) 7.5-month-old infants’ capacity was estimated to be at least three items (the highest set size tested with that age group; Oakes et al., 2006, 2009). In the original version of the task, 10-month-olds preferred the changing display in set size four but not six (Ross-Sheehy et al., 2003). Thus, across these studies, capacity increased from one to at least three items between 6 and 7.5 months, and to about four items at 10 months.

What could account for these seemingly inconsistent trajectories across tasks and development? Cowan and colleagues argued that we should assume developmental change in capacity is truly monotonic, and the discrepancy between infant and child results arose from underestimation of capacity in children and/or overestimation in infants (Cowan, Morey, Chen, Gilchrist, & Saults, 2008). Simmering (2012) tested whether the change detection task was underestimating capacity by modifying the task to be easier for young children. Even with these modifications, capacity estimates from change detection were lower for 3- to 7-year-olds than for infants in the change-preferences task (see Figure 2). Cowan et al. (2008 see also, Cowan, 2007) proposed that the infant task could overestimate capacity if memory for only a subset of items in the array was sufficient to drive a change preference. Further studies with infants, however, suggested this was unlikely. In particular, Oakes et al. (2009) tested 6-month-olds in the change-preference task, but replaced all three items in the change display with new colors after each delay. If memory for a single item within the array is sufficient to drive a change preference, young infants should reliably prefer this “all change” display. This was not the case: 6-month-olds showed no preference, suggesting that memory for a single item within the multi-item displays was not sufficient to drive a robust change preference (but see Perone et al., 2011, and discussion below for evidence that robust change-preferences do not require all items be held in memory).

Cowan (2007) also proposed that the discrepant estimates across tasks could be due to different attentional demands, with the change detection task
being more demanding than the change-preference task. Support for this interpretation depends on how “attentional demands” are defined. In the change detection task, participants must attend to the display during specific windows of time within each trial to encode (during the memory array) and compare (during the test array) the items. By contrast, the change-preference task requires a participant to attend to two displays repeatedly across a 20-s test trial, with multiple opportunities to encode and compare items, then notice that items in one display change while items in the other display remain the same. Although these tasks arguably capture different notions of attention, it is not straightforward to determine a priori which would be more demanding.

As another explanation for the discrepancy between change detection and change-preference results, Riggs et al. (2006) suggested that the infant task either measures an aspect of visual working memory that does not scale with capacity, or measures a more passive type of memory. This suggestion is similar to competence-performance arguments proposed in other areas of cognitive development, suggesting that looking tasks reveal the underlying “competence,” while other response types are more demanding and may limit performance (e.g., Baillargeon & Graber, 1988). Although this type of argument may be intuitively appealing, such arguments are essentially unfalsifiable (see Thelen & Smith, 1994, for further discussion), and may be explained through more specific accounts of how performance relates to underlying cognitive processes (e.g., Munakata et al., 1997; Samuelson, Schutte, & Horst, 2009).

Oakes and colleagues suggested that developmental changes in infants’ performance in the change-preference task depends primarily on their ability to individuate items by binding the colors to locations in the task space (Oakes et al., 2006, 2009). In addition to the “all change” version described above (Oakes et al., 2009), 6-month-olds showed no preference in a condition in which the change display included the same three colors swapping positions within the display after each delay (Oakes et al., 2006). In both cases, however, 7.5-month-old infants exhibited reliable preferences for the change displays (Oakes et al., 2006, 2009). In these two conditions, the three items on the no-change display remained the same over delays, while all of the items on the change display change with each blink (either to new colors or to new color-location bindings). Oakes and colleagues argued that young infants failed to detect these changes because they cannot bind colors to locations—essentially, this lack of binding means that the three colors on the no-change display appeared to be changing because infants are unable to make the appropriate comparison to memory (i.e., without remembering which color was where, it is impossible to determine whether the colors are changing locations). Although this binding explanation could account for infants’ performance across conditions in the change-preference task, Oakes and colleagues (2006, 2009) have not specified what developmental mechanism
underlies the emergence of color-location binding between 6 and 7.5 months, beyond noting that brain areas related to feature binding in adults undergo significant developmental change during this period.

Simulations by Perone et al. (2011) provided an alternative explanation for development in the change-preference task. Perone et al. argued that the processes of memory formation and comparison, combined with the dynamics of looking, could explain how preferences arise in the change-preference task (as well as other visual-paired comparison and habituation tasks; Perone & Spencer, 2013a,b, 2014). In this account, developmental improvements arose from changes in the robustness of these underlying processes (through strengthening connectivity within the model; described further below). Using this approach, Perone et al. simulated performance by younger (6 months) and older (7.5 to 10 months) infants to capture three empirical findings: (i) developmental increases in capacity in the standard task (cf. Ross-Sheehy et al., 2003, Experiments 1 and 3); (ii) young infants’ preferences at higher set sizes in a non-memory version of the task (i.e., with no delays; cf. Ross-Sheehy et al., 2003, Experiment 4); and (iii) developmental change in the “all change” condition (cf. Oakes et al., 2009). These simulations suggested that young infants’ lack of preference in both the standard and “all change” conditions arose through weak memory representations of the colors within the no-change display. In some ways, this is similar to Oakes and colleagues’ explanation: both accounts attribute young infants’ failure to inadequate memory of the no-change display. Oakes and colleagues suggest this is because young infants cannot individuate multiple items in memory due to inadequate color-location binding. By contrast, the Perone et al. model did not represent the locations of specific items within the displays (only which display they were on), indicating that color-location binding is not necessary for older infants to show a preference in the “all change” condition. Rather, changes in the robustness or stability of memory processes underlying the memory of multiple colors were sufficient to explain both younger and older infants’ performance across conditions (Perone et al., 2011). Indeed, the young infant model could hold multiple items in memory, just not robustly enough to support a preference over the course of a trial. Furthermore, because the no-delay version did not require maintenance of items in the absence of input, younger infants could exhibit a robust preference in set sizes two and three, as the continued input provided more support for items in memory (discussed further in Chapter 5).

Could the developmental mechanism underlying improvements in infants’ performance in the change-preference task also account for capacity increases in change detection? The current monograph uses two existing models as a starting point to test this possibility: the Perone et al. (2011) model that can explain development in the change-preference task, and a model using a similar architecture developed to explain adults’ capacity
limits in the change detection task (Johnson, Simmering, & Buss, 2014). As described in the section that follows, the Johnson et al. model provides an alternative to “slot” and “resource” explanations of capacity limits in adults. The current implementation of this model tests whether it can also account for developmental improvements in change detection using the type of mechanism in Perone et al.’s model. This developmental mechanism has successfully accounted for a number of phenomena in infant visual memory beyond capacity (e.g., Perone & Spencer, 2013b, 2014), and a similar modeling framework and developmental mechanism has accounted for spatial cognitive development from 2 to 6 years and into adulthood (e.g., Simmering et al., 2008).

The following section explains the current theoretical perspective, which emphasizes the importance of understanding cognitive dynamics, and the modeling framework in which this theory is implemented. Sample simulations demonstrate how the model performs both the change detection and change-preference tasks. Implementing a developmental change similar to Perone et al.’s (2011) in the model illustrates a collection of consequences referred to as real-time stability. Model simulations of both tasks with “child” and “adult” parameters lead to three specific empirical predictions, as well as a computational prediction, tested in the following chapters.

A NEW THEORY OF VISUAL WORKING MEMORY CAPACITY DEVELOPMENT: COGNITIVE DYNAMICS AND REAL-TIME STABILITY

In this section, I describe the model implementation of my theoretical perspective to illustrate key aspects of the theory. In particular, model simulations show how both the formation of memory representations and the influence of memory on behavior differs between the change detection and change-preference tasks due to the different task structures. These results illustrate the importance of cognitive dynamics, that is, how cognition functions within different task contexts (Simmering & Perone, 2013). Furthermore, simulations of development show that strengthening the connections that underlie these cognitive processes in the model increases the real-time stability of the memory system. In contrast to long-term notions of developmental stability in which early behavior predicts later behavior, real-time stability refers to how robustly the memory system functions in the moment of a behavioral task (see Simmering & Perone, 2013, for related discussion). The term real-time stability denotes a collection of consequences that arise from strengthening connectivity (described further in Chapters 4 and 5): faster encoding speeds; more accurate representation of stimuli; decreased interference and decay during maintenance; increased capacity; more reliable use of representation in service of different behaviors; and less influence of task context.
Before reviewing the model implementation in detail, it is important to note the specific formalization presented here is only a tool for testing the concepts embodied in the cognitive dynamics account. Other types of computational models could potentially produce the same patterns of performance by implementing similar characteristics. One goal in using any computational model is to achieve a level of specificity necessary to generate and test behavioral predictions. Chapter 5 includes discussions of the implications of this model implementation for other models and theories of working memory development.

**Formalization of Cognitive Dynamics in a Computational Model**

The cognitive dynamics theory and dynamic neural field model implementation (referred to as “the dynamic model” in the current monograph for simplicity) that are central to this monograph fall within the broader perspective of dynamic systems theory. A primary emphasis of dynamic systems theory is that behavior is not rigidly preprogramed; rather, behavior is emergent, organized in the moment in response to the demands of the task, the details of the environment, and the history of the organism (Thelen & Smith, 1994). Furthermore, theorists within this approach across domains have argued that knowledge cannot be dissociated from process (e.g., McMurray, Horst, & Samuelson, 2012; Schöner & Thelen, 2006; Smith, Colunga, & Yoshida, 2010; Smith, Thelen, Titzer, & McLin, 1999; Spencer, Clearfield, et al., 2006). A dynamic systems approach brings a new perspective to the question of the development of visual working memory because many previous theories have focused mostly on the whether and how information is represented (e.g., capacity, Cowan, 2013; Oakes et al., 2006, 2009; Riggs et al., 2006, 2011, or precision of memory, Burnett Heyes et al., 2012), not how that information is used to support behavior across multiple task contexts (see Simmering & Perone, 2013, for further discussion).

Within dynamic systems theory, the focus is on how behavior emerges from multiple underlying causes, encouraging researchers to explore the various contributions to behavior, and to evaluate the robustness of behavior relative to the circumstances required to support them. This approach highlights the influence of task details, and seeks to specify the processes and mechanisms by which such details should affect behavior (e.g., Schutte, Simmering, & Ortman, 2011). Consistent with this dynamic systems view, a primary goal of the cognitive dynamics theory presented here is to reunite cognition and behavior by explicitly modeling the connection between underlying memory processes and the overt behavioral responses generated in specific memory tasks. The current project focuses on two tasks designed to assess visual working memory capacity, but ultimately this approach can be expanded to incorporate more tasks and domains (as discussed in Chapter 5).
The dynamic model presented here was derived from the Dynamic Field Theory framework (see Schôner, Spencer, & the DFT Research Group, 2015, for review). Amari (1977) first proposed the class of dynamic neural field models to capture the real-time dynamics of neural activation in visual cortex. Specific values along metric dimensions are represented as localized peaks of activation (also called “bumps” in biophysical models; see, e.g., Edin et al., 2009; Wei, Wang, & Wang, 2012) within dynamic neural fields. Figure 4 shows a simple two-layer model architecture for illustration. In this architecture, the top layer consists of a single field of excitatory nodes, which are reciprocally coupled to the bottom layer of inhibitory nodes. Along the x-dimension in this figure is the continuous dimension to which both fields of nodes are “tuned” (e.g., color or space). Specifically, nodes that respond preferentially to similar values along this dimension in the excitatory field engage in locally excitatory connections (see solid-curved arrow in Figure 4), as well as projecting excitation to similarly tuned nodes in the inhibitory field. These inhibitory nodes, in turn, project broad lateral inhibition back to the excitatory field (see dashed arrows in Figure 4). Critically, the spread of inhibitory connections is wider than the spread of excitatory connections, which keeps excitation from propagating throughout the excitatory field.

![Figure 4](image_url)

**Figure 4.**—A two-layer dynamic neural field model consisting of a single layer of color-selective excitatory nodes (top layer) reciprocally coupled to a similarly tuned layer of inhibitory nodes (bottom layer). Nodes coding for similar values along the metric dimension in the excitatory field engage in locally excitatory interactions (curved solid arrow), and transmit excitatory activation to the inhibitory layer (not shown). Nodes in the inhibitory layer transmit broad lateral inhibition back to the excitatory field (dashed arrows), which keeps activation localized within a “peak” in the excitatory field. Thick black lines within each field show the activation level across the metric color dimension, with thin black lines indicating the threshold (0) for interactions among nodes.
Through this local excitation-lateral inhibition profile of connectivity, activation in the excitatory field is able to form into localized peaks. These peaks of activation have characteristics that can be specifically applied to model different cognitive processes. In particular, peaks may exist in an input-driven or self-stabilizing state, in which peaks remain active only as long as the input is present; once input is removed, however, activation falls back to the resting state (see Supplementary Video 1). This state can capture perceptual processes that serve to detect items in the environment, yet return to a resting state after the input is removed. Note that this state emerges when local excitatory connections are weak relative to lateral inhibitory connections. If local excitation is strengthened, fields operate in a self-sustaining state, in which peaks of activation remain above-threshold (i.e., above zero) even after input has been removed (see Supplementary Video 2). This self-sustaining state is the form of working memory that is central to the framework presented here.

This form of working memory—self-sustaining peaks of activation within a dynamic field—highlights a central characteristic of the cognitive dynamics theory: an emphasis on stability. Stability is necessary for memory representations to be used in service of behavior or other cognitive tasks. At its most basic level, stability means resistance to change. In the context of behavior, stability refers to the ability to repeatedly produce approximately the same behavior at different times. Reliably repeated behavioral forms are not static; for example, the ways in which one’s hand moves to sign one’s name are not identical across repetitions. However, stability in the organization of the muscles in the hand produces signatures that are similar enough across repetitions to differentiate the true signature from a forgery.

In the context of cognition, this notion of stability can be applied to the repeated use of a cognitive process to support behavior. For example, in spatial recall, participants show variability in their memory for the same location when it is tested throughout an experimental session (e.g., Spencer & Hund, 2002)—the same memory processes produce approximately the same outcome over repetitions (recall of the location), but this memory is not static (rather, it varies across trials). This notion of stability is referred to here as real-time stability to emphasize the organization of behavior in the moment of the task; as discussed further below and in Chapter 4, changes in real-time stability play an important role in the cognitive dynamics account of development.

The two-layer model shown in Figure 4 can operate in a perceptual state (with the formation of self-stabilizing peaks as shown in Supplementary Video 1) or a working memory state (with the formation and maintenance of self-sustaining peaks as shown in Supplementary Video 2). Taken separately, these states capture important components of visual working memory tasks: detecting inputs in the environment and maintaining specific information over the course of brief delays. To perform the change detection and change-
preference tasks, however, these processes must be integrated. Such integration is achieved by combining two of these two-layer architectures—one with connectivity tuned to a perceptual state and one tuned to a working memory state—into a single three-layer architecture by pairing two excitatory fields with a single, shared inhibitory field, shown in Figure 5 (see Johnson & Simmering, 2015, for full description). This three-layer architecture combines perceptual and working memory processes to perform a third critical process in visual working memory: comparison.

The simulation shown in Figure 5 illustrates this comparison process in the three-layer architecture (see Appendix for details on the architecture and parameters). In this simulation, three items have been encoded in the working memory field (see peaks in bottom layer of Figure 5). Excitation associated with these peaks projects from the working memory field to similarly tuned nodes in the inhibitory field (see bumps in the middle layer of Figure 5). These nodes, in turn, project inhibition to similarly tuned nodes in both excitatory fields; local excitation in the working memory field is strong enough for the three peaks to sustain, but this inhibition suppresses further perceptual processing of the specific items that are held in memory (see top layer of Figure 5). Through this shared inhibition, detection of new items is driven by above-threshold activation in the perceptual field: input corresponding to a new color (i.e., one not held in memory) produces a self-stabilized peak in the perceptual field (see circle in Figure 5).

![Figure 5](image)

**Figure 5.**—The three-layer dynamic neural field architecture that forms the foundation of visual working memory processes in the cognitive dynamics theory. As in Figure 4, each layer is comprised of nodes selectively tuned to represent a continuous color dimension. Two excitatory layers perform perceptual (top layer; PF, perceptual field) and working memory (bottom layer; WM, working memory field) processes, while activation is mediated by a shared inhibitory field (middle layer; Inhib). This figure shows three self-sustaining peaks of activation within the working memory field, and a self-stabilized peak in the perceptual field. Axes for each field are as in Figure 4. See text for further details and interpretation.
further in the sections below, this activation in the perceptual field guides task-relevant behavior.

According to the cognitive dynamics account, performance in both change detection and change-preference tasks depends on the same basic processes of encoding, maintenance, and comparison embodied in the three-layer architecture. However, these processes are not sufficient to perform these tasks. The change detection task requires an explicit “same” or “different” judgment on each trial, and the change-preference task requires looking back and forth between displays. Furthermore, trials in the change-preference task are relatively extended in time (20 s compared to the 1–2 s trials in change detection), which brings long-term memory processes to bear on performance. Thus, in addition to the three-layer architecture, the dynamic model includes two Hebbian fields (to incorporate long-term memory) as well as a fixation system and a “same”/“different” response system.

The full model architecture used for this monograph is shown in Figure 6 (see Appendix for equations). The model consists of an excitatory perceptual

![Diagram of model architecture](image)

**Figure 6.**—The dynamic model used to simulate performance in the change detection and change-preference tasks. Shown within the dashed box, the five-layer dynamic neural field architecture consists of an excitatory perceptual field, an inhibitory layer, and an excitatory working memory field, with Hebbian layers connected to both excitatory fields. This architecture is connected to a fixation system (top) with nodes corresponding to looks to center (C), left (L), right (R), and away (A) locations within the task space, as well as a response system (left) with nodes corresponding to different (D) and same (S) decisions. Arrows indicate excitatory (solid) and inhibitory (dotted) projections between layers or nodes.
field, an inhibitory field, and an excitatory working memory field; the two excitatory fields are also reciprocally coupled to Hebbian fields. Each field consists of a collection of nodes tuned to represent a 360˚ color space. These fields pass excitation and inhibition as indicated by solid and dashed arrows, respectively, to capture the local excitation and lateral inhibitory connections described above. As activation builds in the perceptual or working memory field, traces of activation are laid down at associated sites in the respective Hebbian fields. These traces feed back onto the perceptual and working memory field, biasing the model to build peaks of activation at feature values that were activated before, leading to a form of long-term memory. Together, these five fields provide the perceptual, working memory, and long-term memory processes necessary to encode, maintain, and compare items presented over time in visual working memory tasks. To generate the behaviors relevant to the change detection and change-preference tasks, two simple behavioral systems are coupled to this system: a “same”/“different” responses system sends and receives activation from the perceptual and working memory fields; and a fixation system sends and receives activation from the perceptual field. How these systems function is described below in the context of the specific tasks.

Using this model as a formalization of the concepts in the cognitive dynamics theory, I address two key questions raised in Chapter 1: what real-time processes support performance in the change detection and change-preference tasks, and what developmental processes underlie visual working memory capacity increases? Similar models within the Dynamic Field Theory framework have been used to account for developmental changes in spatial cognition with a simple developmental mechanism, the spatial precision hypothesis (see Simmering & Schutte, 2015, for review). According to this hypothesis, connections within and between layers of the model strengthen over development. The current goal is to explore whether similar changes in visual working memory processes could account for capacity increases in both the change detection and change-preference tasks through increases in real-time stability. Testing this real-time stability hypothesis first requires understanding the processes that operate within each trial in the change detection and change-preference tasks. To this end, the following two sections explain how this dynamic model performs each task. Next, the final section of this chapter describes the consequences of strengthening connectivity within the model in the context of these tasks, leading to four specific predictions regarding how performance relates across tasks over development.

Change Detection Performance

To model performance in the change detection task, the necessary cognitive processes must first be identified. Recall the task design from
Chapter 1: each trial in the change detection task begins with a brief presentation (100–500 ms) of a memory array that includes a small number of colored squares to be encoded. After a short delay (250–1000 ms) during which colors must be maintained in memory, a test array is presented in which either all of the colors match the memory array (no-change trials) or one color has changed to a new color value (change trials). The participant’s task is to compare the test array to the items held in memory and respond “same” or “different.” Capacity is estimated from the number of correct responses on each trial type (using a formula described in Chapter 3).

The general processes of encoding, maintenance, and comparison were described above and illustrated in Figure 5, leaving only the response to be explained. As shown in Figure 6, nodes representing “same” and “different” responses are coupled to the working memory and perceptual fields, respectively. As such, activation from items (i.e., self-sustaining peaks) held in the working memory field supports a “same” response, while activation from the detection of new items (i.e., self-stabilized peaks) in the perceptual field drives a “different” response. Activation to these nodes is controlled by a “gating” node that autonomously controls the projection of activation from the excitatory fields to the response nodes; this prevents a response from being generated at the incorrect time in the trial (i.e., during presentation of the memory array). Once activation reaches the response nodes, they compete in a “winner-take-all” fashion such that only one node achieves above-threshold activation, thereby generating a discrete response on each trial.

Figure 7 shows activation within the three-layer architecture of the dynamic model during critical points during two trials to illustrate these processes. Although these figure panels could be construed as separate “stages” in the task, patterns of activation in the dynamic model evolve continuously throughout the trial. The stage-like character of performance arises through the timing of task events, rather than from different processes across stages. First, at the end of the memory array presentation, the three colors from the memory array have been encoded into the working memory field (Figure 7A). Next, these peaks were maintained throughout the delay, with shared inhibition producing troughs in the perceptual field at the same color values (Figure 7B). For a no-change trial, the same three colors were presented in the test array, which produced a slight increase in activation within the working memory field, but no above-threshold activation in the perceptual field due to the inhibitory troughs (Figure 7C). As such, activation from the working memory field projected to the “same” response node, leading to a correct response (Figure 7D). For a change trial, one color in the test array was new, leading to above-threshold activation in the perceptual field at this color value (Figure 7E). This activation projected to the “different” response node, leading to a correct response (Figure 7F).
How does this characterization of visual working memory compare to the “slot” and “resource” accounts that currently dominate the adult literature? As explained in depth by Johnson et al. (2014), this dynamic model encompasses components of these alternative accounts, but it is not a hybrid account. One of the key contrasts between these perspectives is in the proposed source of errors within the change detection task. From a slots perspective, errors
primarily arise from items not being held in memory (i.e., capacity is lower than the number of items in the array), although a smaller number of errors may arise through guessing (e.g., Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988) or attentional lapses (e.g., Rouder, Morey, Morey, & Cowan, 2011). From a resource perspective, on the other hand, errors are attributed to insufficient resolution only (e.g., Bays & Husain, 2008). By contrast, errors within the dynamic model can arise through any of the underlying processes: encoding, maintenance, comparison, or response (see Johnson et al., 2014, for further comparisons with slot and resource accounts).

Failures in encoding and maintenance are similar to the capacity limitations proposed by slot accounts, as the number of peaks that can be held in the working memory field at one time is limited. However, as illustrated by Johnson et al. (2014), two important contrasts with slots accounts arose through simulations of the dynamic model. First, the capacity limit in the dynamic model is not fixed: the number of peaks that were maintained across trials varied some due to noise, but also showed systematic influence of the stimuli that were presented. Their simulations showed that, at high set sizes, more similarity among the items within the memory array led to fewer peaks in the working memory field sustaining through the memory delay. The influence of similarity on memory illustrate why it is important to consider the dynamics of cognition—what is being remembered cannot be separated from the processes of remembering. In particular, the lateral inhibitory connections within the dynamic model produce a spread of inhibition surrounding each peak in the working memory field. When multiple peaks form near one another, their related inhibition combines, making it more difficult to form new peaks. Inhibition in the dynamic model also includes a global component that is independent of the peak locations; this global inhibition contributes to capacity limits by preventing new peaks from forming once activation has reached a certain point (see Johnson et al., 2014, for related simulations and behavioral analyses).

The second important contrast that arose through simulations of the dynamic model account of visual working memory is that the capacity estimate derived from behavioral responses was not equal to the number of peaks in the working memory field. Specifically, Johnson et al. (2014) calculated the mean number of peaks held on each trial in each set size for simulations of adults’ performance in the change detection task. Their behavioral results showed a mean capacity estimate of 4.58 items, while their model simulations produced a mean capacity estimate of 4.64 items, providing a close fit to the behavioral data. Analysis of the peaks within the model, however, showed that, on average, 5.79 peaks were held in the working memory field during the delay on set size six trials. Johnson et al. also calculated the mean number of peaks separately for each response type (correct versus incorrect “same” and “different” responses) and found that the model nearly always held all of the
items in the working memory field on “miss” trials (incorrect “same” responses) and sometimes had fewer peaks on “hit” trials (correct “different” responses), with the lowest number of peaks on “false alarm” trials (incorrect “different” responses). These simulation results highlight the importance of addressing more than just the nature of representations: how memory is used in service of behavior can produce errors even when all of the items are held in memory.

To illustrate capacity limits and potential sources of errors in the dynamic model, Figures 8 and 9 shows time slices through the three-layer architecture on incorrect trials. Figure 8 shows that all colors were represented in the working memory field at the end of the memory array (8A) and delay (8B). When the test array was presented (Figure 8C), though, the new color happened to be at a value that fell between two of the colors in memory (i.e., at 0° between −40° and 40°). Although activation pierced threshold in the perceptual field at this new color value, the lateral inhibition associated with the nearby peaks in the working memory field kept the overall level of activation too low to generate a correct “different” response (see Figure 8D). This “miss” error could be construed as a combined failure in

![Figure 8](image-url)
comparison—the surrounding inhibition made it difficult for the new input to be detected—and response—a signal too weak to generate the correct response.

Figure 9 shows a simulation of the dynamic model producing a false alarm error on a set size four no-change trial. On this trial, the four items were initially encoded, with peaks forming in the working memory field (Figure 9A). However, during the delay, these peaks were not all maintained: the peak at 80° died out (dashed arrow in Figure 9B). Without this peak in the working memory field, there was no corresponding trough of inhibition in the perceptual field at this color value (circle in Figure 9B). When the test array was presented with the same four colors, an input-driven peak formed at −80° in the perceptual field (circle in Figure 9C), leading to a “different” response (Figure 9D). This false alarm error reflects a failure in maintenance of the items from the memory array, leading to a repeated item appearing novel at test.

As these simulations illustrate, formalizing the processes of encoding, maintenance, comparison, and response in a dynamic model can shed
light onto capacity limits and errors in the change detection task. Through this formalization, Johnson et al. (2014) identified conceptual gaps in previous accounts of capacity limits, in that they did not address how representations were compared and how responses were generated in the task. This level of formalization also allows for testing whether and how the same processes could support performance in the change-preference task, how performance relates across tasks, and what developmental changes in these processes might capture the seemingly discrepant capacity estimates from the literature. The following sections consider these questions in turn.

**Change-Preference Performance**

The same general processes of encoding, maintenance, and comparison that were described above and illustrated in Figure 5 also support performance in the change-preference task. Recall the task design from Chapter 1: infants are presented with two displays, each of which contains a small number of colored squares. The colors are presented simultaneously for 500 ms, followed by a 250-ms delay in which both displays show a blank gray background, then a new 500-ms presentation of colors. On the no-change display, all of the colors remain the same following each delay. On the change display, one randomly selected color changes to a new color (not already present within the change display). Over the course of a 20-s trial, infants’ looking to each of the displays is recorded, and a change preference score is calculated from the proportion of time spent looking at the change display (out of the total time looking to either display). For infants to show a preference they must *encode* the colors on a given display, *maintain* them in memory through the delay(s), then *compare* them to the new display to determine if any colors have changed. If no colors have changed, the familiarity of the display leads to *decreased looking* to the current display; if a color has changed, the novelty supports *continued looking* to the current display.

This characterization of the relation between memory formation, recognition, and looking is formalized in the dynamic model. In addition to the processes of encoding, maintenance, and comparison shown in Figure 5, the fixation system (shown in Figure 6) allows the model to use these processes to generate looking behavior. This system consists of four self-excitatory and mutually competitive nodes: one associated with the left (no-change) display, one for the right (change) display, one for the center (attention-getter) light, and one for looks away from the testing apparatus. Activation to this system comes from two sources. First, input to the left and right nodes represents the presence of stimuli on those displays, with weaker inputs to the center and away nodes, where no task-relevant information is presented. With this input, activation builds in parallel for all four nodes until...
one of the nodes “wins” this mutual competition by entering a stable “on” state (i.e., above zero activation). Above-threshold activation of a fixation node represents the related fixation within the task (i.e., the left display, right display, center, or away). Fixating one of the displays in the task allows task-relevant input (i.e., the colors within that display) to enter the five-layer memory architecture.

Once activation enters the five-layer memory architecture, the encoding, maintenance, and comparison processes are engaged. These processes provide the second source of input to the fixation nodes: when activation is above-threshold within the perceptual field, it projects back to the fixation nodes (see bi-directional arrow in Figure 6), increasing the likelihood that the current fixation will continue. Thus, activation in the perceptual field—which corresponds to perceptual processing of new inputs—is fundamentally linked to fixation. Within the dynamic model, perceptual processing is suppressed for familiar items, which ultimately produces a novelty preference in the model.

Figure 10 illustrates how these processes operate in the context of a set size four trial of the change-preference task. Figure 10A shows activation of the fixation nodes over the 20-s trial. As this figure shows, the model primarily fixated the two displays, with some “away” fixations. Over the course of the trial, as the items were encoded in the working memory field and associated Hebbian field, the model fixated the change display (blue line) more than the no-change display (red line), resulting in a robust change preference. To illustrate the model’s performance, time-slices through the perceptual and working memory fields are shown during the dynamic model’s fixations of the no-change display (Figure 10C), the change display (Figure 10D and E), and back to the no-change display (Figure 10F). The stimuli presented at these points in the trial are shown in Figure 10B.

During the first fixation of the no-change display (Figure 10C) the four colors are encoded, forming peaks in the working memory field. The inhibition associated with these peaks suppresses activation at these values in the perceptual field, which reduces the activation sent back to the fixation system. This reduction in support for fixation, combined with the dynamics of the fixation system (see Appendix for description), leads the model to look away from the no-change display (see decrease in the node’s activation in Figure 10A). Through the dynamics of the fixation system, the activation of the node corresponding to the change display rises above threshold next, leading the model to fixate this display and receive those four color inputs. Note that, through random selection, one of the colors in the change display matched one from the no-change display (cyan, localized at 80˚; see Figure 10B). Because this input matched an item in memory, there was no above-threshold activation in the perceptual field for this color (Figure 10D). Similarly, the input at −160˚ did not pierce threshold due to surrounding
inhibition from nearby colors held in the working memory field (at $-120^\circ$ and $160^\circ$; note that the color space “wraps” from $-180$ to $180^\circ$). Thus, even though there were four colors in this display, only two produced peaks in the perceptual field.

Because fixation is supported by above-threshold activation in the perceptual field, there was less support to the right/change node during this fixation than in the previous fixation. As a result, the activation of this fixation node began to fall (see Figure 10A). Before the fixation ended, however, one item in the display changes (cyan to yellow), which led to a peak in the perceptual field (circle in Figure 10E). Note that the repetition of the other
colors in the display allowed the peak at $-160^\circ$ to pierce threshold in the perceptual field (Figure 10E). With more activation in the perceptual field, this maintained the models’ fixation of the change display (see D and E in Figure 10A). As the peaks built more strongly in the working memory field during this display presentation, more inhibition projected to the perceptual field and eventually drove activation below threshold, releasing fixation. Next, the model fixated the no-change display again. The model had maintained all of these colors in the working memory field during the fixations to the change display (see peaks at $-120$, $-80$, $80$, and $160^\circ$ in Figure 10D and E). As a result, activation at these values did not pierce threshold in the perceptual field, and fixation quickly released (see F in Figure 10A).

The dynamic model’s performance in this simulation highlights how the same visual working memory processes that support change detection performance contribute to looking behavior in the change-preference task: if peaks that correspond to the colors in the no-change display build quickly in the working memory field (with support from the associated Hebbian field), this inhibits activation in the perceptual field and removes support of the active fixation node. Essentially, the model tends not to fixate items that are held in memory—rather, it looks to items that are new. Over the course of many fixations and stimulus presentations within a change-preference trail, a preference for the change display emerges.

These two sections have illustrated my formalization of the processes of encoding, maintenance, and comparison in visual working memory, as well as how these processes interact with the response required in the change detection task and looking behavior in the change-preference task. This formalization of both tasks within a single computational framework provides the opportunity to bridge between infancy and later childhood to advance a more general theory of working memory development that can span multiple tasks and developmental periods. The next question addressed in this chapter is whether a single developmental mechanism within this unified computational framework could account for behavioral changes in both tasks. These simulations also address the discrepancy in capacity estimates across tasks and development, leading to empirical predictions described in the final section.

**Predictions Across Tasks Over Development**

As described briefly above, Perone et al. (2011) simulated developmental changes in infants’ performance in the change-preference task using a similar dynamic model architecture. They implemented development following the spatial precision hypothesis, which has been used extensively in the Dynamic Field Theory framework, especially with models of spatial cognition (see Simmering & Schutte, 2015, for review). In particular, to capture developmental improvements in visual working memory, Perone
et al. increased the strength of excitatory connections within the perceptual and working memory field, as well as the inhibitory connections from the inhibitory field to these two fields. By strengthening these connections the encoding, maintenance, and comparison functions within the dynamic model became more stable and robust for the “older” infant parameters. Their simulations showed that stronger connectivity led to a range of consequences: peaks in the working memory field formed more quickly and sustained more reliably (i.e., better encoding and maintenance), activation in the perceptual field was suppressed more effectively through shared inhibition (i.e., better comparison), and a closer correspondence between fixations and the contents of memory (i.e., better mapping of memory to behavior). These consequences are components of the notion of real-time stability introduced above. Put simply, with stronger connections, the dynamic model forms and uses memories more reliably in the context of the change-preference task.

The current monograph tests the hypothesis that increasing real-time stability can account for improvements in both the change-preference and change detection tasks through early childhood into adulthood. How would changes in real-time stability affect performance in the change detection task? This question can be answered through sample model simulations that scale connectivity over development. Because only adults’ performance has been fit before, this requires modeling development “backward” by decreasing the strength of connections to simulate earlier development. I conducted exploratory simulations with the parameters used to generate Figures 7–10 scaled to approximate children’s performance (see Appendix for details). Because these simulations are qualitative rather than quantitative in nature, I simply scaled the strength of local excitatory and lateral inhibitory connections by 0.5 (i.e., making these “child” parameters half the strength of the “adult” parameters), as well as scaling the input to be weaker and broader (multiplying the strength by 0.5 and width by 1.5; cf. Simmering et al., 2008).

Figure 11 shows a sample simulation of the dynamic model, using these “child” parameters, performing a set size four change detection trial. When the four items were presented, four peaks initially formed in the working memory field (Figure 11A). In comparison to Figures 7–9 with the adult parameters, the peaks formed by the child parameters are notably weaker and the inhibition that projects back to the perceptual field is quite weak; in Figure 7, inhibitory troughs are apparent in the perceptual field, whereas almost no inhibition is seen in the perceptual field in Figure 11. These differences are due to the relatively weak connections within and between fields with the child parameters. Another consequence of these weaker connections is shown in Figure 11B when two peaks “die out” during the memory delay (circle in Figure 11B). Essentially, reduced real-time stability
led to a failure in maintenance for these items. As such, when the same four items were presented at test, the model made a false alarm error, indicating that these two items were new (see peaks in the perceptual field, circled in Figure 11C). This sample simulation suggests that increasing real-time stability through strengthening connectivity within the dynamic model could capture developmental improvements in change detection performance.

Figure 12 shows the same “child” parameters in the context of a set size four change-preference trial, as Figure 10 showed with the adult parameters. Figure 12A shows activation of the fixation nodes over the course of the trial, with the stimuli from the first two presentations in Figure 12B. The model first fixated the no-change display (Figure 12C), and the four inputs formed peaks in the perceptual and working memory fields, although relative to the adult simulation (cf. Figure 10C) activation in these fields was lower and the inhibition projected into the perceptual field was much weaker. After the release of fixation to the no-change display, the model fixated the change display. As in the prior simulation (Figure 10), one color overlapped between the two displays (black, localized at $-160^\circ$; see Figure 12B), which supported...
the continued maintenance of this peak in the working memory field (see arrow in Figure 12D) and suppression of the related peak in the perceptual field. Activation corresponding to the other colors was again near threshold in the perceptual field, which led to a rapid decline in fixation (see blue line in Figure 12A). Just before the model looked away from the change display, however, one color changed (red to blue, localized at $-120^\circ$ and $160^\circ$, respectively), which produced a peak in the perceptual field at the new color value (circled in Figure 12E). This activation was relatively weak, however, and activation of the fixation node dropped below threshold (labeled E in Figure 12A). Next, the model fixated away, but the four peaks in the working memory field corresponding to the previous input remained above-threshold (Figure 12F).
This simulation with the child parameters illustrates the consequences of decreased real-time stability. Relative to the simulation with adult parameters, fewer peaks were maintained in the working memory field at one time, and the peaks were more likely to die out without support from input (i.e., without a repeated presentation of the item). However, over the course of a complete trial, items repeat regularly (due to the relatively small number of colors available) and Hebbian learning increases the support for peaks in the working memory field. Thus, although performance with the child parameters differed notably from that with the adult parameters in the moment-by-moment interactions, over the course of the entire task, performance may be more similar across parameter sets.

To test the consequences of these changes more concretely in the change-preference task, I conducted 40 simulations each with the child versus adult parameters (80 simulations total) in this task. In particular, I used the details from Ross-Sheehy et al.’s (2003) Experiment 3 (see Chapter 4 for details of the implementation in the model), in which they tested infants in two trials each in set sizes two, four, and six. The results of these simulations revealed change-preference scores that were significantly above chance (0.5) at all three set sizes for both parameter sets ($M_{SS2\text{-}child} = 0.55$, 95% CI [0.53, 0.56]; $M_{SS4\text{-}child} = 0.54$, 95% CI [0.52, 0.55]; $M_{SS6\text{-}child} = 0.55$, 95% CI [0.52, 0.56]; $M_{SS2\text{-}adult} = 0.57$, 95% CI [0.55, 0.59]; $M_{SS4\text{-}adult} = 0.55$, 95% CI [0.53, 0.57]; $M_{SS6\text{-}adult} = 0.56$, 95% CI [0.53, 0.59]). Thus, the dynamic model’s capacity as estimated from the change-preference task would be at least six items for the child and adult parameters.

These simulation results illustrate my first behavioral prediction: capacity estimates in the change-preference task should continue to increase over development, with both children and adults showing capacity estimates above the four items estimated at 10 months. According to the cognitive dynamics account, this increase in capacity relative to infants arises through continued increases in real-time stability: the processes that Perone et al. (2011) simulated through infancy continue to improve through early childhood into adulthood. It is not clear whether this prediction would be unique to the cognitive dynamics account and dynamic model presented here. The binding account put forth by Oakes et al. (2006, 2009) clearly explains the increase in capacity from one to two or more items, but it is less clear whether it would predict that capacity should increase from two to three to four items gradually during infancy, or beyond four items later in development. Rather, it might predict different developmental processes underlying the transition in memory from one item to more during infancy versus increases in capacity later in development.

The next question addressed by these simulations is how performance should compare across tasks when tested within the same individuals. Answering this question requires enumerating the differences between the
tasks contexts. First is the number of stimulus presentations: a trial in the change detection task includes colors in only two arrays (memory and test) separated in time; many trials are presented to the participant over the course of a session, but the colors are only meant to be maintained and compared within a single trial (indeed, failing to update memory from one trial to another would lead to errors). By contrast, a trial in the change-preference task includes two arrays presented simultaneously across displays 27 times throughout the trial, with ongoing comparisons that can occur within a single display across presentations, or between displays within a single presentation and/or across presentations. Second, the delays in the change detection task (1 s) are longer than in the change-preference task (250 ms), which could tax maintenance processes more, especially early in development when memory representations are relatively unstable.

The structure of the change-preference task allows for more items to be held in memory, even by the same memory system: the field parameters that could sustain a maximum of six peaks in the change detection task, as shown by Johnson et al. (2014), were used to generate Figure 10, which shows seven peaks sustaining in the working memory field. This difference in the dynamic model’s performance across tasks comes from both the long-term memory contribution (through the Hebbian field) as well as the repeated presentation of some of the same stimuli. Together, these two sources of activation support more items in the working memory field. Thus, the different task structures lead to memory for more items in the change-preference task than the change detection task.

The third important contrast between tasks is in the behavioral response. The change detection task requires a discrete “same” or “different” response on each trial, whereas preference scores in the change-preference task are calculated from the accumulation of looking to the two displays over the course of each 20-s trial (see Chapter 5 for discussion of a new “one-shot” version of the infant looking task by Oakes, Baumgartner, Barrett, Messenger, & Luck, 2013). The all-or-none response for change detection requires a robust signal from the corresponding component of the visual working memory system: activation in the perceptual field for a “different” response and activation in the working memory field for a “same” response. This mechanism of producing responses biases the dynamic model toward responding “same” (i.e., there is virtually always at least one peak in the working memory field to send activation to the “same” node), but it fits adults’ pattern of responses across trial types and set sizes well (Johnson et al., 2014). As shown in Figure 8, however, the response system is not error-proof, as activation in the perceptual field may be too weak to generate a response.

Looking behavior, in contrast to the “same”/“different” responses, can play out in a number of ways over a trial and still lead to a preference for the change display. As Figure 10A illustrated, fixations of both displays continued
throughout the 20-s trial. A robust preference could arise through multiple differences between looks to the change versus no-change displays—more fixations, longer duration of fixations, or both. The preference magnitudes reported by Ross-Sheehy et al. (2003) were around 0.55–0.58, with infants looking to both displays for about 10 s total per trial. This translates to infants looking, on average, about 1–1.5 s longer to the change display than the no-change display within a trial. This more graded measure of the underlying visual working memory system allows for a weaker—or less stable—memory system supporting preferences (see Chapter 4 for further discussion and illustration with model simulations).

A related point becomes apparent when comparing quantitative simulations across the change-preference (Perone et al., 2011) and change detection (Johnson et al., 2014) tasks: the mapping from peaks in the working memory field to capacity estimates is not direct in either task. Perone et al. found that a robust preference could be supported without representing all of the items from either display. In particular, they showed that the dynamic model could exhibit a significant change preference even when only half of the items from either display were held in the working memory field at any given time. This suggests that change-preference scores may overestimate the number of items held in memory. By contrast, Johnson et al. found that capacity calculations were lower than the number of peaks in memory (as reported above: a mean of 5.79 peaks yielded a mean capacity estimate of 4.64).

To summarize, these simulations suggest that the structure of the tasks allows more items to be encoded during the change-preference task than the change detection task. Furthermore, capacity estimates may be higher than the number of items in memory for the change-preference task, but lower than the number of items in memory for the change detection task. Together, these differences across tasks lead to a second empirical prediction: capacity estimates should differ across tasks even when tested in the same individuals. Is this prediction unique to the cognitive dynamics account? The most likely answer is “no,” although other theories have not specifically addressed this question. Cowan’s (2007) proposal that the change detection task could be more attentionally demanding than the change-preference task suggests he would predict this difference within individuals. Furthermore, the suggestion by Riggs et al. (2006) that the change-preference and change detection tasks tap different memory processes could also lead to a predicted discrepancy across estimates within individuals.

Although other theories may converge with the cognitive dynamics account on the predicted difference in capacity estimates across tasks, the dynamic model simulations specify that the same memory processes support performance across tasks. This leads to my third empirical prediction: performance should be correlated across tasks. Put differently, if the
predicted difference in capacity estimates across tasks is supported empirically, does performance in the change-preference task provide any index of capacity estimates in change detection? Based on prior empirical work, two possibilities arise. First, as reviewed in Chapter 1, Rose et al. (2012) found that novelty preferences during infancy predicted later cognitive functions, including working memory. Because the processes that support novelty preferences in infant visual-paired comparison tasks lead to preferences for the change display in the change-preference task, it is possible that preference scores could relate to capacity from change detection. Second, Rose et al. (2012) have suggested that switch rates are an important individual difference, as they also predict later cognitive functions; from their perspective, switch rates provide an index of attention.

Beyond these empirical results, however, the dynamic model simulations can provide further insight into potential relations in performance across task. In particular, the formalization of cognitive processes into a computational model allows for the assessment of the underlying processes that give rise to the variables suggested by prior research—change preference scores and switch rates—to predict whether and how they might relate to change detection performance. These two variables are affected most prominently by different factors within the dynamic model: excitation supporting fixation for change preference scores and inhibition releasing fixation for switch rates. This suggests that these factors could provide separable contributions to performance in the change detection task (discussed further in Chapter 4).

Consider change preference scores first. Because maintaining fixation of a given display in the dynamic model is driven by continued activation in the perceptual field (see Figures 10 and 12), change preference scores can be interpreted as the representation of new items from the change display. These representations are realized as self-stabilized peaks piercing threshold driven by excitation in the perceptual field. In the context of the change detection task, better representation of new items would correspond to higher hit rates (i.e., better performance on change trials) and therefore higher capacity estimates.

In contrast to change preference scores, switch rates in the dynamic model depend more on inhibition. In particular, fixation switches when activation projecting from the perceptual field decreases to the point that one of the competing fixation nodes can pierce threshold. As demonstrated in Figures 10 and 12, activation in the perceptual field is suppressed through inhibition corresponding to the peaks held in the working memory field. Thus, the more quickly peaks form in the working memory field, the more quickly inhibition suppresses activation in the perceptual field, and the more quickly fixation switches. In the context of the change detection task, this formation of peaks and suppression of activation is important for recognition of items already in memory, which should contribute most prominently to
correct rejection rates (i.e., better performance on change trials) and therefore higher capacity estimates in change detection.

These analyses of the dynamic model’s performance reveal the potentially separable contributions of excitatory and inhibitory connections to produce higher change preference scores and switch rates, respectively. Both excitatory and inhibitory processes are fundamental to how the dynamic model performs the change detection task, suggesting that preference scores and switch rates could both relate to capacity estimates from the change detection task. Although these predicted correlations may seem intuitive given some evidence for these looking measures’ predictive power, only through the level of specificity achieved in the dynamic model formalization can these predictions be attributed to the specific underlying processes (i.e., detecting novelty vs. familiarity, realized through excitatory vs. inhibitory connections). Prior explanations for performance and development in the change detection and change-preference tasks have not addressed the processes of comparison for detecting novelty versus familiarity in either task, but rather have focused on the nature of representations (mostly through slot-like characterizations). Through the simulations and predictions described here, the cognitive dynamics account provides both more specificity—detailing the processes of encoding, maintenance, comparison, and response—and more generality—predicting how these processes operate across tasks—than previous explanations of visual working memory development.

These first three predictions test key components of the cognitive dynamics account of performance in the change-preference and change detection tasks. The final prediction tested in this monograph focuses on the implementation of this theory in the dynamic model. Specifically, I present quantitative simulations to test whether developmental changes in both of these tasks could arise through the same underlying mechanism, namely, strengthening excitatory and inhibitory connections that support increasing real-time stability. Prior theories have put forth different explanations for the capacity increases in each task. If the same underlying mechanism could explain performance improvements in both tasks, however, this would provide a critical step toward identifying a more general source of developmental change in visual working memory. As outlined in Chapter 1, performance on working memory tasks in infancy and childhood predict more general cognitive skills, but to date the nature of these relationships is poorly understood. Most critically, it is currently unknown whether the processes that drive these correlations with general cognitive skills are the same between infant working memory tasks and tasks assessing working memory during childhood. If the dynamic model implementation provides evidence that a common developmental mechanism can span these two tasks from infancy through later childhood, this provides an important first step
toward establishing a more general theoretical account of working memory development.

SUMMARY

The goal of this chapter was to put forth a specific theory of performance in the change-preference and change detection tasks over development. This account focuses on cognitive dynamics, that is, the processes that support encoding, maintenance, comparison, and responses within various behavioral contexts. This approach brings a new perspective to the study of developmental changes in capacity limits, as most prior theories have focused on the nature of representations (i.e., capacity vs. resolution, whether colors are bound to locations) rather than how representations support different behaviors across task contexts. A central goal of the cognitive dynamics account is to use a computational implementation to illustrate the connections between cognition and behavior in the real-time context of laboratory tasks. By specifying these processes, this theory can provide an entry point to understanding how behavioral improvements over development relate to underlying cognitive changes.

Simulations of the change-preference and change detection tasks within the dynamic model led to four specific predictions to be tested with behavioral experiments and simulations in the subsequent chapters. A series of experiments tested children’s (ages 3, 4, and 5 years) and adults’ performance in both change-preference and change detection tasks. These experiments tested three specific predictions. First, when tested in the change-preference task, children and adults should show capacity estimates higher than infants—at least six items. Second, estimates of capacity measured from the same individuals in the change detection task should be lower than estimates from the change-preference task. Third, despite divergent capacity estimates, performance should be correlated across these tasks because they rely on the same underlying visual working memory system. In particular, model simulations suggest that both change preference scores and switch rates could provide potentially separable indexes of detection of novelty versus familiarity, which support performance on change and no-change trials, respectively, in the change detection task.

In Chapter 4, I use these empirical data to test my fourth prediction, that strengthening connectivity within the dynamic model can quantitatively fit developmental changes in visual working memory capacity across both tasks. Results supporting these four predictions demonstrate the utility of a computational implementation of the cognitive dynamics theory to explain how performance relates across tasks and development. Moreover, this project provides a test of the real-time stability hypothesis as an explanation of
developmental increases in visual working memory capacity from infancy through early childhood and into adulthood.

NOTES

i. I use the term “Dynamic Field Theory framework” to refer to a collection of theoretical and computational principles that have been implemented in dynamic neural field models. Although the model used for the current monograph is similar to other implementations within Dynamic Field Theory framework, the different theoretical context leads me to refer to the model separately from the larger framework.

ii. Although the original formulation by Amari (1977) was developed to model recordings of neural activation, the model presented here is abstracted away from this level of detail. Rather than focusing on neural dynamics directly, this type of model builds upon principles derived from neural data without specifying precise neural analogs (see Schneegans, Lins, & Schöner, 2015; Schoner, Reimann, & Lins, 2015, for further discussion). For this reason, the terminology conventionally used to describe neurons within fields and neural interactions within the model is not used in the current monograph.

iii. The term “spatial precision hypothesis” has been used extensively within the Dynamic Field Theory framework to refer to similar changes in connectivity within and between layers in dynamic neural field models (Perone et al., 2011; Perone & Spencer, 2013b, 2014; Schutte & Spencer, 2009, 2010; Simmering & Patterson, 2012; Simmering et al., 2008; Spencer & Hund, 2003; Spencer, Simmering, & Schutte, 2006; Spencer, Simmering, Schutte, & Schöner, 2007). Earlier implementations of the spatial precision hypothesis also included changes in the spread of connections (i.e., how broadly connections spread along the spatial dimension; Schutte, Spencer, & Schöner, 2003; see also Schutte & Spencer, 2009, Experiment 1). The current monograph includes slightly different implementation of these changes than has been used previously (see Appendix for details) and therefore uses the terminology “real-time stability” to describe the cognitive consequences of these types of changes in connectivity in order to keep the focus at the conceptual level rather than the specific model implementation.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher’s website.
III. EMPIRICAL TESTS OF PREDICTIONS COMPARING CAPACITY ESTIMATES ACROSS TASKS AND DEVELOPMENT

Vanessa R. Simmering

The experiments in this chapter test three behavioral predictions of the cognitive dynamics account, generated by implementing both change detection and change-preference tasks in a dynamic neural field model. First, capacity estimates for children and adults should be greater than four items in the change-preference task. Second, capacity estimates for the same participants should be lower in the change detection than change-preference task. Third, despite these divergent capacity estimates, performance should be correlated across tasks over development because both depend upon the same visual working memory system. Experiment 1 includes a large sample of 3-, 4-, and 5-year-olds and adults in the change-preference task. A subset of these participants also participated in Experiment 2 or 3, which tested color and shape versions of the change detection task, respectively (using modifications described in Simmering, 2012). Note that all participants completed the change-preference task first to prevent potentially introducing a bias to look for changes in the displays based on experience in the change detection task (where attention to change is explicitly instructed). Both tasks were completed within a single experimental session. To alleviate potential
concern that the order of tasks impacted performance, I compared the current Experiment 3 in a replication of this method with 3- and 5-year-olds that did not follow the change-preference task (Simmering, Miller, & Bohache, 2015); an ANOVA with Age Group (3 years, 5 years) and Change-Preference Task (with, without) comparing capacity estimates ($K_{\text{max}}$, described in Experiment 3 Results) showed no difference associated with completing the change-preference task (Task main effect $p = .63$, interaction $p = .62$). This suggests that the task order did not systematically affect children’s change detection performance.

EXPERIMENT 1: CHANGE-PREFERENCE TASK

Method

Participants

A total of 154 children participated in this study: 56 three-year-olds ($M_{\text{age}} = 3$ years, 4.46 months; $SD = 2.52$ months; 26 girls, 30 boys), 44 four-year-olds ($M_{\text{age}} = 4$ years, 2.87; $SD = 1.93$ months; 22 girls, 22 boys), and 54 five-year-olds ($M_{\text{age}} = 5$ years, 0.83 months; $SD = 1.75$ months; 28 girls, 26 boys). An additional eight children participated but were not included in analyses for the following reasons: incomplete data (1 five-year-old; see below for details), a visual abnormality (1 five-year-old), equipment failure (1 five-year-old), or experimenter error (1 three-year-old, 2 four-year-olds, 2 five-year-olds). Forty adults also participated in this study ($M_{\text{age}} = 21$ years, 5.35 months; $SD = 3.53$ years; 22 women, 18 men). An additional three adults participated but were not included in analyses because they wore glasses; the glare from the two displays made it impossible to code their looking direction.

Potential child participants were identified from a database of birth records maintained at a large Midwestern university. Parents of these children were contacted by mail and phone to schedule participation. Children received a small gift after participating. Adult participants were recruited through an introductory psychology course and received research participation credit, or were volunteers from the community. Adult participants and parents of child participants all reported normal or corrected-to-normal vision and no history of color-blindness.

Apparatus

The apparatus and stimuli were identical to those used in the infant study described previously (Ross-Sheehy et al., 2003, Experiment 3). Stimuli were presented on two 17-inch ViewSonic monitors using a Macintosh G3 computer. The monitors were positioned side by side with a 22 cm gap between them. Participants were seated approximately 100 cm from the displays; viewing distance and angle differed across participants due to
variations in height. From this distance, each monitor subsumed approximately 18.26° (w) by 13.50° (h) of visual angle. The total eccentricity of the displays was 88 cm (approximately 47.5° of visual angle).

As shown in Figure 3, a solid gray background was continuously present on both monitors, with arrays of colored squares blinking on and off together in time; one monitor contained the change display and the other monitor contained the no-change display. Each display contained two, four, or six colored squares that measured approximately 5 cm by 5 cm each. The set size (i.e., the number of squares on each monitor) matched across the two displays and remained constant throughout a trial. The initial square colors on each trial were selected at random from a set of nine colors: green, brown, black, violet, cyan, yellow, blue, red, and white. RGB values for the colored squares and background were black (0, 0, 0), blue (0, 0, 255), cyan (0, 255, 255), green (0, 255, 0), gray (150, 150, 150), red (255, 0, 0), violet (238, 130, 238), white (255, 255, 255), yellow (255, 255, 0). See Vogel et al. (2001) for approximate CIE values. Following Ross-Sheehy et al. (2003), the colors always differed from others within the same display, but could be repeated across the two displays.

The squares on both monitors appeared simultaneously for 500 ms, disappeared for 250 ms, and continued appearing and disappearing at these intervals for a total of 20 s per trial. For the no-change display, the colors remained unchanged throughout the 20-s trial. For the change display, one color would change to a new color at each onset; the location of this square was selected at random for each presentation, and the new color was selected at random from the set of colors that was not currently included in that display.

Procedure and Experimental Design

The procedure for this task was adapted from infant studies (Ross-Sheehy et al., 2003, Experiment 3) for use with older children. After completing the consent form, the participant was taken into the testing room; a parent sometimes accompanied child participants, but most children completed the task alone. The participant sat on a chair (or, for some of the younger children, in a parent’s lap) facing the two computer monitors. If a parent sat with the child in the testing room, she or he wore occluding glasses during the session. A large black curtain hung from ceiling to floor with openings to display the two monitors, as well as a small black box used as an attention-getter. The black box produced a red light and tone to draw the participant’s attention at the beginning of each trial.

In contrast to the infant version of this task, participants were given brief instructions to ensure that they would attend to the stimuli during the task. Specifically, they were told that they were going to watch six short videos, and that the experimenter would ask them questions about the videos at the end. No further instructions were provided to prevent potential bias in looking
behavior. After the task, the experimenter asked the participant whether he or she noticed any difference between the two displays in the videos. Most children’s responses referred to changes in the numbers, positions, and colors of the items across trials; only one child (a 5-year-old) correctly reported that the colors would change on one screen while the colors on the other screen did not change. Approximately, 70% of adults reported this difference across displays.

Each session included two trials in each of the three set sizes—one with the change display on the left, and one with the change display on the right—for a total of six trials. The order of trials, as well as which side the change display was on, was determined randomly for each participant. The experimenter was seated behind the curtain to initiate each trial. For most sessions, a trained observer recorded the duration of participants’ looking to each of the two monitors online using a program developed for the Macintosh computer (Habit; Cohen, Atkinson, & Chaput, 2000); when she was not available, however, she coded sessions offline using the video recording. In all cases, the coder could not see the images being presented on the monitors.

Each trial began with the attention-getter (the black box with a flashing red light and synchronous beeping); once the participant attended to the light, the experimenter pressed a key to begin presentation of the displays. For the duration of the 20-s trial, she then coded the participant’s gaze as looking to the left or right monitor. If the participant looked away from the displays, no look was recorded. Upon completion of each trial, the attention-getter was reactivated, beginning the same sequence for the next trial. This sequence was repeated across all six trials. The entire task took less than 5 min to complete.

Results

Because the same trained observer coded every behavioral session, typical calculation of interrater reliability is not possible. However, to ensure that this observer was sufficiently trained, a set of trained observers recoded approximately 25% of the sessions (49 of 195); the Pearson’s correlation between these coders’ results and the original coder were $r = 0.93$ for total looking time, $r = 0.92$ for change-preference scores, and $r = 0.90$ for switches. These interrater correlations are slightly lower than those typically reported in infant studies because coding children’s and adults’ looking is more challenging due to their more subtle eye movements; however, these values are similar to those reported in studies with participants from a comparable age range (e.g., $r = 0.91$ reported by Overman, Bachevalier, Sewell, & Drew, 1993).

Before analyzing the results, I inspected the data for trials on which the participant fixated only one display for the entire duration of the trial.
I excluded these trials because it was impossible for the participant to detect a difference between the two displays if only one display had been fixated. This led to the exclusion of 16 trials from 12 children’s data; these trials were roughly evenly distributed across set sizes. Half of these trials were fixations of only the change display, while the other half were fixations of only the no-change display. Five of these sixteen trials were from one 5-year-old child’s data, leading to only one trial of usable data from this participant; as a result, this participant was excluded from analyses (as listed Participants section above). Following Ross-Sheehy et al. (2003), the data were analyzed by comparing the total duration of looking per trial across age groups and set sizes, as well as computing change-preferences scores for comparison across age groups and set sizes. Additionally, I included a third analysis of switch rates across age groups and set sizes. As described in Chapter 2, switching may provide an index of how quickly and robustly memory representations are formed (Perone & Spencer, 2014).

Total looking time was computed automatically by Habit as the time the participant looked at either display on each trial. Scores were then averaged across the two trials to the same set size. Mean total looking times are shown in Figure 13A. Total looking time varied little across set size and age group, although adults’ looking time was slightly longer than children’s. Ross-Sheehy et al. (2003) reported that 10-month-old infants in Experiment 3 looked just under 10 s per trial on average for each set size ($M_{SS2} = 9.82$ s, $M_{SS4} = 9.48$ s, $M_{SS6} = 9.22$ s); children and adults in the present study looked considerably longer overall, but showed a similar pattern across set sizes.

I analyzed these data in a two-way ANOVA with Set Size (2, 4, 6) as a within-subjects factor and Age Group (3 years, 4 years, 5 years, adult) as a between-subjects factor. This analysis revealed only a significant main effect of Age Group ($F_{3, 190} = 15.62$, $p < .001$, $\eta_p^2 = 0.20$). Follow-up Tukey HSD tests ($p < .05$) showed that adults’ total looking time ($M = 16.57$ s) was greater than children’s total looking time for all three age groups, with no differences across the child age groups ($M_{3y} = 14.54$ s, $M_{4y} = 14.44$ s, $M_{5y} = 14.63$ s).

Change-preference scores were computed as the time each participant looked to the change display divided by total looking time to both displays for each trial. Scores were averaged across the two trials to the same set size. Mean change-preference scores for each age group are shown in Figure 13B. As this figure shows, change-preference scores were generally higher for set size two (SS2), with little change across age groups. Compared to the 10-month-old infants’ scores reported by Ross-Sheehy et al. (2003) in Experiment 3 ($M_{SS2} = 0.56$, $M_{SS4} = 0.58$, $M_{SS6} = 0.52$), children and adults showed similar preference scores (means reported below).

These scores were analyzed in an ANOVA with Set Size as a within-subjects factor and Age Group as a between-subjects factor. This analysis revealed only a significant main effect of Set Size ($F_{2, 380} = 6.04$, $p = .003$, $\eta_p^2 = .03$).
Follow-up Tukey HSD tests ($p < .05$) showed that change-preference scores in SS2 ($M = 0.58$) were significantly higher than in SS4 ($M = 0.55$) and SS6 ($M = 0.54$); change-preference scores did not differ between SS4 and SS6. To estimate capacity, I also compared mean change-preference scores to chance (0.50) using two-tailed $t$-tests (see Table 1). As indicated by the confidence intervals, the difference in change-preference scores is statistically significant.
intervals shown in Figure 13B, change-preference scores were significantly above chance for all set sizes for all age groups. These analyses yielded a capacity estimate of at least six items for all age groups; capacity estimates may have been higher if trials with higher set sizes had been included (discussed further below). The capacity estimates from this task showed an increase relative to 10-month-olds, but also suggest a higher capacity than the typical three to four items estimated in change detection studies with adults (e.g., Vogel et al., 2001), confirming my first prediction.

For the final analysis, switches between displays were tabulated as the number of times the participant looked from one display to the other during each of the six trials. Because there was a significant difference in total looking time between children and adults, I normalized these values by dividing the number of switches by the looking time on each trial to arrive at a switch rate. Switch rates were then averaged across the two trials to the same set size. Mean switch rates are shown in Figure 13C. As this figure shows, adults switched more frequently than children; in addition, children switched more frequently in SS2 than SS4 and SS6, whereas adults switched more frequently in SS2 and SS4 than in SS6.

I analyzed these data in an ANOVA with Set Size as a within-subjects factor and Age Group as a between-subjects factor. This analysis revealed significant main effects of Set Size ($F_{2, 380} = 20.13, p < .001, \eta^2_p = 0.10$) and Age Group ($F_{3, 190} = 44.66, p < .001, \eta^2_p = 0.41$). Follow-up Tukey HSD tests ($p < .05$) showed that switch rates were higher in SS2 ($M = 0.67$) than in SS4 ($M = 0.58$) and SS6 ($M = 0.57$), which did not differ. As can be seen in Figure 13C, this difference across set sizes is driven primarily by children’s pattern of performance, although the interaction did not approach significance ($p = .278$). Additional Tukey HSD tests showed that adults ($M = 0.86$)
switched more frequently than children did, and switch rates did not differ significantly across the child age groups ($M_{3y} = 0.48$, $M_{4y} = 0.55$, $M_{5y} = 0.59$).

**Discussion**

Children and adults showed change-preference scores significantly above chance for all set sizes tested, leading to an estimated capacity of at least six items for all age groups. Analyses showed no developmental differences in preference scores; both total looking time and rates of switching between displays were higher for adults than for children. Interestingly, the only significant differences found across set sizes in any measure were between set sizes two and four—roughly equivalent to expected capacity in change detection (cf. Figure 2). Both change-preference scores and switch rates were significantly higher in set size two, suggesting a possible link between capacity in change detection and performance in this task. I return to this question when comparing performance across tasks.

One important finding from this experiment is the comparison between children and infants tested by Ross-Sheehy et al. (2003). In their study, 10-month-old infants showed a significant change preference in set size four but not six, suggesting an adult-like capacity of four items. In the current experiment, children and adults’ change preference scores in set size six were significantly above chance, implying a higher capacity of at least six items. This suggests that capacity—as estimated in the change-preference task—continues to increase beyond infancy. It remains for future studies to explore this developmental transition further, mapping how performance changes across intermediate ages, as well as testing the upper limits of capacity in this task with older children and adults. The current study could not yield capacity estimates higher than six items because larger set sizes were not tested.

Set sizes larger than six were not tested in this experiment for two reasons. First, only nine colors have been included as stimuli in this task when used with infants; if set size was increased to, for example, eight items, each presentation would include at least seven colors that matched across the two displays. That amount of similarity across displays would likely alter results relative to the smaller set sizes that have been tested. Although more colors could be included, it is possible that memory performance would differ for a larger range of colors: multiple studies with adults have shown that change detection performance is affected by similarity between colors (Johnson, Spencer, Luck, & Schönner, 2009; Lin & Luck, 2009; Makovski, Watson, Koutstaal, & Jiang, 2010), and initial evidence with children suggests this effect could change developmentally (Grahn, Cooper, Miller, & Simmering, 2015; Simmering, Andrews, & Cooper, 2016). Second, increasing the number of items in each display would require altering the sizes and/or proximity of
the stimuli. The effect of shrinking the stimuli is untested, and it is unclear whether this would affect performance in this task differentially over development. Increasing the spatial distribution of the stimuli would require more fixations per display in order to encode all of the stimuli. Although this would probably not affect adults’ performance, it is possible that infants and/or children could have a harder time integrating across fixations and associating the stimuli within the correct display. These questions are important to consider in future research, and suggest possibly adapting new methods for testing the upper limits of performance in looking tasks over development.

EXPERIMENT 2: COLOR CHANGE DETECTION

To allow for comparisons across tasks, a subset of the participants from Experiment 1 also participated in the color change detection task that has been modified for use with young children (Simmering, 2012). By testing the same participants in both tasks, I can directly compare capacity estimates across the change-preference and change detection tasks to test the prediction that performance should be correlated across tasks.

Method

Participants

A subset of the participants from Experiment 1 also participated in this experiment: 14 three-year-olds (M age = 3 years, 4.68 months; SD = 2.35 months; nine girls, five boys), 14 four-year-olds (M age = 4 years, 2.56 months; SD = 1.77 months; six girls, eight boys), 14 five-year-olds (M age = 5 years, 0.36 months; SD = 0.78 months; five girls, nine boys). An additional 19 children were excluded from analyses for the following reasons: eight children did not understand the task (7 three-year-olds, 1 five-year-old); nine chose to end early (4 three-year-olds, 5 five-year-olds); one experimenter error (3-year-old); and one equipment failure (3-year-old). Fourteen adults from Experiment 1 also participated in this experiment (M age = 24 years, 0.91 months; SD = 4.67 years; eight women, six men).

Percentage correct and capacity estimates from these children were described previously (Simmering, 2012). The current monograph presents a more complete description of the data, including analyses of accuracy (A'), statistical comparisons to adults’ performance, and correlations with performance in the change-preference task. Simmering (2012) also included data from 7-year-olds, as well as comparison between this task and a control condition replicating the exact method used by Riggs et al. (2006) with 5- and 7-year-olds.
Apparatus

The apparatus and procedure were identical to Simmering (2012), and differed slightly for children versus adults, as described below when relevant. Children were familiarized with the change detection task using flashcards (3” × 3”) that showed SS1, SS2, and SS3; adults were given a verbal description of the task with sample trials illustrated on a sheet of paper. The task was completed on an 18” Macintosh computer monitor, which a female experimenter controlled with a keyboard and mouse that were kept out of the child’s view. The experimenter entered responses for child participants on the keyboard following their verbal responses, and adult participants entered their own responses while holding the keyboard in their lap. All participants were seated in an adjustable-height chair at a viewing distance of approximately 24–28” from the screen.

In contrast with the standard version of this task (shown in Figure 1), the items were presented within a rectangular frame to facilitate the description as a card-matching game (see Figure 14). The background of the computer screen was black, and the frame was a 5.75” tall × 4.75” wide gray rectangle. The position of the “cards” alternated between the left to right halves of the screen across trials. This helped reduce interference across trials. Stimuli were 1” × 1” colored squares as used by Simmering (2012), drawing from the set of colors used in Experiment 1: red, green, blue, yellow, cyan, violet, black, and...
white. Colors were selected randomly on each trial without replacement (i.e.,
a color could not appear twice within an array). Stimuli could appear at any of
five equally spaced positions in a 3" diameter invisible circle around the center
of the frame. For SS2–SS4, the stimuli appeared in neighboring positions; all
five positions were filled for SS5. For each set size, the positions were chosen
randomly for the first trial, but remained constant across the 12 trials in that
set size.

Procedure

After completing the change-preference task, child participants took a
short break before starting the change detection task to ensure that fatigue
would not impair performance. Adult participants were also offered a break,
though none chose to take one. Once the participant was seated in the testing
room, the task was explained and the necessary training completed. For
adults, sample trials were shown on a sheet of paper while the experimenter
described the task. For children, the task was described as a matching game
in which they had to help a teddy bear find cards that matched. The
experimenter then demonstrated the task using flashcards for training.
Flashcards were placed on the right or left side of a large sheet of cardstock in
the experimenter’s lap, alternating across trials. The first (memory array) card
was shown for approximately 2 s, and the child was instructed to “Look at the
picture and remember the colors.” The card was then removed and after a
brief delay, the second (test array) card was shown in the same location. The
experimenter then asked the child if the two cards matched. After the child
responded, the experimenter placed the cards face up next to each other on
the cardstock and praised or corrected the child as needed. Flashcard trials
were presented in the following order for all child participants: SS1 no-
change, SS1 change, SS2 change, SS2 no-change, SS3 change, SS3 no-change.
If the experimenter felt the child still did not understand the task, flashcard
trials were repeated as needed.

Once the participant understood the task, the experimenter began the
computerized version of the task. The first block of trials were practice, with a
2 s duration of the memory array and a 900 ms delay. After the delay, the test
array remained visible until a response was entered. Children responded
verbally, and the experimenter entered the response on a keyboard; adults
entered their response on the keyboard. The a key corresponded to a “same”
response and the l key to a “different” response, with stickers marking the
keys to help participants remember the correspondence. After a response was
entered, a chime played if the response was correct. This positive feedback
was included to help children stay motivated. If children seemed fatigued,
they were offered a break in between set size blocks.

The practice block included eight trials in random order: four trials in
SS1 and four trials in SS2 (two trials in each set size were change trials, two
were no-change trials). Within all blocks, a key press was required to initiate each trial; this allowed the experimenter to pace the task to each child’s attention, and reduced the likelihood that errors came from children not attending to the stimuli. Between each block of trials, the child was offered a short break. The order of test blocks was semi-randomly assigned, such that the first two blocks were either SS1 or SS2; for 3- and 4-year-olds, the third block was always SS3; for 5-year-olds and adults, SS3–SS5 were randomly ordered across the third, fourth, and fifth blocks. A different order was used for younger children because they often chose not to complete SS4 and SS5; these high set size blocks were difficult for all children, and young children tended to become discouraged as their performance declined. Each block of test trials included six change and six no-change trials in random order. For 5-year-olds, the duration of the memory array on test trials was shortened to 500 ms, to be comparable with previous studies (Riggs et al., 2006, see Simmering, 2012, for direct comparison); the longer duration was used for younger children to ensure adequate time for encoding, as young children tended to look away from the screen more frequently during stimulus presentation. The total duration of the task was approximately 20–30 min.

One additional feature of the task differed for adult participants relative to child participants. Because this design includes a longer memory array, which may encourage verbal recoding and/or rehearsal in adults, adults were given a verbal load to rehearse. At the beginning of each block, a three-digit number appeared on the computer screen, and adults were instructed to repeat this number throughout all of the trials for that block. The verbal load was not included for children because previous research suggests that children this young do not spontaneously verbally recode or rehearse visual stimuli (e.g., Hitch, Woodin, & Baker, 1989; Miles, Morgan, Milne, & Morris, 1996, see Pickering, 2001, for review).

Results

Participants’ responses were classified as correct rejections, hits, misses, and false alarms, as described in Chapter 2. Figure 15 shows the distributions of these response types for each age group separately. Note that correct rejections and false alarms sum to 1.0 (all no-change trials) and hits and misses sum to 1.0 (all change trials); overall percent correct can be estimated as the average of correct rejections and hits for each age group. As this figure shows, these two classes (i.e., correct responses) were the most common responses across ages and set sizes. In addition, performance generally decreased (i.e., errors increased) as set size increased, especially for children.

For each set size, two measures were computed from participants’ responses: accuracy ($A'$) and capacity estimates ($K$). Although most 5-year-olds and all adults completed SS4 and SS5, many 3- and 4-year-olds did not due
to difficulty keeping children motivated to complete blocks with harder trials. Therefore, I analyzed accuracy across all age groups for SS1–SS3 only, but across all set sizes for 5-year-olds and adults. For each participant, however, I included data from all completed set sizes to compute the maximum capacity estimate (described below).

**Accuracy (A’)**

Accuracy was calculated based on rates of hits (H) and false alarms (FA), using Grier’s (1971) formula, updated by Aaronson and Watts (1987) to accommodate below-chance performance:

If $H \geq FA$ : $A' = \frac{1}{2} + \{[(H - FA) * (1 + H - FA)]/[4*H*(1 - FA)]\}$

If $H < FA$ : $A' = \frac{1}{2} - \{[(FA - H) * (1 + FA - H)]/[4*FA*(1 - H)]\}$
An accuracy score of 1 indicates perfect performance, and a score of 0.5 indicates chance performance. This measure was chosen because, like $d'$, it is affected differentially by errors on change versus no-change trials, but unlike $d'$ it allows for cases in which the false alarm rate exceeds the hit rate (as was the case for some children) and follows a more intuitive range (similar to percent correct). One limitation of this approach is that this equation is invalid when H and FA are equal to each other and 0 or 1, producing 0 as the denominator. In these rare cases (only 2 of the 56 $A'$ scores calculated here), I set $A'$ to 0.50 (chance) because they reflect participants repeating a single response across all trials within that set size.

Accuracy scores were computed separately for each set size block for each participant. Mean accuracy scores across set sizes for each age group are shown in Figure 16. As this figure shows, accuracy was highest for adults and lowest for 3-year-olds, and decreased across set size for all age groups. Error bars showing 95% confidence intervals indicate that, even at their lowest accuracy, children were still performing above chance (0.50); also, adults’ performance was not distinguishable from ceiling (1.0) across SS1–SS4.

An ANOVA with Set Size (1, 2, 3) as a within-subjects factor and Age Group (3 years, 4 years, 5 years, adult) as a between-subjects factor revealed significant main effects of Set Size ($F_{2, 104} = 41.35, \ p < .001, \ \eta^2_p = 0.44$) and Age Group ($F_{3, 52} = 17.71, \ p < .001, \ \eta^2_p = 0.51$), which were subsumed by a significant Set Size $\times$ Age Group interaction ($F_{6, 104} = 5.75, \ p < .001, \ \eta^2_p = 0.25$). Tests of simple effects for each age group separately revealed significant set size main effects for 3-, 4-, and 5-year-olds ($F_{2, 26} = 12.43, 21.59, 14.69$, respectively; all $p < .001; \ \eta^2_p = 0.49, 0.62, 0.53$, respectively), but not for adults ($p = .55$). For each group of children, follow-up Tukey HSD tests ($p < .05$) showed that accuracy was lower in SS3 than in SS1 and SS2, although SS1 and SS2 did not differ from each other (see Figure 16). Thus, the Set Size by Age Group interaction was driven by a different pattern across set sizes for

---

**Figure 16.**—Mean accuracy scores ($A'$) in Experiment 2 across set sizes (SS), separately for each age group; note that chance equals 0.5. Points are offset horizontally to improve visibility. Error bars show 95% confidence intervals.
children than adults, with children’s performance declining significantly in SS3.

Because most 5-year-olds (n = 12) completed all set sizes, I was also able to compare their performance and adults’ in an ANOVA with Set Size (1, 2, 3, 4, 5) as a within-subjects factor and Age Group (5 years, adults) as a between-subjects factor. As with the analysis of smaller set sizes and all age groups, this ANOVA showed significant main effects of Set Size ($F_{4, 96} = 18.37, \ p < .001, \ \eta^2_p = 0.43$) and Age Group ($F_{1, 24} = 107.70, \ p < .001, \ \eta^2_p = 0.82$), which were subsumed by a significant Set Size x Age Group interaction ($F_{4, 96} = 9.69, \ p < .001, \ \eta^2_p = 0.29$). Tests of simple effects revealed significant set size effects for both 5-year-olds ($F_{4, 44} = 20.63, \ p < .001, \ \eta^2_p = 0.65$) and adults ($F_{4, 52} = 3.76, \ p = .009, \ \eta^2_p = 0.22$). Follow-up Tukey HSD tests ($p < .05$) showed that 5-year-olds’ accuracy in SS4 and SS5 was significantly lower than in SS1 and SS2; all other differences were not significant. For adults, accuracy in SS5 was lower than only SS2 and SS3 (see Figure 16). Again, the Set Size by Age interaction is driven by a different pattern over set sizes for 5-year-olds than adults, with the latter group not showing a decline in performance until SS5.

**Capacity Estimates (K)**

Capacity estimates were calculated using Pashler’s (1988) formula, separately for each set size (SS): $K = \text{SS}^* (H - FA) / (1 - FA)$; note that capacity can, at maximum, equal the set size in each block. Because of this limit on capacity estimates, it is not sensible to analyze these data across set sizes. Moreover, calculating an average capacity estimate across blocks would result in an artificially low number by including set sizes below capacity. To avoid this limitation, I chose to use each participant’s highest estimate across set size blocks, $K_{\text{max}}$ (for further discussion, see Simmering, 2012; see also Olsson & Poom, 2005; Todd & Marois, 2005). Mean $K_{\text{max}}$ estimates were, respectively, across age groups, 1.90 (95%CI [1.55, 2.25]), 2.20 (95%CI [1.92, 2.48]), 2.90 (95%CI [2.51, 3.28]), and 4.62 (95%CI [4.38, 4.86]) items. I analyzed mean $K_{\text{max}}$ estimates in a one-way ANOVA with Age Group as a between-subjects factor, revealing a significant main effect ($F_{3, 52} = 55.76, \ p < .001, \ \eta^2 = 0.76$). Follow-up Tukey HSD tests ($p < .05$) showed that $K_{\text{max}}$ estimates did not differ between 3- and 4-year-olds, but all other differences were significant.

**Discussion**

Results from the color change detection task were analyzed using accuracy and capacity estimates. Accuracy ($A'$) showed that adults performed at ceiling for all set sizes except five, and children’s performance was below ceiling and above chance for all set sizes. For all groups of children, accuracy decreased significantly from set size two to three; for 5-year-olds, accuracy also
decreased to set sizes four and five. Adults’ performance in set size five was only significantly lower than set sizes two and three. These decreases in accuracy relate to capacity estimates, with significant decreases in accuracy near each age group’s estimated capacity. The key result from this experiment was the increase in maximum capacity estimates over development: children’s capacity increased significantly from approximately two items for 3- to 4-year-olds, to approximately three items for 5-year-olds, and again to between four and five items for adults. In all of these cases, capacity estimates from change detection were lower than the six items estimated from the change-preference task, supporting the difference across tasks predicted by the cognitive dynamics account and model simulations.

EXPERIMENT 3: SHAPE CHANGE DETECTION

A second subset of the participants from Experiment 1 participated in a shape change detection task. The goal of this experiment was to provide a second test of the predicted differences across change-preference and change detection tasks, and to see if the capacity limits found in Experiment 2 are specific to color, or if similar patterns occur in memory for other visual features. Research with adults has shown small but reliable differences across stimulus dimensions; in particular, capacity tends to be slightly lower for shapes than for colors (e.g., Wheeler & Treisman, 2002). Thus, results may be lower in this experiment than in Experiment 2.

Method

Participants

Forty-two children from Experiment 1 also participated in this experiment: 14 three-year-olds (M age = 3 years, 6.39 months; SD = 1.22 months; eight girls, six boys), 14 four-year-olds (Mage = 4 years, 3.93 months; SD = 0.61 months; nine girls, five boys), 14 five-year-olds (Mage = 5 years, 2.21 months; SD = 2.09 months; nine girls, five boys). An additional 10 children were excluded from analyses for the following reasons: seven children did not understand the task (4 three-year-olds, 3 four-year-olds); two chose to end early (3-year-olds); and one experimenter error (3-year-old). Fifteen adults from Experiment 1 also participated in this experiment (M age = 21 years, 4.30 months; SD = 2.90 years; 10 women, 5 men).

Apparatus

The apparatus was identical to that used in Experiment 2 with one exception: rather than colored squares, the stimulus set comprised eight white shapes from Wheeler and Treisman (2002), shown in Figure 17.
Procedure

The procedure was identical to Experiment 2 with two exceptions. First, the memory array was presented for 2 s for all children (rather than only 500 ms for 5-year-olds). Second, all participants completed set size blocks in the same order: SS2, SS1, SS3, SS4, SS5.

Results

As in Experiment 2, participants’ responses were classified as correct rejections, hits, misses, and false alarms. Figure 18 shows the distributions of these response types for each age group separately. As this figure shows, correct responses were most common across ages and set sizes. In addition, performance generally declined as set size increased, especially for children. Relative to Experiment 2, performance is slightly lower, consistent with adult studies comparing memory for shapes and colors (e.g., Wheeler & Treisman, 2002).

For each set size, two measures were computed from participants’ responses: accuracy ($A'$) and capacity estimates ($K$). As in Experiment 2, many 3- and 4-year-olds did not complete SS4 and SS5 due to difficulty keeping children motivated to complete blocks with harder trials. Therefore, I analyzed accuracy across all age groups for SS1–SS3 only, but across all set sizes for 5-year-olds and adults. Again, data from all completed set sizes were used to compute capacity estimates.

**Accuracy ($A'$)**

Accuracy was computed as in Experiment 2. Figure 19 shows mean accuracy scores across set sizes separately for each age group. As can be seen, accuracy was highest for adults and lowest for 3-year-olds, and generally decreased across set size for all age groups. Error bars showing 95% confidence intervals suggest that, even at their lowest accuracy, children were still performing above chance (0.50); adults’ performance was not distinguishable from ceiling (1.0) in SS1 and SS2.

An ANOVA with Set Size (1, 2, 3) as a within-subjects factor and Age Group (3 years, 4 years, 5 years, adult) as a between-subjects factor revealed
FIGURE 18.—Response distributions across set sizes (SS) and trial types for each age group in Experiment 3. Note that \( n = 10 \) for 5-year-olds in SS5; in all other cases, \( n = 14 \). Corr rej, correct rejections; NC, no-change; CH, change.

FIGURE 19.—Mean accuracy scores (\( A' \)) in Experiment 3 across set sizes (SS), separately for each age group; note that chance equals 0.5. Points are offset horizontally to improve visibility. Error bars show 95% confidence intervals.
significant main effects of Set Size ($F_{2, 106} = 91.27, p < .001, \eta_p^2 = 0.63$) and Age Group ($F_{3, 53} = 46.80, p < .001, \eta_p^2 = 0.73$), which were subsumed by a significant Set Size \times Age Group interaction ($F_{6, 106} = 4.31, p = .001, \eta_p^2 = 0.20$). Tests of simple effects for each age group separately showed significant set size main effects for all age groups ($F_{2, 26} = 31.62, 24.16, 33.26, 8.48$, respectively; all $p < .001; \eta_p^2 = 0.71, 0.65, 0.72, 0.38$, respectively). Follow-up Tukey HSD tests ($p < .05$) revealed different patterns across the age groups (see Figure 19). For 3-year-olds, accuracy in SS1 was higher than in SS2 and SS3, but did not differ between SS2 and SS3. For 4-year-olds, accuracy decreased significantly from SS1 to SS2 and from SS2 to SS3. For 5-year-olds and adults, accuracy did not differ between SS1 and SS2, but both were higher than SS3. Thus, the Set Size by Age interaction was driven by different patterns of Set Size effects across the different age groups.

Next, as in Experiment 2, I compared 5-year-olds’ ($n = 10$) and adults’ performance in an ANOVA with Set Size (1, 2, 3, 4, 5) as a within-subjects factor and Age Group (5 years, adults) as a between-subjects factor. This ANOVA also showed significant main effects of Set Size ($F_{4, 92} = 36.18, p < .001, \eta_p^2 = 0.61$) and Age Group ($F_{1, 23} = 77.46, p < .001, \eta_p^2 = 0.77$), which were subsumed by a significant Set Size \times Age Group interaction ($F_{4, 92} = 2.63, p = .039, \eta_p^2 = 0.10$). Tests of simple effects for each age group separately revealed significant set size main effects for both 5-year-olds ($F_{4, 36} = 32.08, p < .001, \eta_p^2 = 0.78$) and adults ($F_{4, 56} = 14.27, p < .001, \eta_p^2 = 0.50$). Follow-up Tukey HSD tests ($p < .05$) showed that 5-year-olds’ accuracy in SS4 and SS5 was lower than in SS1 and SS2, but did not differ from SS3; for adults, accuracy in SS4 and SS5 was lower than all others and did not differ from each other (see Figure 19). Again, the Set Size by Age Group interaction is driven by a different pattern over set sizes for 5-year-olds than adults.

**Capacity Estimates [K]**

Capacity estimates were calculated as in Experiment 3 in each set size for each participant. Mean $K_{\text{max}}$ estimates were, respectively, across age groups, 1.75 (95%CI [1.45, 2.06]), 2.61 (95%CI [1.88, 3.33]), 2.26 (95%CI [1.92, 2.61]), and 3.92 (95%CI [3.57, 4.28]) items. I analyzed mean $K_{\text{max}}$ estimates in a one-way ANOVA with Age Group as a between-subjects factor, revealing a significant main effect ($F_{3, 53} = 16.10, p < .001$). Follow-up Tukey HSD tests ($p < .05$) showed $K_{\text{max}}$ estimates differed among all age groups except 5-year-olds did not differ from 3- and 4-year-olds.

**Discussion**

As in Experiment 2, accuracy scores ($A'$) showed that change detection performance improved over development, although $K_{\text{max}}$ estimates did not
reflect this improvement statistically during childhood. This is likely due to the fact that performance on change trials (i.e., hits) was considerably lower than in Experiment 2. A key finding from this experiment was again the $K_{\text{max}}$ estimates: children’s capacity was estimated around two items between 3 and 5 years, with adults showing significantly higher capacity of approximately four items. In all of these cases, capacity estimates from change detection were lower than the six items estimated from change-preference, supporting the difference across tasks predicted by the cognitive dynamics account.

**COMPARISONS ACROSS TASKS**

Experiments 1–3 confirmed the first two predictions of the cognitive dynamics account of visual working memory capacity: children and adults showed capacity estimates of six items in change-preference, which is higher than infants show in this task, as well as the commonly accepted adult-like capacity of three to four items; these capacity estimates were also higher than the capacity estimates derived from the change detection task for the same individuals (approximately two to three items for children, between four and five items for adults). The third prediction of the cognitive dynamics account is that performance in both tasks depends on the same underlying visual working memory system, which should be evident in correlated performance across tasks.

This prediction was tested in a multiple regression analysis across tasks for all child participants who completed both the change-preference and color change detection tasks. I chose to include only the color version of the change detection task because it is not clear how separable memory is for colors versus shapes. As noted previously, in adults capacity for shapes is often slightly lower than capacity for colors (e.g., Wheeler & Treisman, 2002); the results from Experiments 2 and 3 show this pattern extends into childhood as well (see also Simmering et al., 2015). Thus, for the cleanest comparison across tasks, it seems best to focus only on comparison within a single feature dimension to begin with.

I also chose to exclude adults’ data from these analyses for two reasons. First, there were clear differences between children’s and adults’ performance in Experiments 1 and 2. These relatively large developmental differences may mask more subtle developmental and/or individual differences within the child age groups. Second, 7 of the 12 adults who completed both tasks had $K_{\text{max}}$ estimates at ceiling (i.e., five items), presenting the possibility that this measure was artificially limited for these participants and therefore would not present a complete picture of how performance relates across tasks.
Method

Participants

The 51 children who participated in the change-preference task and successfully completed the color change detection task were included in these analyses: 14 three-year-olds (as reported in Experiment 2), 14 four-year-olds (as reported in Experiment 2), and 23 five-year-olds (13 from Experiment 2 plus 10 from a condition reported in Simmering, 2012, which replicated the method from Riggs et al., 2006 $M$ age = 5 years, 0.37 months, $SD = 0.69$ months; 11 girls, 13 boys). An additional 20 children participated in these tasks but were excluded from analyses: 19 (13 three-year-olds and 7 five-year-olds) for reasons described in Experiment 2, and 1 five-year-old whose $K_{\text{max}}$ estimate was five items. Because it was not possible for $K$ estimates to exceed five in this task, this child’s $K_{\text{max}}$ estimate may have been artificially limited by the task design, leading me to exclude his data from this analysis.

Before considering the results of this analysis, it is important to note potential limitations of the current sample. First, the inclusion of only children who successfully completed both change-preference and change detection tasks was not random. Inclusion in the analysis of Experiment 1 but not here resulted from random assignment for most children (i.e., they were assigned to Experiment 3 or a change detection condition not reported here; $n = 49$ and 34, respectively). Of the 19 children who completed Experiment 1 but were assigned to Experiment 2 and then excluded from analyses of the change detection task, 17 either did not understand the task ($n = 8$) or did not complete enough trials for a capacity estimate to be calculated ($n = 9$). This could result in the current sample including only high-performing children, especially for 3-year-olds. Second, the final sample size for this cross-task comparison is relatively small for addressing individual differences. This is an unfortunate limitation that arises in studies recruiting populations that are not easily accessed (i.e., young children). Both of these limitations make the current analyses more likely to produce a null effect, however, suggesting that any significant findings are worth pursuing for replication with larger, more representative samples.

Variable Selection

The primary measure of interest from change detection is capacity; as such, the $K_{\text{max}}$ estimate for each individual was the dependent measure used in these analyses. In the change-preference task, however, there is no direct calculation of capacity per participant. As described in Chapter 2, simulations of the dynamic model revealed how looking behavior relates to the underlying cognitive processes. Of particular relevance for this analysis are change preference scores, which reflect the detection of novelty, and switch rates, which reflect the detection of familiarity. Because these two measures each have three scores, one derived from each set size, all six scores were included.
in these analyses. Finally, because capacity in change detection was significantly higher for 5-year-olds in Experiment 2, I included age (as a continuous variable) as well. Thus, in this analysis I tested which of the following variables predicted change detection capacity: age; change preference scores in SS2, SS4, and SS6; and switch rates in SS2, SS4, and SS6.

Results

Simple correlations among these factors (Pearson’s $r$) are shown above the diagonal in Table 2 (note that correlations with $K_{\text{max}}$, which are of most interest, are in bold). As expected, $K_{\text{max}}$ estimates from change detection were most strongly positively correlated with age, although age did not correlate significantly with any of the measures from the change-preference task. In addition, positive correlations between $K_{\text{max}}$ and change preference scores from SS2 and switch rate in SS6, were significant, as well a trend toward a correlation with SS4 switch rate. Table 2 also includes partial correlations (below the diagonal; correlations with $K_{\text{max}}$ in bold) controlling for age. Notably, the only effect that changed significance was the correlation with SS4 switch rate, which reached significance with age partialled out.

To test the robustness of the relations among these variables further, I conducted a hierarchical regression analyses with three models including Age, Change Preference Scores, Switch Rates to predict $K_{\text{max}}$. Table 3 shows these three models for comparison. Model 1 included only Age; as expected, it accounted for a significant proportion of the variance in $K_{\text{max}}$ (see Table 3). Next, for Model 2, I entered Change Preference Scores from SS2, SS4, and SS6 to test whether the scores used to estimate capacity correlated across tasks. Similar to the simple correlations shown in Table 2, in this model only SS2

### Table 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.58***</td>
<td>-.08</td>
<td>.02</td>
<td>-.05</td>
<td>.04</td>
<td>-.01</td>
<td>-.06</td>
</tr>
<tr>
<td>1. Capacity ($K_{\text{max}}$)</td>
<td>.40***</td>
<td>-.25*</td>
<td>-.07</td>
<td>-.16</td>
<td>-.02</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>2. SS2 change pref.</td>
<td>-.08</td>
<td>-.25*</td>
<td>.20†</td>
<td>.05</td>
<td>-.07</td>
<td>-.13</td>
<td></td>
</tr>
<tr>
<td>3. SS4 change pref.</td>
<td>-.00</td>
<td>-.07</td>
<td>.21†</td>
<td>.09</td>
<td>-.04</td>
<td>-.16</td>
<td></td>
</tr>
<tr>
<td>4. SS6 change pref.</td>
<td>.15</td>
<td>.17</td>
<td>.04</td>
<td>.09</td>
<td>.44***</td>
<td>.35**</td>
<td>.54***</td>
</tr>
<tr>
<td>5. SS2 switch rate</td>
<td>.25*</td>
<td>-.02</td>
<td>-.07</td>
<td>-.04</td>
<td>.44***</td>
<td>.54***</td>
<td></td>
</tr>
<tr>
<td>6. SS4 switch rate</td>
<td>.43***</td>
<td>.13</td>
<td>-.13</td>
<td>-.16</td>
<td>.35**</td>
<td>.54***</td>
<td></td>
</tr>
</tbody>
</table>

Note. $N=51$. Values above the diagonal indicate bivariate correlations (Pearson’s $r$); values below the diagonal indicate partial correlations corrected for age. Correlations with $K_{\text{max}}$ (the dependent variable of primary interest) are in bold. SS2, set size two; SS4, set size four; SS6, set size six.

†$p<.10$, *$p<.05$, **$p<.01$, ***$p<.001$ (one-tailed).
Change Preference Scores accounted for additional variance beyond Age alone (see Table 3). Further, this model provided a significantly better fit than Model 1. Finally, for Model 3, I added Switch Rates in SS2, SS4, and SS6. In this model, only SS6 Switch Rates accounted for a significant additional proportion of the variance (see Table 3). This model also provided a significantly better fit than the previous models.

These results provide initial evidence that performance is related across the change-preference and change detection tasks, even though the capacity estimates differed across tasks. As an exploratory follow-up analysis, I also tested whether these measures from the change-preference task relate differentially to hits versus correct rejections in the change detection task. As described in Chapter 2, the dynamic model simulations suggest that change preference scores should relate more strongly to hits (i.e., detecting novelty) and switch rates should relate more strongly to correct rejections (i.e., detecting familiarity). However, the standard implementation of the change detection task is not ideal for addressing this question: unlike $K_{\text{max}}$ estimates, which can be calculated for each individual regardless of the number of set sizes completed, rates of hits, and correct rejections are linked to set size (see Figure 15). Furthermore, children completed a relatively small number of trials per trial type in each set size, which could artificially reduce variability across participants. As such, this analysis is a limited approach to testing this question, but results consistent with the cognitive dynamics account would be encouraging for further investigation into how these measures and processes relate across tasks.

For this analysis, I chose to compare children’s performance on highest set size they completed in the change detection task with the significant

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ($\beta$)</td>
<td>.58***</td>
<td>.60***</td>
<td>.62***</td>
</tr>
<tr>
<td>SS2 preference scores ($\beta$)</td>
<td>.33**</td>
<td>.32**</td>
<td>.32**</td>
</tr>
<tr>
<td>SS4 preference scores ($\beta$)</td>
<td>.03</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>SS6 preference scores ($\beta$)</td>
<td>.00</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>SS2 switch rates ($\beta$)</td>
<td></td>
<td>-.08</td>
<td></td>
</tr>
<tr>
<td>SS4 switch rates ($\beta$)</td>
<td></td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>SS6 switch rates ($\beta$)</td>
<td></td>
<td>.31*</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.33***</td>
<td>.40***</td>
<td>.48***</td>
</tr>
<tr>
<td>$\Delta R^2$</td>
<td>.07</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>F change (df)</td>
<td>2.92* (3, 47)</td>
<td>3.24* (3, 42)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Parameter estimates are standardized ($\beta$) and $R^2$ adjusted. F change results are in comparison to the prior model.

*p < .05, **p < .01, ***p < .001.
change-preference measures from the regression analysis above. As a first step, I tested whether hit rates and correct rejection rates from the highest set size completed were related to the $K_{\text{max}}$ estimates used in the hierarchical regression analysis. They were not significantly correlated ($r_{49} = 0.10, p = .23$, for both measures), indicating that these analyses could reveal different relations to the change-preference task than the regression on $K_{\text{max}}$ estimates. Next, I tested the predictions laid out in Chapter 2: SS2 change preference scores were significantly correlated with hit rates ($r_{49} = 0.23, p = .048$) and SS6 switch rates were marginally correlated with correct rejection rates ($r_{49} = 0.22, p = .065$). These exploratory correlations suggest that both the change-preference and change detection tasks can provide separable measures of the processes that support detection of novelty versus familiarity, and that these same processes operate across tasks.

**Discussion**

Using hierarchical regression, I tested whether capacity estimates from the color change detection task were predicted by change preference scores and/or switch rates from the change-preference task. After controlling for age, both set size two change preference scores and set size six switch rates accounted for significant variance in capacity estimates. As predicted by the dynamic model simulations in Chapter 2, these contributions to performance were separable, as set size two change preference scores did not correlate with set size six switch rates (see Table 2) and they explained unique variance in the regression model. Exploratory analyses comparing these measures separately to hit versus correct rejection rates suggest that change preference scores could provide an index of novelty detection (correlated with hit rates) and switch rates could provide an index of familiarity detection (correlated with correct rejection rates). However, a more stringent test of this predicted relation between tasks is needed to determine whether these effects are reliable.

An important remaining question is why the predictive factors in the regression model were limited to only set size two for change preference scores and set size six for switch rates. Model simulations can provide some insight into this question: in the following chapter, I show how the formation and use of memory representations differs across set sizes in the change-preference task, which could lead to behavioral differences that contribute to these correlations. Together these two measures can be considered components of the real-time stability of the memory system, with more stable memory processes supporting better performance in the change detection task.

As noted above, these results should be interpreted with some caution due to limitations in the sample. However, further investigation of the
relation between looking measures and working memory seems warranted based on these findings. The capacity estimates derived across Experiments 1–3 alone would not suggest that such correlations would exist; there were no developmental differences in change preference scores (and therefore capacity estimates) in Experiment 1, even between children and adults. This could have been interpreted as evidence that this task measures a different type of memory than change detection (as had been proposed by Riggs et al., 2006). Instead, by implementing both tasks within a formal model, the cognitive dynamics theory explained how the different patterns of performance across tasks could result from the same underlying system, and that capacity in change detection could be predicted by change-preference performance. Moreover, this comparison across tasks in preschool-aged children suggests that looking tasks may still be a viable method for identifying individual differences beyond infancy.

**SUMMARY**

The experiments presented in this chapter supported the first three predictions of the cognitive dynamics account that were outlined in Chapters 1 and 2. First, Experiment 1 showed that capacity as estimated by the change-preference task continues to develop beyond infancy: 3- to 5-year-old children and adults all demonstrated significant change preferences in set size six. Second, the same participants showed higher capacity estimates in the change-preference versus change detection task, specifically estimated to be at least six items for all age groups in the change-preference task, but only two to three items for children and between four and five items for adults in the change detection tasks. Third, and most critically for the cognitive dynamics account and dynamic neural field model implementation, even though capacity estimates differed across tasks, performance was positively correlated, with set size two change preference scores and set size six switch rates in the change-preference task predicting capacity estimates in color change detection task, even after controlling for age-related increases. These results provide an answer for the question of why previous results seemed inconsistent across these two visual working memory tasks over development, and whether they rely on the same underlying memory system. Furthermore, they align with the cautious interpretation put forth by Ross-Sheehy et al. (2003), in that capacity estimates were not isomorphic across tasks. The next, deeper question to address is what mechanism underlies developmental change in these tasks. The following chapter tests my fourth prediction: whether implementing a variant of the spatial precision hypothesis—strengthening connectivity within the model—can capture the pattern of behavioral results described here.
NOTE

i. One participant (a 5-year-old) responded “same” on every trial in set size five, leading to a hit rate of 0 and a correct rejection rate of 1. For this participant, set size five was also the last set size tested within the random order, suggesting that fatigue may have contributed to this pattern of responding. Thus, I chose to use this participant’s data from set size four in this analysis.
The review of the literature in Chapter 1 identified the need for a general explanation of working memory development. Chapter 2 introduced the cognitive dynamics theory and real-time stability hypothesis as a way to synthesize findings across multiple tasks and developmental period. Implementing this theory in a computational model led to three empirical predictions that were supported by behavioral results in Chapter 3. The current chapter describes tests of the fourth prediction from the model implementation, specifically that developmental improvements in both change-preference and change detection performance could arise from a common underlying source. According to the real-time stability hypothesis, behavior improves developmentally through increasing stability in the underlying visual working memory system. In the type of model introduced in Chapter 2, increasing stability in infants’ memory has been simulated through strengthening connections within and between layers of the model (Perone et al., 2011). Here, I test whether this implementation can be expanded to account for development through early childhood and into adulthood in the change-preference task, as well as the change detection task.
In the sections that follow, I first present quantitative simulations of the dynamic model showing that strengthening connections within the three-layer architecture, combined with changes to accumulation of long-term memory and the task-specific behavioral systems, can fit the pattern of data reported in Chapter 3. Next, I analyzed how the processes within the model differ as a function of this developmental change to provide a further explanation of changes in real-time stability. Then, I use the model’s performance to explain the correlations in Chapter 3, with capacity estimates in color change detection predicted by change preference scores from only set size two and switch rates from only set size six. Finally, in the last section I discuss potential sources of this developmental change in connectivity as it relates to questions of maturation versus experience. This section also considers whether individual and developmental differences could arise from the same sources.

QUANTITATIVE SIMULATIONS OF EXPERIMENTS 1 AND 2 RESULTS

Experiments 1–3 showed that the change-preference and change detection tasks lead to different capacity estimates when testing the same individuals in both tasks. In particular, capacity estimates were higher in the change-preference task (at least six items) and did not change over development from 3 to 5 years or to adulthood. Capacity estimates in change detection ($K_{\text{max}}$), on the other hand, were low for young children (about two items) and increased over development (to about three items for older children to between four and five items for adults). In addition, comparisons across tasks revealed that set size two change preference scores and set size six switch rates in the change-preference task independently and significantly predicted capacity in change detection, even after controlling for variance due to age.

These results support the first three predictions generated with the dynamic model implementation of the cognitive dynamics theory: capacity estimates for older children and adults were greater than four items in the preferential looking task; capacity estimates for the same participants were substantially lower in change detection; and performance was related across tasks despite divergent capacity estimates. This support for the model prediction suggests that these two tasks rely on the same underlying visual working memory system. But can the dynamic model produce this pattern of results across tasks and age groups with a single developmental mechanism? This is an important question because a common source of developmental change from infancy through childhood could shed light onto the processes that relate working memory at these different points in development to higher cognitive skills. This section includes quantitative simulations of both
tasks testing the fourth prediction generated in Chapter 2: development in these tasks can be captured within the dynamic model through strengthening connectivity as an implementation of the real-time stability hypothesis.

Before considering the simulation results in detail, it is important to note that the precise implementation of the model architecture and the parameter values are not the ultimate goal of this exercise. Rather, these simulations are a demonstration proof that many of the cognitive concepts that have been invoked to account for developmental changes in memory, either independently or in opposition, could arise through a shared mechanism. Similar model implementations within the Dynamic Field Theory framework have led to novel behavioral predictions in visuospatial cognition over development (see Johnson & Simmering, 2015; Simmering & Schutte, 2015, for reviews), but such predictions do not depend on specific details of the simulations: there could be minor variations in architecture and/or parameters that still produce the same general outcome. Thus, rather than focusing on whether simulations like these capture behavior to the exclusion of other model implementations, the question of interest is whether the theoretical constructs of cognitive dynamics and real-time stability can explain performance and development in these tasks and, ultimately, whether it could be generalized more broadly (as discussed in Chapter 5). Using a formal model to test this theory provides a level of specificity not often achieved in developmental theories (see Schlesinger & McMurray, 2012; Simmering & Patterson, 2012; Simmering & Spencer, 2008, for further discussion).

Quantitative simulation of the behavioral data from Experiments 1 and 2 is a formidable task, given the different developmental trajectories across tasks (e.g., few differences between age groups in the change-preference task but many differences between age groups in the change detection task). Furthermore, the data set provides a large number of means to fit: in the change-preference task, I fit three behavioral measures (total looking times, change preference scores, switch rates) across three set sizes and four age groups, for a total of 36 means; in the change detection task, I fit percent correct on two trial types (change and no-change trials) in three set sizes for 3-year-olds, four set sizes for 4-year-olds, and five set sizes in 5-year-olds and adults, for a total of 34 means. Thus, varying a relatively small number of parameters in the model (described further below), I was able to fit the dynamic model to 70 behavioral means across tasks and development.

To test whether the dynamic model could fit this pattern of data through strengthening connectivity over development, I began with parameters tuned to fit adults’ performance in other tasks (described further below; Johnson et al., 2014; Lipinski et al., 2010) and simulated adults’ change detection performance. I chose to fit change detection performance first because there was more variance across data points, providing more constraints on the
parameter space that could produce this developmental pattern. Because the dynamic model used by Johnson et al. to fit adults’ performance did not include Hebbian fields, I first modified their parameters to fit the results with adults from the current Experiment 2. Next, I tested the effect of modifying the strength of connectivity over development by scaling a subset of the parameters to capture children’s performance from Experiment 2. Lastly, I held the parameters in the five-layer architecture constant to simulate performance in the change-preference task, and modified a subset of the parameters in the fixation system relative to those reported by Perone et al. (2011) to fit the full data set from Experiment 1. In all cases, parameters were tuned “by hand,” that is, by modifying the values based on previous implementations of the real-time stability hypothesis, then evaluating whether the means resulting from the simulations produced a satisfactory fit across behavioral measures, then iterating these parameters as needed (slightly stronger or slightly weaker) to approximate the behavioral data better. As a starting point, I defined a satisfactory fit as a mean absolute error (MAE) between the behavioral and simulation means, averaging across set sizes within each measure, less than 0.50; this corresponds to one-half of a unit of measure (i.e., trial for change detection, second in looking time, switch per second for switch rate). As results show, most fits surpassed this threshold and provided a closer fit to the behavioral data.

**Simulation Method**

**Model Architecture and Parameters**

The architecture consisted of the five-layer architecture described in Chapter 2 (shown in Figure 6) with the response nodes for the change detection task and the fixation system for the change-preference task (see Appendix for equations). Each of the five fields in the field model contained 361 nodes. Parameter values for the “adult” fits are shown in Table 4 and discussed within the simulation results sections below. Simulations were conducted in Matlab 2013 (Mathworks, Inc.) on a Dell Optiplex 330 with an Intel Core2 Duo 2660 MHz processor. The dynamic field equations were integrated using the Euler method with one time step = 2 ms. For these simulations, the model was presented with one to six inputs on each trial, depending on the set size, out of nine possible inputs that corresponded to different colors selected from a hypothetical continuous 360° color space (i.e., 1° in color space = one unit in the model), in which color stimuli were uniformly distributed with separations of 40°.

**Change Detection Task Structure**

Each simulation trial began with a 100-ms relaxation period that allowed the model to reach a stable resting state. This was followed by the 2-s
<table>
<thead>
<tr>
<th>Field/Nodes</th>
<th>( \tau )</th>
<th>Resting Level</th>
<th>Self Excitation</th>
<th>Excitatory Projection(s)</th>
<th>Inhibitory Projection(s)</th>
<th>Input</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CF (u)</strong></td>
<td>( \tau_u = 80 )</td>
<td>( h_u = -6.75^a )</td>
<td>( c_{ua} = 2 )</td>
<td>( c_{uH} = 0.2^b )</td>
<td>( c_{up} = 1.85 )</td>
<td>( c_{stim} = 50 )</td>
<td>( q_u = 0.04^e )</td>
</tr>
<tr>
<td></td>
<td>( \tau_{uH} = 50^a )</td>
<td>( \sigma_u = 3 )</td>
<td>( \sigma_{uH} = 6.4^b )</td>
<td>( \sigma_{up} = 26 )</td>
<td>( k_u = 0.05^d )</td>
<td>( \sigma_{stim} = 3 )</td>
<td>( q_{uH} = 6^e )</td>
</tr>
<tr>
<td><strong>Inhib (v)</strong></td>
<td>( \tau_v = 10 )</td>
<td>( h_v = -12^a )</td>
<td>( c_{va} = 2 )</td>
<td>( c_{vH} = 1^d )</td>
<td>( c_{vp} = 0.04c )</td>
<td>( \tau_{vH} = 50^a,c )</td>
<td>( q_v = 1^c )</td>
</tr>
<tr>
<td><strong>WM (w)</strong></td>
<td>( \tau_w = 80 )</td>
<td>( h_w = -4.5^a )</td>
<td>( c_{wa} = 3.15 )</td>
<td>( c_{wH} = 1.95 )</td>
<td>( c_{wp} = 0.04c )</td>
<td>( \tau_{wH} = 50^a,c )</td>
<td>( q_w = 1^c )</td>
</tr>
<tr>
<td><strong>Hebb. (H)</strong></td>
<td>( \tau_{Hbuild} = 3000^b )</td>
<td>( \tau_{Hdecay} = 100,000^b )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Resp. (R)</strong></td>
<td>( \tau_r = 80 )</td>
<td>( h_r = -5 )</td>
<td>( c_{rR} = 1.85 )</td>
<td>( c_{rH} = 1.4 )</td>
<td>( c_{rp} = 0.04c )</td>
<td>( \tau_{rH} = 50^a,c )</td>
<td>( q_r = 1^c )</td>
</tr>
<tr>
<td><strong>Same (s)</strong></td>
<td>( \tau_{sH} = 80 )</td>
<td>( h_{sH} = -4.92 )</td>
<td>( c_{sH} = 4.8 )</td>
<td>( c_{sp} = 0.04 )</td>
<td>( \tau_{sH} = 50^a,c )</td>
<td>( q_{sH} = 0.025 )</td>
<td></td>
</tr>
<tr>
<td><strong>Gate (g)</strong></td>
<td>( \tau_g = 80 )</td>
<td>( h_g = -4.92 )</td>
<td>( c_{gH} = 4.8 )</td>
<td>( c_{gp} = 0.04 )</td>
<td>( \tau_{gH} = 50^a,c )</td>
<td>( q_{gH} = 0.025 )</td>
<td></td>
</tr>
<tr>
<td><strong>Fix. (F)</strong></td>
<td>( \tau_f = 80 )</td>
<td>( h_f = -5^c )</td>
<td>( c_{fH} = 4.8 )</td>
<td>( c_{fp} = 0.04 )</td>
<td>( \tau_{fH} = 50^a,c )</td>
<td>( q_{fH} = 0.025 )</td>
<td></td>
</tr>
<tr>
<td><strong>Left / right (l)</strong></td>
<td>( \tau_{lf} = 80 )</td>
<td>( h_{lf} = -5.48^c )</td>
<td>( c_{lfH} = 1 )</td>
<td>( c_{lfp} = 0.04 )</td>
<td>( \tau_{lfH} = 50^a,c )</td>
<td>( q_{lfH} = 0.025 )</td>
<td></td>
</tr>
<tr>
<td><strong>Center (c)</strong></td>
<td>( \tau_{cf} = 80 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Following mathematical convention, parameter subscripts indicate the field receiving the projection followed by the field/node sending the projection (i.e., \( c_{uH} \) indicates the strength of a projection into \( w/WM \) from \( u/CF \)). Parameters shown in the Response (Resp.) row were identical for the Same and Different (Diff.) nodes; any parameters that differed between these nodes are shown in the following rows. Similarly, parameters shown in Fixation (Fix.) row were identical across Left, Right, Center, and Away nodes; any parameters that differed across fixation nodes are shown in the following rows.

- The resting levels of these three fields also included colored noise; the equation for these resting levels, which contains parameters \( q_h \) and \( r_{gh} \), is reported in the Appendix.
- Parameters were identical for both Hebbian (Hebb.) fields (i.e., \( c_{uH} = c_{wH}, \sigma_{uH} = \sigma_{wH}, \tau_{uHbuild} = \tau_{wHbuild}, \tau_{uHdecay} = \tau_{wHdecay} \) and are shown only once for simplicity.
- These parameters were identical across all three fields, and are reported only once in the table for simplicity.
- These projections were applied uniformly across the fields, thus there are no corresponding \( \tau \) values.
- The fixation node equations used dynamic resting levels; the equation for the resting level change, which contains the parameters \( h_f, h_{down}, \) and \( \tau_{fH} \), is reported in the Appendix.
- Noise for the fixation system was applied to the constant input strength to the Left and Right nodes (\( q_{lC}, q_{rC} \)) rather than the node activation at each time step; see Appendix for details.
presentation of the inputs corresponding to the memory array, then a 900-ms delay interval with no input, as in Experiment 2. Next, the inputs corresponding the test array were presented, in which either all items matched the memory array (a no-change trial) or one item changed to a new value (a change trial). As in Experiment 2, the new color value was chosen randomly from the remaining colors, and therefore could not match a color currently in the display. As shown in Table 4, inputs to the model projected strongly into the perceptual field, weakly into the working memory field (i.e., $c_{\text{stim}} > c_{\text{wstim}}$), and to the gate node within the response system; the gate node also received a projection from the working memory field, with parameters tuned such that only the combination of items held in the working memory field and input (as occurs during presentation of the test array) were strong enough to drive activation of the gate node above threshold. Activation of the gate node was passed through a sigmoid function and multiplied by activation projecting from the perceptual and working memory fields to the response nodes (see Appendix for details); as a result, these nodes only received input from the respective fields if the gate node activation was above-threshold ($>0$). This gating mechanism served as an autonomous control that resulted in the model only generating a response once the test array was presented. Because the response nodes were coupled via strong inhibitory connections, only a single node surpassed threshold on each trial (with rare exceptions, as described in the Appendix).

In the behavioral task, the test array disappeared from the computer screen after a response key was pressed. This was approximated in the simulations by setting the end time of the test array presentation to be 250 time steps (500 ms) after the activation of one of the response nodes rose above threshold. To parallel the design of Experiment 2, each “session” for the model included 12 trials (six change, six no-change) in each of the five set sizes, presented in the following order: two, one, three, four, five. Note that 3-year-olds’ performance was not fit for set sizes four and five because no children completed these blocks (as reported in Experiment 2). For 4-year-olds’ performance, I included data from set size four to provide more constraints on the model’s performance, although data from this block were not analyzed in Experiment 2 due to the small sample ($n = 3$). The model was then tested 20 times for each parameter set (comparable to running 20 participants per age group).

Change-Preference Task Structure

As with the change detection task, each simulation trial began with a 100-ms relaxation time that allowed the model to reach a stable resting state. This was followed by a 400-ms presentation of an attention-getter, which projected only to the center node. After the attention-getter ended, a 150-ms delay preceded the beginning of the 20-s trial. At the beginning of each trial,
the illumination of the displays was modeled as a transient (250 ms) “attentional capture” input to the left and right fixation nodes. Through the entire duration of the trial, a constant input (with added white noise; see Table 4) was projected to the left, right, and away fixation nodes, indicating the presence of stimuli other than the attention-getter. Lastly, a transient “onset” input was projected to the left and right nodes for the first 250 ms of each stimulus presentation (i.e., 27 times throughout the trial), signaling the presence of the color stimuli on the displays. These inputs follow those used by Perone et al. (2011) and are described further in the Appendix.

When activation of either the left or right fixation node surpassed threshold (>0), the specific input from that display—that is, the colors that were presented on the corresponding display—was projected to the model (strongly into the perceptual field and weakly into the working memory field, as in the change detection task; see Table 4). Input streams for each display were constructed following the details of Experiment 1: each consisted of two, four, or six colors (depending on set size) that were on for 500 ms with 250 ms blank intervals in between. For the no-change display, the colors did not change over the course of the 20-s trial; for the change display, one of the colors, selected randomly, would be replaced with another color that was not currently present in that display. Note that, as in Experiment 1, it was possible for colors to be repeated across displays; because spatial location is not represented in this architecture, such colors would be represented with identical inputs to the fields. For each “session,” the model was presented with six 20-s trials, two at each set size, in random order. The model was tested 40 times for each parameter set, reflecting the larger sample size from this experiment.

**Simulation Results and Discussion**

In this section, I describe the results of the simulations in the change detection task first, followed by the result in the change-preference task. Within each subsection, I discuss how I arrived at the parameter to capture adults’ performance, then children’s performance and report the closeness of the fit in MAE between the behavioral and computational means. This section concludes with an evaluation of the fits across tasks.

**Change Detection Task Performance**

To evaluate the fit of the model simulations to the behavioral data from Experiment 2 (plus data from ten 5-year-olds who completed a similar color change detection task reported by Simmering, 2012, and the cross-task comparison above), I chose to compare the number of correct trials, separately for no-change and change trials (i.e., correct rejections and hits); because the other measures of performance (accuracy and capacity estimates)
are computed from these distributions, the fits would be similar across measures. Furthermore, considering the fit separately for each trial type provides more constraints than fitting a combined measure, such as accuracy, in which similar scores can result from different contributions of each trial type (e.g., an $A'$ score of 0.60 would be derived from a hit rate of 0.67 and false alarm rate of 0.17, as well as a hit rate of 0.83 and false alarm rate of 0.33).

Figure 20 shows the mean number of correct trials for the change detection simulations for each parameter set (fitting the different age groups), with behavioral data for comparison. As this figure shows, the final simulations captured the relative rates of errors on each trial type. Moreover, the simulations yielded mean maximum capacity estimates similar to the behavioral estimates: 1.99 items with the 3-year parameters (1.90 in Experiment 2), 2.83 with the 4-year parameters (2.20), 2.93 with the 5-year parameters (2.87 combining Experiment 2 with the 5-year-olds from Simmering, 2012), and 4.50 with the adult parameters (4.62).

To arrive at these results, I began with the parameters from Johnson et al. (2014) for the three-layer architecture (i.e., the perceptual, inhibitory, and
working memory fields) and response nodes; for the Hebbian fields, I used parameters tuned for adults’ spatial recall performance (Lipinski et al., 2010), as no previous fits of change detection performance have incorporated the Hebbian fields. These parameters provided a good fit to adults’ performance on no-change trials across set sizes (MAE = 0.10 trials), but an inferior fit of adults’ performance on change trials across set sizes (MAE = 0.67), with the model performing worse than adults at higher set sizes. This deficit in the model’s performance can be characterized as a bias to respond “same” too often as more items were held in the working memory field. Within the full model architecture, there are multiple parameters that could be modified to decrease this bias; the simplest way was to adjust the response system to be less likely to respond “same.” I did this by slightly decreasing the resting level of the “same” node \(h_{ss}\) from \(-4.35\) (the value from Johnson et al., 2014) to \(-4.75\) (see Table 4). With this minor adjustment, the fit of the model’s performance on no-change trials remained excellent (MAE = 0.05), and the fit on change trials improved to a satisfactory level (MAE = 0.20; overall MAE = 0.13). These fits are shown in Figure 20A, with the related parameters shown in Table 4.

The next step in my quantitative fits was to simulate the developmental improvements in change detection performance between 3 and 5 years and into adulthood. To this end, I modified the adult parameters according to the real-time stability hypothesis, specifically decreasing the strength of excitatory and inhibitory connections to capture children’s poorer performance. Given the small difference in performance between 3 and 5 years (relative to the change between 5 years and adulthood), I began by simulating 3-year-olds’ performance, then adjusted the parameters proportionally stronger as needed to approximate older children’s performance. To achieve this fit, I made three types of changes to the three-layer architecture, two changes to the Hebbian fields, and five changes to the response system; the resulting parameter values are shown in Table 5. Parameters within the three-layer architecture were changed following previous fits of spatial recall performance over development (e.g., Schutte & Spencer, 2009). First, relative to the adult parameters, inputs were weaker and more diffuse (smaller \(c_{stim}\), larger \(\sigma_{stim}\)) in the child parameters. Second, the local excitatory (\(c_{uu}, c_{ww}\) and inhibitory (lateral, \(c_{uv}\) and \(c_{wv}\), global \(k_{uv}\) and \(k_{wv}\)) connections were weaker (smaller) for the “child” parameters. Third, the child parameters included stronger noise relative to the adult parameters (larger \(q_u/q_v/q_w\)).

In addition to these established changes within the three-layer architecture, changes were needed in the Hebbian fields and response system to fit children’s performance. The Hebbian fields were changed to build more quickly and decay more slowly (smaller \(\tau_{build}\), larger \(\tau_{decay}\)) relative to the adult parameters, to reflect the greater influence of long-term memory...
<table>
<thead>
<tr>
<th>Field/Node</th>
<th>$\tau$</th>
<th>$\gamma uu = 1.7$</th>
<th>$\gamma wu = 1.26$</th>
<th>$\gamma wu = 1.4175$</th>
<th>$\gamma wu = 1.6695$</th>
<th>$\gamma wv = 1.1788$</th>
<th>$\gamma wv = 1.885$</th>
<th>$\gamma wv = 0.2015$</th>
<th>$\gamma wv = 0.06$</th>
<th>$\gamma wv = 0.06$</th>
<th>$\gamma wv = 0.06$</th>
<th>$\gamma wv = 0.06$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF (u)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WM (w)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hebb. (H)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resp. (R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff. (d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same (s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gate (g)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix. (F)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left (l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right (r)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Only parameters that were changed relative to the adult values are shown in this table; all other values were identical to those reported in Table 4. Numbers at the end of each subscript indicate the corresponding age group in cases where the value changed between parameters tuned to 3 and 5 years. As in Table 4, parameters that were identical across Same and Different (Diff.) nodes are shown in the Response (Resp.) row; parameters that were identical across Left, Right, Center, and Away nodes are shown in the Fixation (Fix.) row.

*This parameter was identical across all three fields, and are reported only once in the table for simplicity.*

*As in Table 4, parameters were identical for both Hebbian (Hebb.) fields (i.e., $\tau uHbuild = \tau wHbuild$, $\tau uHdecay = \tau wHdecay$) and are shown only once for simplicity.*
on children’s performance (indicated by reports from experimenters, e.g., that children sometimes tried to compare a new memory array to the previous test array). The response system was modified in parallel to the three-layer architecture, with weaker inputs (smaller $c_{du}$, $c_{sw}$), weaker self-excitation (smaller $c_{rr}$) and competition (smaller $c_r$), and stronger noise (larger $q_r$) relative to adult parameters. Two additional changes to the response system were driven by evaluation of simulation fits. First, to compensate for weaker excitation in the working memory field with the child parameters, I increased the strength of the projection from the working memory field to the gating node (larger $c_{gw}$) relative to the adult parameters. Second, the resting level of the “same” node was higher (smaller $h_is$; recall that resting levels are negative) for the child versus adult parameters, reflecting children’s overall bias to respond “same” more often.

With these changes, the model provided an excellent fit of 3-year-olds’ performance on both no-change and change trials (MAE = 0.10 and 0.22, respectively; overall MAE = 0.16). These fits are shown in Figure 20B, and provide the first demonstration that changes in real-time stability could account for developmental differences in change detection performance. Next, to fit the model to 4- and 5-year-olds’ performance, I made four changes relative to the 3-year-old parameters. Specifically, I increased the strength of the local excitatory and lateral inhibitory connections for the working memory field (larger $c_{ww}$ and $c_{wv}$), the input (larger $c_{ustim}$ and $c_{wstim}$), and the projection to the “different” node (larger $c_{du}$); additionally, I decreased the build rate of the Hebbian fields (larger $\tau_{build}$). The fits for 4- and 5-year-olds’ performance, shown in Figure 20C and D, were also quite good (4 years: MAE = 0.16, 0.21, and 0.19, respectively, for no-change trials, change trials, and overall; 5 years: MAE = 0.35, 0.19, and 0.27, respectively).

**Change-Preference Task Performance**

To evaluate the fit of the model simulations to behavioral data from Experiment 1, I compared total looking times, change preference scores, and switch rates across set sizes two, four, and six for each age group. To fit the model’s performance in the change-preference task, I began with the four parameter sets for the five-layer architecture that were tuned to fit change detection performance across the four age groups, then tested performance with the fixation system parameters tuned to capture infants’ performance by Perone et al. (2011). With these “infant” fixation parameters (averaging across all set sizes and age groups), the model provided a good fit of change preference scores (MAE = .024; equivalent to a difference of 0.36 s looking to the change display), an acceptable fit of switch rates (MAE = 0.45 switches per second), but an inferior fit of total looking time (MAE = 2.76 s). For all parameter sets, the model produced less total looking than in the behavioral data. This poor fit suggest that the developmental changes in the fields were
not enough to produce the changes in looking behavior; rather, the dynamics of fixation appear to change over development as well.

To achieve a better fit to the behavioral data from the change-preference task, I held the parameters within the fields constant (i.e., as tuned for change detection performance for each age group) and scaled the infant fixation parameters in three ways. First, in parallel to the changes within the fields, I increased the strength of self-excitation and mutual inhibition ($c_{ff}$ and $c_{cf}$) for all parameter sets relative to the infant parameters. Second, I changed the relative influences of inputs to the fixation nodes by increasing the strength of the constant input from the displays ($c_{fC}$) and decreasing the projection from the perceptual field ($c_{fu}$); conceptually, this made the fixation system more responsive to the task context and less driven by novelty compared to the infant parameters. Lastly, I modified the dynamic resting level to make each of the fixation nodes more likely to transition between “on” and “off” states (through a lower negative attractor value, $h_{down}$; see Appendix for description of these dynamics). Table 5 shows these parameter values. Note that local excitation was equivalent across the four parameter sets, whereas the other parameters were all lower for the child versus adult simulations; additionally, only the low attractor varied across the three child parameter sets.

With these changes, the fits improved for total looking (MAE = 0.64, 0.74, 0.57, and 0.32 across parameter sets, respectively), change preference scores (MAE = .017, .010, .011, and .026, respectively), and switch rates (MAE = .08, .05, .06, and .03, respectively). Figure 21 shows these behavioral measures and the model fits with parameter values tuned for each age group. As this figure shows, fits for mean total looking time (Figure 21A) were satisfactory, although the model showed slightly more of an increase in looking over set sizes than seen in the behavioral data. Change preference scores (Figure 21B) were also fit well by the model across set sizes and age groups, and all of the mean change preference scores from the simulations were significantly greater than chance (dotted line). Lastly, the pattern of switch rates across set sizes—more switching in set size two—was fit well by the model (Figure 21C).

**Evaluation**

This chapter presented quantitative simulations to test my fourth prediction, that implementing the real-time stability hypothesis in the unified model could capture developmental changes in both change detection and change-preference tasks. To test this hypothesis, I conducted a series of simulations to capture the full range of behaviors across tasks and development. Across the two tasks, I compared 70 means between the simulations and behavioral data (34 in change detection, 36 in change-preference) and achieved an excellent fit overall, with the model performing well within a single unit of measure of the behavioral data (i.e., averaging across age groups: MAE = 0.19 trials in change detection; in the
change-preference task, MAE = 0.57 s total looking, 0.016 preference score, and 0.06 switches per second).

These simulations provide support for the cognitive dynamics account of capacity estimates arising from both change detection and change-preference tasks, as well as the proposed developmental mechanism, increases in realtime stability through strengthening connectivity. The simulations presented here are a particularly stringent test for a computational model—few models are tested across multiple tasks, multiple measures within a task, and multiple

Figure 21.—Simulation fits to behavioral data (behav.) across set sizes (SS) from Experiment 1: (A) mean total looking time, (B) mean change preference scores, (C) mean switches between displays. Error bars show 95% confidence intervals. Dashed line in (B) indicates chance performance.
points in development. As such, the simulations presented here move beyond many cognitive and developmental models in providing both specificity—producing a good fit to behavior on a trial-by-trial basis within each task—and generality—across tasks and development (see also Simmering & Spencer, 2008).

CONSEQUENCES OF STRENGTHENING CONNECTIVITY: INCREASES IN REAL-TIME STABILITY

One benefit of using formal models to evaluate theoretical constructs is the ability to “look inside” the model to understand the underlying causes of changes in the model’s performance. In this section, I present an analysis of changes in the dynamic model’s encoding and maintenance of items to illustrate how strengthening connectivity within the model gave rise to developmental improvements in performance through increases in real-time stability. In the following sub-sections, I consider developmental changes in seven components of real-time stability in turn: (i) strength of representations; (ii) encoding speed; (iii) resistance to interference and (iv) decay; (v) capacity limits; (vi) correspondence between memory and behavior across tasks; and (vii) effects of task context on memory.

Evaluating the formation and maintenance of memory representations requires considering how the representations change over time. To assess representations over time, I plotted the timecourse of above-threshold activation within the working memory field during the presentation of the memory array (encoding) and delay (maintenance) during the change detection task. I chose to limit this assessment to the context of the change detection task because the timing of events puts more of a demand on memory processes and can therefore reveal differences in real-time stability more clearly than the change-preference task (which also provides more support from long-term memory).

Figure 22 shows the timecourse of activation during encoding in the left column and during maintenance in the right column. Consider the encoding period first (recalling that the processes within the model do not differ between “stages” of the task; all that differentiates encoding from maintenance is the presence vs. absence of input). Figure 22A shows the rise time of activation for the adult parameters during encoding. As this figure shows, total activation generally increased with set size, but not linearly—that is, the amount of activation in set size three was not three times more than in set size one. As discussed by Johnson et al. (2014), this characteristic of the dynamic model contrasts with resource accounts of capacity limits: if activation is considered the resource that is shared among items, the amount allocated to each item does decrease as the number of items increases;
FIGURE 22.—Above-threshold activation over time in the working memory field during the change detection memory array presentation (left column) and delay period (right column), separately for each set size (SS) in simulations with (A) adult, (B) 3-year-old, (C) 4-year-old, and (D) 5-year-old parameter sets, averaged across trials within a run.
however, the total amount of resource (activation) across items is not fixed (i.e., total activation for set size five was twice that of set size one).

With weaker connectivity corresponding to earlier in development, four changes in the dynamic model’s memory representations are evident. First, overall activation increased considerably over development, for all set sizes (see scaling of $y$-axes in Figure 22); note that this applies to both the encoding and maintenance phases of the trials. A second, related consequence is that the child parameters showed less of a difference in activation across set sizes. This is partly due to the lower number of peaks being formed (discussed below), but the mean activation per peak was also lower with the child parameters (see set size one in Figure 22). These differences are due to weaker individual representations (i.e., peaks) earlier in development. As such, total activation illustrates the first component of real-time stability, stronger representations over development, that arose through strengthening connectivity in the dynamic model.

A third notable difference across parameter sets is that, although the slope of the initial rise in activation appears comparable across parameter sets in Figure 22, the different amount of activation reached indicates that the child parameters built activation more slowly. For instance, consider the time to reach activation of 30 in set size one across parameter sets: 475 ms with the 3-year-old parameters, 360 ms with the 4-year-old parameters, 295 ms with the 5-year-old parameters, and 125 ms with the adult parameters. Thus, the adult parameters built activation nearly four times as quickly as the 3-year-old parameters, demonstrating a second component of real-time stability: developmental increase in encoding speed resulting from strengthening connectivity within the dynamic model.

Lastly, activation was less stable over time during the presentation of the memory array for the child parameters compared to the adult parameters. This is most strikingly demonstrated with the 3-year-old parameters (Figure 22B), which showed substantial fluctuation at the higher set sizes. The variation in activation can be attributed to the set size nearing the upper limit for the number of peaks that could build simultaneously with these parameters (discussed further below), even with continued input. This variation results from competition (or interference) between items being encoded in the working memory field: in set size one, activation was stable over time for all parameter sets; only when multiple items were being encoded at once did activation fluctuate notably over time. These simulations demonstrate a third component of real-time stability, reduced interference among items over development.

I next examined how above-threshold activation in the working memory field changed after the stimulus input was removed as peaks sustained through the memory delay, plotted in the right column of Figure 22. With the adult parameters (Figure 22A), activation decreased slightly after the removal
of the input, then remained at a fairly constant rate as the self-sustaining peaks were maintained through the delay. Similar to the strength differences noted during encoding, the child parameters also resulted in less activation during the delay period relative to the adult parameters. Also, the variability in activation over time was greater with the child parameters, reflecting reduced real-time stability as peaks “die out” during the memory delay (cf. Figure 23B–D). Thus, the relatively small fluctuations over delay with increasingly strong connectivity illustrate increased resistance to decay over development.1

For the next assessment of the model’s performance, I analyzed the number of peaks held in the working memory field at the end of the delay periods across tasks. I first calculated the mean number of peaks for each response type in the change detection task (i.e., correct rejections, hits, misses, false alarms) for each parameter set, shown in Figure 23. Note that the simulations with the 3- and 4-year-old parameters include set sizes four and five; even though there were no behavioral data for comparison at these set

![Figure 23](https://example.com/figure23.png)

**Figure 23.**—Mean number of peaks sustaining in the working memory field at the end of the change detection delay period, separately for each parameter set (simulating different age groups). Lines show means across all trials, bars show means separated by response type. Error bars show 95% confidence intervals. SS, set size; corr rej, correct rejections.
sizes, it is useful to consider the model’s performance on these trials to see the full effects of developmental changes in the parameters. As Figure 23 shows, the adult parameters maintained an average of 4.99 peaks in the working memory field on all set size five trials except the small number of trials that resulted in false alarms (on which the mean was 4.50; similar to results from Johnson et al., 2014). By contrast, fewer peaks were maintained in the working memory field with the child parameters, for example, reaching only 3.23 peaks on average across trials in set size five with the 3-year-old parameters. The number of peaks held in the working memory field at the lower set sizes also shows more variability with the child parameters relative to the adult parameters. This analysis demonstrates that capacity increases over development with increased real-time stability.

An important consideration for understanding the correspondence between memory and behavior in the change detection task is how the number of peaks in the working memory field differed across trial types. In particular, all parameter sets held fewer peaks on false alarm trials than on other trial types, suggesting that these errors result from “forgetting,” that is, the loss of a peak in the working memory field (cf. Johnson et al., 2014). Sample simulations in Chapter 2 (Figures 9 and 11) demonstrated how this arises in the model. This type of error makes intuitive sense: a forgotten item would appear to be new, when presented in the test array.

Unlike the adult parameters, for the child parameters, the number of peaks in the working memory field was also lower on hit trials than miss trials, suggesting that correct “different” responses in the model sometimes arose when at least one item was forgotten during the delay, and miss errors could occur when all items were held in the working memory field (cf. Figure 8 in Chapter 2). Thus, although capacity was indeed lower with the child parameters—the model was able to hold fewer items in the working memory field than with the adult parameters—this difference is not the sole contribution to errors in the model (as discussed in Chapter 2; see also Johnson et al., 2014, for further discussion). Rather, more errors must arise through the processes of comparison and response within the child parameters. Consider, for example, that all parameter sets maintained the item presented in set size one on every trial (i.e., all means were 1.00 across response types). However, performance was below ceiling in this set size for the child parameters (cf. Figure 20). Thus, the same number of peaks in the working memory field did not produce the same level of performance, highlighting another component of real-time stability: improved correspondence memory and behavior over development.

For the last analysis of the model’s representations, I calculated the number of peaks in the working memory field during each delay period in the change-preference task. Because there is not a single delay period for maintenance in this task, it is less straightforward to assess how many peaks
the model could hold within a trial. To address how capacity limits are realized in the change-preference task, I plotted the mean number of peaks held in the working memory field during the first delay period (Figure 24A), the final delay period (Figure 24B), and averaged across all delay periods (Figure 24C).

Figure 24A shows that the parameter sets varied little in the number of peaks maintained during the first delay period in set size two, but in set sizes four and six the “older” parameter sets held more peaks in the working memory field. This difference can be attributed not only to the faster encoding with stronger connectivity (cf. Figure 22), but also to the fixation dynamics: with stronger interactions within the fixation system, the model with the adult parameters was more likely to look at both displays during the first presentation. As a result, the model received more input with the adult parameters. The consequences of increased inputs can also be seen by comparing Figure 24A with B; during the final delay period, the model had received input across many stimulus presentations (27 per trial) and fixations (the “adult” model averaged 15.95 fixations per trial; cf. $M = 15.30$ fixations per trial in the behavioral data). As a result, the number of peaks in the
working memory field was more similar across set sizes during the final delay period compared to the first delay period.

When comparing parameter sets, however, the capacity limits on the total number of peaks that can sustain in the working memory field are evident. Consider, for example, the adult parameters in set size six: although all nine stimulus colors were present across displays in each presentation (as described in Experiment 1), the model still maintained fewer than six peaks on average ($M = 5.58$) by the final delay period. This converges with simulations by Johnson et al. (2014) suggesting that a similar dynamic model has an upper capacity limit of six items, and shows that even the more supportive context of the change-preference task cannot infinitely increase the capacity of the working memory field in the model. By contrast, however, the peaks held in the working memory field with the 3-year-old parameters in set size six show a divergent pattern from the change detection simulations above. In change detection, these parameters averaged 3.23 peaks in set size five; the same parameters held an average of 4.38 peaks by the final delay period in the change-preference task. This demonstrates how the task context influenced the contents of the working memory field in the model: with repeated presentations and short delays, the child parameters could maintain, on average, 1.15 more peaks simultaneously during the change-preference task. By analogy to the behavioral task, this would imply that children’s capacity was actually higher—remembering more items during the delay—in the change-preference task than the change detection task. The large difference in capacity limits across tasks early in development, combined with the relatively small difference in capacity limits later in development, point to another developmental change that is not often a focus of theories: the dynamic model’s memory representations were less affected by the task context over development.

Lastly, note that the average number of peaks held across all delay periods (Figure 24C) was only at or above the number of items presented per display in some cases. Specifically, in set size two means ranged from 2.22 to 4.65 peaks over parameter sets, and in set size four the adult parameters held a mean of 5.52 peaks. In all other cases, the model was maintaining fewer items than the set size, but still showed a robust preference (Figure 21); these results converge with Perone et al.’s (2011) findings that a similar model with infant parameters held fewer peaks in the working memory field than suggested by capacity estimates (see McMurray et al., 2012, for simulations suggesting a similar divergence between vocabulary and fast mapping performance). This also contrasts with analyses of performance in the change detection task, which showed that the mean number of peaks in the working memory field was higher than the estimated capacity. These findings across tasks highlight the importance of linking representations to behavior when assessing constructs like capacity—even when similar numbers of items were held in
the working memory field across tasks, the behavioral results yielded notably different estimates of capacity. The discrepancy across tasks decreased over development, however, providing a further example of improved correspondence between memory and behavior.

These simulations showed that the dynamic model can capture developmental improvements in change detection performance by implementing the real-time stability hypothesis as strengthening connectivity over development. Analyses of the processes underlying the model’s performance provided evidence for the following consequences that comprise the notion of increasing real-time stability over development: stronger representations; faster encoding speed; reduced interference among items; increased resistance to decay; higher capacity; improved correspondence memory and behavior; and less effect of the task context on behavior. Thus, the real-time stability hypothesis not only provides a specific mechanism underlying developmental change—strengthening connections supporting encoding, maintenance, and comparison processes—but also specifies the process by which these changes influence performance in two different task contexts. As discussed further in Chapter 5, these characteristics of the real-time stability hypothesis converge with empirical findings and other theoretical accounts proposed to explain working memory development during infancy and developmental changes in complex tasks during childhood.

These results also provide an important qualification to the possibility Ross-Sheehy et al. (2003) raised, that the change-preference task might assess “relative” capacity while the change detection task assesses “absolute” capacity. One could conclude from the current behavioral results that the tasks indeed measure different types of capacity, as estimates differed notably across tasks. However, the correlations in Chapter 3 suggest shared underlying processes, and can be used to argue against dissociating the behavioral tasks. By taking a process-oriented approach and formalizing the theory of cognitive dynamics into a computational model, these simulations showed that capacity differences emerged in the context of the tasks; that is, the same cognitive system will be limited in different ways depending on the task context. Simulations showed that such differences arose simply through the stimulus timing, manner of presentation (i.e., on one vs. two displays), and behavioral responses required in the tasks. In the literature, these differences between tasks, as well as the different task instructions (i.e., no explicit instructions in the change-preference task vs. specifically reporting changes in change detection), have led theorists to propose that the tasks recruit different strategies or memory processes. For example, Riggs et al. (2006) suggested that looking tasks might rely on passive memory, and Oakes et al. (2013) questioned whether habituation or long-term memory contributed to preferences. The formalization presented here shows that invoking different processes to explain diverging results is not necessary; rather, such findings can be understood by
considering representations in the context of the processes through which representations are created and used in the service of behavior (cf. McMurray et al., 2012; Schöner & Thelen, 2006; Smith et al., 2010, 1999; Spencer, Clearfield, et al., 2006).

RELATING PERFORMANCE ACROSS TASKS

The next important question to consider is whether these simulations can help explain the correlations across tasks described in Chapter 3. The rationale for the correspondence between change preference scores and hits in change detection (detecting novelty), and between switch rates and correct rejections in change detection (detecting familiarity), was laid out in Chapter 2. However, the sample simulations did not lead to a priori predictions for differences across set sizes. Recall the empirical findings from the cross-task comparison: capacity estimates in color change detection were predicted independently by (a) change preference scores from set size two and (b) switch rates from set size six (see Table 3). What would lead to these differences across set sizes in the cross-task relationships in children’s performance? I consider these two measures in turn below.

As illustrated in Figure 24C, the child parameters maintained an average of 1.52 more peaks in set size six than set size two. How would the number of peaks in the working memory field affect change preference scores? As explained in Chapter 2, continued fixation of the change display depends on activation in the perceptual field. More peaks in the working memory field contribute to stronger inhibition projected into the perceptual field, resulting in lower activation. Thus, as more items are being maintained, the magnitude of the change preferences should be reduced. Restated conceptually, the more items that are being held in memory, the less novel the change display appears. This could explain why only change preference scores set size two—with the smallest number of peaks and therefore greatest activation (i.e., novelty) in the perceptual field—were predictive. This also could explain the main effect of set size from the change-preference task, in which change preference scores were significantly higher in set size two than four and six.

The effect of the number of peaks in the working memory field on switch rates would be opposite the effect on change preference scores because switching depends on a decrease of activation in the perceptual field. In particular, as more items are encoded in the working memory field, more inhibition projects to the perceptual field via the inhibitory field, and fixation is released more quickly. Thus, switch rates should increase as the number of items in memory increases over time. Although this would seem to imply that switch rates should be higher for set size six than two (opposite the empirical and computational results), it is important to note that the model’s switch
rates depend on the rate of increase in the number of peak over time within the trial, not just the overall number of peaks. The rate at which peaks increase is greater for set size two than six (cf. Figure 24A and B), leading to the higher switch rate in the lower set size.

Essentially, change preference scores and switch rates depend on more versus less activation in the perceptual field, which depends on inhibition projected into the perceptual field, which is stronger when more items are held in memory. This mechanism also explains why switch rates differed over development: the “adult” parameters held more peaks than the “child” parameters, and adults switched significantly more often than children. Why, then, did change preference scores not differ between age groups? In the context of this particular task, there were only nine possible colors that could be presented across both displays within the trial. As such, on a set size two trial the colors in the change display would begin to repeat after, at maximum, eight presentations (perhaps earlier—how soon the same color could reappear was not constrained); on a set size six trial, the colors in the change display would begin to repeat after four presentations at most. Furthermore, duplication of colors across displays was not controlled, meaning a new item on the change display could be present within the no-change displays. Indeed, in set size six, a minimum of three colors had to be present in both displays, as the total number of colors presented across displays on a single set size six trial (12) was greater than the number of unique colors used (nine). Together, these details of the task limit how “new” a changed color can be; the accumulation of long-term memory of the colors within a trial, or even the sustained working memory for one or more presentations, would lead to a general suppression of the effect of novel colors on fixation of the change display. This could also explain why the effect of set size on change preference scores was relatively small ($\eta_p^2 = .03$).

The results of these simulations shed new light onto how memory representations form and influence behavior in the change-preference task. This level of formularization has revealed gaps in previous conceptual theories of performance in this task. In particular, Oakes et al. (2009) claimed that long-term memory could not easily be used to detect changes in the change-preference task due to the brief presentations and the frequent changes. However, as these simulations and those by Perone et al. (2011) showed, a robust memory for the no-change display is just as important for a change preference to emerge. Because the colors on this display remain the same over the duration of the 20-s trial, long-term memory can indirectly support a robust change-preference (but see Oakes et al., 2013, for a task that eliminates this characteristic, as discussed in Chapter 5). Furthermore, the original conceptualization of the task (Ross-Sheehy et al., 2003) focused on how items compared within a single display over time, but not across displays. These simulations suggest that duplications of colors across displays could
help support memory representations and, by extension, a robust change preference. In the current formalization, however, there is no representation of which colors were on which display—it is possible that adding such information would change this characteristic of the model’s performance.

POTENTIAL SOURCES OF DEVELOPMENTAL CHANGES IN CONNECTIVITY

The final question to consider in the context of these model simulations is what could cause the strengthening of connectivity that produces increases in real-time stability. Within the larger Dynamic Field Theory framework, most developmental simulations have included parameters modified “by hand” to capture different developmental levels (see Simmering & Schutte, 2015, for review), as was the case for the current simulations. However, in a recent series of simulation experiments, Perone and Spencer (2013a) showed how the accumulation of specific experiences over a sufficiently long timecourse approximated changes that have previously been implemented through strengthening connectivity. In particular, they exposed a number of “infant” models to multiple series of visual inputs to simulate real-world interactions with objects distributed through time (the equivalent of more than 500 hr of experience per “infant”). Importantly, these models included Hebbian layers similar to those in the current model, which allowed experience to build over these extended exposures.

Perone and Spencer (2013a) evaluated the effects of this accumulation of experience by testing the models in a processing speed task modeled after Rose, Feldman, and Jankowski (2002). They showed that allowing the model to develop autonomously through experience could produce the same types of outcomes as simulations in which development was implemented “by hand” (e.g., Perone & Spencer, 2013b). In particular, models with more experience exhibited a more mature pattern of performance, including stronger novelty preferences and more frequent gaze switching. Similar to the simulations presented here, models with more experience (i.e., “older” models) showed stronger representations and faster encoding speeds. Thus, Hebbian learning across a long range of specific experiences could result in general improvements in performance in a dynamic neural field model.

Perone and Spencer (2013a) also showed that this mechanism could address individual differences across “infant” models. They noted that individual differences across infants often parallel group level developmental differences (also discussed in Chapter 1). To test whether stable individual differences could emerge through the accumulation of experience, they conducted a second simulation experiment in which they assessed performance at three points during development for the individual models. Results showed that these models reproduced the same general pattern as in their
first simulation experiment when tested as a group, and also revealed stable differences across models. Specifically, hierarchical regression showed that the models’ performance on the processing speed task “early” in development (approximating 7 months of age) predicted their performance “later” in development (approximating 12 months of age). These simulations provided the first demonstration that models set to the same initial conditions could autonomously develop stable individual differences in visual memory performance.

How do the results from Perone and Spencer (2013a) relate to the empirical and computational results presented here? The simulations presented in this chapter included no sense of “individuals,” in that each separate run of the model within a given age group had the same parameter values. Individual runs of the model varied slightly due to noise, and accumulated different experiences within the context of each task, but these would not be stable across tasks. Although the simulations by Perone and Spencer show a promising interpretation of individual differences, such an approach to the current empirical results is not practical due to the computational power required: Perone and Spencer reported that each individual model’s development took more than 8 hr to simulate just to arrive at 12 months of experience. Simulating 3–5 years of experience for even only 10 individuals per age group, therefore, would require an unreasonable time commitment. Furthermore, this demonstration of how individual and developmental differences could arise through a common source does not necessarily imply that they must arise in this manner.

A more promising route to explore the possible explanatory power of such a mechanism conceptually is to test the effects of experience behaviorally. Perone and colleagues have shown one such application in the context of the dimensional change card sort task (Perone, Molitar, Buss, Spencer, & Samuelson, 2015). They tested 3-year-old children in the standard version of the task, in which sorting was tested first along a shape dimension, then a color dimension, and found typical rates of perseveration (i.e., many children continued to sort by shape on the color trials). They compared performance in this condition to a “memory” condition, in which children played a memory matching game with colored stimuli between the pre- and postswitch card sort trials. Given this extra experience along the color dimension, children in the memory condition were significantly less likely to sort perseveratively. That is, exposure to the colors in a memory game allowed children to show the more mature pattern of performance, correctly sorting by color on the postswitch trials. Perone et al. showed model simulations demonstrating how the accumulation of long-term memory of the colors supported performance in the card sorting task.

Even with these computational and behavioral demonstrations that relevant experiences can support more mature behavior, there remains the
possibility of experience-general maturational improvements underlying behavior. These sources of change are not mutually exclusive and likely both play a role. Various studies have shown that there are changes in multiple brain regions over development that might produce effects similar to the changes in connectivity implemented in the dynamic model here. As an example, consider development of the prefrontal cortex, which plays a central role in a variety of working memory tasks (e.g., Crone, Wendelken, Donohue, van Leijenhorst, & Bunge, 2006; Diamond & Goldman-Rakic, 1989; Luciana, Lindeke, Georgieff, Mills, & Nelson, 1999). The prefrontal cortex continues to develop throughout childhood and adolescence (e.g., Huttenlocher, 1979, 1990, 1994; Jernigan, Trauner, Hesselink, & Tallal, 1991), suggesting that its development plays a likely role in developmental changes in working memory capacity. Myelination in the frontal lobe continues through early childhood (Sampaio & Truwit, 2001), and reductions in gray matter volume reflecting pruning of neural connections occurs though early childhood into the postadolescent years (Gogtay et al., 2004), in parallel with the developmental trajectory seen in the change detection task (Figure 2).

These types of neurophysiological changes during childhood relating to working memory improvements are consistent with changes in the dynamic model connectivity over development. For example, Edin and colleagues generated predictions for developmental changes in BOLD signal during a visuospatial working memory task based on the consequences of synaptic pruning, synaptic strengthening, or myelination in a neural network model (Edin, Macoveanu, Olesen, Tegner, & Klingberg, 2007). They compared the model’s predicted BOLD activity to fMRI results from frontal and parietal regions with 13-year-olds and adults. The best correspondence between the model and neural data resulted from synaptic strengthening in the model—a mechanism likely to arise through experience similar to the implementation by Perone and Spencer (2013a).

The Representational Acuity Hypothesis proposed by Westermann and Maeschal (2004, 2012) also shows similar changes in processing and behavior. According to their hypothesis, developmental narrowing of receptive fields in visual cortex results in object representations that become more precise over development (i.e., fewer neurons respond to a given stimulus). This narrowing of receptive fields and resulting increase in precision is conceptually similar to increases in real-time stability described here (see also discussion of the resolution of visual working memory in Chapter 5). Their hypothesis has been implemented in a neural network model, which has captured developmental differences in infants’ performance in categorization tasks (Westermann & Maeschal, 2004, 2012). These results converge with simulations within the Dynamic Field Theory framework across a range of tasks (described further in Chapter 5) to suggest that these types of neural changes could explain a wide range of behavioral phenomena over development.
Although evidence of these neural changes are promising for our understanding of development, this simply shifts the developmental question to another level of description, leaving the question open as to what drives the changes in cortex. These cortical changes likely arise through complex interactions among a host of factors from the genetic level to the level of large-scale interactions among populations of neurons in different cortical areas, which are probably experience-dependent. Given that the changes made to the dynamic model in these simulations involved primarily changes in the strength of excitatory and inhibitory connections, it is easy to imagine that such changes could arise from a simple Hebbian process that strengthens cortical connections as a function of experience (cf. Perone & Spencer, 2013a).

SUMMARY

This chapter included quantitative simulations of the dynamic model showing that strengthening connections within the three-layer architecture, combined with changes to accumulation of long-term memory and the task-specific behavioral systems, could fit the pattern of behavioral data reported in Chapter 3. Following these successful fits, I analyzed the processes within the model to show how changes in connectivity gave rise to seven characteristics of real-time stability, which contributed to changes in performance over development. Further analysis of the model’s performance provided a potential explanation for the correlations reported in Chapter 3, with the number of items held in memory showing opposite effects on change preference scores versus switch rates across set sizes. Lastly, I described simulations in a similar architecture that suggest individual and developmental differences in performance could arise through a common source, namely extended experience with instances along a given dimension (i.e., color), and a recent empirical test that supports this possibility (see also Simmering et al., 2015 but see Cowan, Ricker, Clark, Hinrichs, & Glass, 2015, for evidence that specific knowledge cannot entirely explain developmental improvements).

NOTE

i. Note that the common sense of decay—a gradual loss of information over time—does not happen at the level of individual representations within the dynamic model. Rather, the “on” and “off” states of this type of bistable attractor network produce all-or-none representations: a peak is either held in the working memory field or not (see Zhang & Luck, 2009, for related behavioral evidence in visual working memory). The strength of the peak may decrease over time and the location of the peak may drift (i.e., change position within the field; see, e.g., Lipinski et al., 2010); these changes often arise due to interactions with other peaks and are therefore considered more to interference than decay.
The goal of this monograph was to advance a process-oriented theory of visual working memory development that can span multiple tasks and developmental periods. As reviewed in Chapter 1, lab-based measures of working memory correlate both concurrently and longitudinally with measures of more general cognitive skills, highlighting the importance of understanding how working memory functions and develops. However, there currently exists a disconnect between research and theories of working memory during infancy and later childhood. As a bridge between these disconnected research areas, the current monograph focused on two tasks considered to measure visual working memory capacity, the change-preference task developed for infants, and the change detection task used with children and adults. Previous studies using these methods revealed seemingly inconsistent estimates of capacity over development, raising the questions of whether the tasks rely on the same underlying processes and whether a common developmental change could explain improvements in both tasks. Here I presented the cognitive dynamics theory, which emphasizes understanding the link between memory
representations to behavior across tasks contexts and development, as a way to reconcile previous behavioral results and theoretical accounts.

Chapter 2 introduced a modeling framework, the Dynamic Field Theory, to formalize the concepts embodied within the cognitive dynamics theory into a computational model. Using a dynamic neural field model implementation, I showed how these two behavioral tasks could be performed by the same cognitive system. Simulations of these tasks led to three specific behavioral predictions and one simulation prediction. First, simulations suggested that capacity estimates from the change-preference task should continue to increase over development beyond Ross-Sheehy et al.’s (2003) estimate of four items at 10 months. Second, the model’s performance indicated that capacity estimates from the change-preference task should be higher than estimates from change detection within the same participants. Third, despite the predicted difference in capacity estimates across tasks, performance should be correlated across tasks because both tasks rely on the same underlying visual working memory system. In particular, model simulations demonstrated how change preference scores and switch rates in the change-preference task could provide indexes of the underlying memory processes, suggesting that both could relate to capacity estimates from change detection. Finally, I predicted that the full pattern of performance across tasks and development could be simulated using a single developmental mechanism—an increase in real-time stability—realized through strengthening connections in the dynamic model.

Chapter 3 presented three behavioral experiments confirming the first three predictions. In Experiment 1, 3- to 5-year-old children and adults completed the change-preference task from Ross-Sheehy et al. (2003) in set sizes two, four, and six. Change-preference scores were significantly above chance for all age groups at all set sizes, suggesting a capacity of at least six items by 3 years of age. Experiments 2 and 3 tested these same participants in the change detection task, in which capacity was estimated between two and three items for children and between four and five items for adults. Thus, as predicted, capacity estimates in the change-preference task increased beyond infancy, and were higher than estimates from change detection within the same group of participants. Furthermore, comparing performance across tasks revealed capacity estimates from change detection were significantly and independently predicted by change preference scores from set size two and switch rates from set size six in the change-preference task, even after controlling for age-related increases.

Chapter 4 presented a test of my fourth prediction, that the suite of behavioral effects across tasks and development could be quantitatively fit by the dynamic model using strengthening connectivity as the primary mechanism of change. These simulations provided a good fit to the empirical data, and analyses of the processes embodied in the model revealed seven consequences of strengthening connectivity, which comprise the notion of
real-time stability in the cognitive dynamics account. Further analysis of the model’s performance across set sizes revealed why the correlations in Chapter 3 were limited to the lowest set size for change preference scores, but the highest set size for switch rates. These simulations show the utility of formalizing theories with computational models, as they can provide insights into potential underlying mechanisms and reveal gaps in previous conceptual explanations of performance and development.

In this final chapter, I first describe remaining questions and challenges for this account of visual working memory. The next section includes discussion of possible generalizations of this perspective beyond visual working memory development, first to the theories and empirical findings reviewed in Chapter 1, and second across a broader range of domains.

REMAINING QUESTIONS AND CHALLENGES IN VISUAL WORKING MEMORY

The dynamic model implementation of the cognitive dynamics theory has answered some questions within visual working memory, addressing how capacity estimates relate to behavior in the change-preference and change detection tasks over development. In this section, I discuss four additional considerations from the literature on capacity limits: (i) a new eye-tracking measure of infants’ capacity; (ii) the resolution of memory representations; (iii) the representation of multiple visual features in “bound” objects; and (iv) the role of attention in visual working memory performance.

A New Eye-Tracking Measure of Capacity in Infancy

Oakes and colleagues developed an infant “one-shot” change detection task to probe capacity without the extended trials used in the change-preference task. In the one-shot task, a single memory array of two colors was presented for 517 s, followed by a 317-ms delay, then a 3,000-ms test array in which one color remained the same and one color changed (Oakes et al., 2013). Oakes et al. (2013) tracked 6- and 8-month-old infants’ eye movements during the presentation of the test array and found that 8-month-olds exhibited a preference for the changed color, while 6-month-olds looked equally to both colors. In a follow-up experiment, they presented memory arrays with two identical colors (essentially requiring the encoding of only a single item), and 6-month-olds exhibited a preference for the changed color at test. These findings converge with results from the change preference task, suggesting that capacity increases from one item at 6 months to two or three items at 8 months (Oakes et al., 2006, 2009; Ross-Sheehy et al., 2003).

How do these results align with the cognitive dynamics account of performance in the change-preference task? Although the dynamics of looking
between two objects within a single display likely differ from the dynamics of looking between two displays that may contain multiple objects, the same notion of familiarity reducing fixation and novelty maintaining fixation would hold. Two phenomena from Oakes et al. (2013) need to be explained: the difference in performance between 6 and 8 months, and the difference in 6-month-olds’ performance when the memory array contained one versus two colors. For the first finding, the mechanism from Perone et al. (2011) is applicable: the “older” infant parameters in their model formed memory representations more quickly, could hold more than one item in memory at once, and also used memory representations in service of behavior more stably.

Even if most 8-month-old infants were only able to stably encode a single item in memory, this may be sufficient to drive a preference for the new item at test. Assuming that infants would not be biased during encoding, about half of the infants would encode the item on the left, and half would encode the item on the right (for the sake of this example, presume the item on the right is the one that changed in the test display). At test, infants who encoded the left item would see one familiar item and one novel item; infants who encoded the right item would see two novel items. Presumably the left-encoding infants would prefer the novel item, but the right-encoding infants could prefer either (randomly) or show no preference. Depending on the strength of each infant’s preference, these different patterns could average together as a significant preference for the changed item at the group level. The data presented by Oakes et al. (2013) were not intended to address individual differences, but this possibility could presumably be tested by analyzing individual infants’ looking over time; fixations during the presentation of the memory array may also shed light onto which color(s) individual infants were likely to encode.

Understanding 6-month-olds’ performance in the second experiment, with two identical colors in the memory array, follows directly from the previous example. In this version of the paradigm, only one color could be encoded during the memory array, and that memory was sufficient to drive a preference for the new color in the test array. In the context of the first experiment, though, which color each 6-month-old encoded would have presumably been random. Could the difference between 6 and 8 months still arise if each age group only encoded a single item in memory? Perhaps: as the simulations in Chapter 4 demonstrated, other changes within the visual working memory system could lead to different levels of performance even with the same memory representations. For example, if 8-month-old infants encode more quickly than 6-month-old infants, the younger infants may need more opportunities to encode a single color, which would occur as they looked between two identical items in the memory array from Experiment 2. This possibility could be tested by extending the duration of the memory
array when it contained two different colors, to see if preferences then
emerged in younger infants. Similar to the simulations presented in
Chapters 2 and 4, this analysis of the one-shot task highlights the importance
of considering not just what is remembered but the processes supporting the
formation and use of memory representations within the specific task
context.

Resolution of Visual Working Memory

The nature of visual working memory capacity limits has been a source of
considerable debate within the adult literature (e.g., Bays & Husain, 2008;
Wilken & Ma, 2004; Zhang & Luck, 2008; see Fukuda, Awh, & Vogel, 2010, for
review). Part of this debate focuses on the resolution or precision of visual
working memory representations. This question has more recently been
addressed with developmental studies. In particular, Simmering and
Patterson (2012) showed that strengthening connectivity within a related
dynamic model led to increased resolution of visual working memory, as
measured by color discrimination performance. In the model, this difference
in resolution did not arise through the width of peaks in the working memory
field, as might be suggested by the characterization of visual working memory
“precision” in the literature (e.g., Burnett Heyes et al., 2012); rather, it was the
strength of the inhibitory troughs projected to the perceptual field that
determined how reliably colors could be identified as “different.” Simmering
and Patterson showed that children’s ability to discriminate very similar colors
separated by a short delay improved between 4 and 6 years, in parallel to the
increases in capacity shown in the current sample, and as predicted by
strengthening connectivity the dynamic model. Following this demonstra-
tion, Simmering and Miller (2016) tested whether these changes in capacity
and resolution were related over development. Their study showed that 5-
to 8-year-old children’s capacity in change detection was predicted by their
discrimination performance, but not by age. These empirical and computa-
tional results indicate an eighth characteristic of increasing real-time stability
not address in Chapter 4: improved resolution of visual working memory over
development.

A related question that has not yet been addressed within this
framework, however, is whether the resolution per item changes as more
items are held in memory. Behavioral evidence shows a trade-off between
resolution and capacity, with the resolution per item decreasing as the
number of items held in memory increases; such findings have been used to
support a “flexible resource” perspective (e.g., Bays & Husain, 2008; Burnett
Heyes et al., 2012). More recent evidence suggests that changes in
resolution asymptote once a maximum number of items (typically three to
four) has been reached; these results support a “discrete resource” account
of capacity limits (see Fukuda et al., 2010, for review). From this perspective, each individual has a limited number of “slots” in which to store items, but differing amounts of a shared resource could be devoted to each item, allowing for differences in resolution among items with a fixed upper limit (Fukuda et al., 2010).

To evaluate how resolution might vary with the number of items held in the dynamic model, it is important to consider how precision or resolution is probed behaviorally (see below for discussion of electrophysiological measures of attention). First, Simmering and Patterson’s (2012) color discrimination paradigm used participants’ ability to detect small changes in colors as the metric of resolution (detecting smaller changes implies higher resolution). As noted in the description of their results above, their model simulations suggested that this effect did not arise through more precise representations (i.e., peak width), but through the associated inhibition that supported the comparison processes needed to detect a change. The simulations in Chapter 4 showed that, as the number of items in the working memory field increases, more inhibition projects to the perceptual field, which makes changes more difficult to detect overall. This leads to the prediction that color discrimination performance should be worse when more items are held in memory, not due to the width of peaks, but due to a reduced ability to detect small changes. Thus, the behavioral estimate of resolution would differ across set sizes without a corresponding change in the underlying representations.

The second type of task used to assess precision or resolution is a recall task, in which participants reproduce a continuous stimulus feature (i.e., color or orientation of a line) from memory, and precision is calculated from the variability in error magnitude across trials. This metric of precision is similar to a common measure in motor control and spatial recall, referred to as variable error in those domains (i.e., the standard deviation of errors across responses to the same target; e.g., Spencer & Hund, 2002). Because similar three-layer model architectures have been used to simulate spatial recall performance, it is possible to extrapolate from those simulation results to interpret precision in color recall. Variable error from behavioral data has been fit through quantitative simulations (Schutte & Spencer, 2009), but does not solely reflect the width of representations (i.e., peaks in the working memory field). Rather, some variation in responding arises through noise in the maintenance process over the memory delay. As such, the behavioral index of “precision” may reflect underlying variability in the real-time processes supporting maintenance, not just the nature of the representation.

Although these spatial recall simulations provide some insight into potential contributions to variable error other than representational precision, they have not addressed the implications for remembering
multiple items at once. Previous simulations of the change detection task, however, have shown that local excitatory and lateral inhibitory connections that underlie encoding and maintenance in the dynamic model allow for metric interactions among items held in working memory simultaneously. In particular, Johnson et al. (2014) showed more similarity of items within the memory array increased the likelihood of losing at item from memory (i.e., a peak “dying out” in the model) and more false alarm errors (in both the model and behavioral data); furthermore, more similarity between the new item on a change trial and other items from the memory array was related to higher miss rates.

How would these interactions among peaks within working memory affect variable error in a color recall task? Johnson (2008; see also Johnson, Ambrose, Dineva, & Spencer, 2016) showed that an uneven distribution of inhibition around a peak—which would arise from a second nearby item held in memory at the same time—caused bias in the peak away from the neighboring item (i.e., repulsion in memory; this effect is well documented between a single item held in memory and perceptual reference frames in spatial recall, both behaviorally and computationally, Spencer, Simmering, Schutte, & Schöner, 2007). Johnson demonstrated that systematically presenting two “near” colors and one “far” color together on each color recall trial led to predictable bias in the direction of recall error (i.e., showing repulsion from nearby colors), providing behavioral evidence for this mechanism of interactions among items in visual working memory. In most color recall tasks the selection of colors is random across trials, which would produce variability in whether another item was nearby or not, and the “direction” of the neighboring item in color-space when one was nearby. Thus, extrapolating from Johnson’s findings, the different directions and magnitudes of the type of repulsion based on the (randomly-selected) distribution of target colors would lead to greater variable error, which in turn would produce a lower calculated precision score.

This section shows how theories that only address the nature of representations may not be sufficiently specific to evaluate against alternative explanations. Only through an analysis of how representations would produce the relevant behavior can we uncover other potential contributions to performance and errors. It remains to be tested whether the types of mechanisms described here—the role of inhibition in the ability to detect changes, or interactions among items in memory producing variability—could account for the full range of behavioral results, or if additional changes in the nature of the representations is also necessary. This highlights a central tenet of the cognitive dynamics theory: the way in which memory is used in service of behavior is equally important to understand as the nature of memory representations.
A second common question with regard to capacity limits is whether multiple features can be bound into so-called “integrated objects,” which are then the units of capacity, or if different features of objects are represented additively, such that one object with two features would require the same capacity as two objects with one feature. Behavioral evidence suggests that neither of these extremes captures adults’ performance in change detection tasks: in some conditions, multiple features can be remembered with no apparent cost (e.g., Luck & Vogel, 1997; Vogel et al., 2001), but in other conditions performance is moderately impaired for multi-versus single-feature items (e.g., Cowan, Blume, & Saults, 2013; Wheeler & Treisman, 2002). There is also some debate over whether binding requires additional processes, such as attention (Johnson, Hollingworth, & Luck, 2008). Furthermore, adults’ performance on recall tasks shows some independence between features (i.e., one feature may be forgotten when the other feature is remembered; Bays, Wu, & Husain, 2011) but also some effect of remembering other features (i.e., resolution for each feature is lower when remembering two features compared to single-feature objects; Fougnie, Asplund, & Marois, 2010). The full pattern of results across tasks and conditions has not yet been explained by a single, comprehensive theory; rather, contradictory results are often presented to refute a theory, but without a thorough description of the source of discrepancies across conditions. This is a promising future avenue for computational work, which can provide the necessary specificity to test which types of representations and processes could capture the details of performance across variations in task parameters.

Two of the developmental theories described in Chapters 1 and 2 have implications for relating features within visual working memory. First, Oakes and colleagues (2006, 2009) proposed that infants’ ability to retain multiple items within an array arises through the emerging ability to bind colors to locations. Second, the neural synchrony model from Raffone and Wolters (2001), cited by Riggs et al. (2011) as a potential mechanism for capacity increases, was originally proposed to account for the similarity in capacity limits between single- and multi-featured objects (e.g., Luck & Vogel, 1997). As the simulations in Chapter 4 demonstrated, such binding was not necessary to explain developmental increases in capacity in the change-preference and change detection tasks when only colors are tested. However, to account for performance when correspondence between color and location is required (e.g., Oakes et al., 2006), or with multiple features per object (e.g., Riggs et al., 2011), the processes involved in binding must be specified.

The processes necessary for binding features and space were not part of the dynamic model presented in Chapters 2 and 4, but initial steps toward this level of representation have been taken within the broader Dynamic Field
Theory framework by Schneegans, Spencer, and Schöner (2015). In their model, separate feature dimensions (e.g., color, shape, space) are represented in parallel four-layer architectures (similar to the three-layer architecture described here, but with the processes capture in the current perceptual field split across an attentional field and a contrast field; see also Schneegans, Lins, & Spencer, 2015). These features are then linked via two-dimensional feature-space fields that receive projections from the one-dimensional fields. For example, the two-dimensional color-space working memory field would receive projections from the color working memory field along one dimension and the space working memory field along the other dimension; where the activation from these fields intersects would specify which color was present at which location. In this model, multiple features of an object (e.g., color and shape) are not linked directly, but rather both features are linked to the shared spatial location. Mis-binding of features is possible when multiple objects are present, but this is limited through attentional processes (see Schneegans, Spencer, et al., 2015, for simulations of the circumstances under which mis-bindings are likely to occur).

The most stringent test of feature binding comes through change detection tasks that test binding changes (i.e., when two features swap between objects) versus changes to new features (i.e., a new color or shape on one object). Schneegans, Spencer, et al. (2015) demonstrated that the expanded Dynamic Field Theory architecture can detect binding changes—the swap in position of one feature between two objects—by virtue of the new feature-location binding. Rather than a new feature value driving a peak in the perceptual field, as shown in the simulations in Chapters 2 and 4, binding changes are detected in this model’s feature-space fields. One challenge for this mechanism would be identifying such changes when an object is presented in a new location (e.g., Cowan et al., 2013); this type of probe could potentially be addressed with an updating mechanism such as that used in a dynamic neural field model of multi-object tracking (Spencer, Barich, Goldberg, & Perone, 2012). Thus far, this expanded model architecture has been used to demonstrate that binding changes can be detected through this manner of representation, but has not been tested quantitatively to see if it could produce the patterns of results shown in behavioral tasks. As such, it is not yet clear how this approach compares to existing explanations of feature binding in adults.

What are the implications of developmental increases in real-time stability for feature binding? Although the types of changes from Chapter 4 have not been implemented in this expanded Dynamic Field Theory architecture, a likely consequence of weaker connections within and between fields would be an increased likelihood of mis-binding features and/or forgetting features as peaks die out during the delay. Moreover, slower encoding would make it more difficult to stabilize peaks in the
two-dimensional fields, as these fields require activation building first in the one-dimensional fields before projecting activation to the two-dimensional fields. Behavioral evidence shows that children as young as 3 years perform similarly on trials with single- versus multi-featured objects when testing only changes to new features; however, when performance was compared between trials requiring the detection of binding changes versus detection of new features, young children show binding impairments (Simmering, Johnson, Patterson, & Spencer, 2009; Simmering & Wood, 2016). Further behavioral and computational tests are needed to evaluate whether there are developmental changes in feature binding beyond the changes in visual working memory for single-feature objects.

Relating Attention and Visual Working Memory

A third area of interest is the role of attention in visual working memory capacity limits. Effects of attention are difficult to summarize due to notable differences in definitions and distinctions for multiple types of attention (see Anderson, 2011, for discussion). In this section, I consider three types of evidence that attention relates to visual working memory capacity: exogenous attentional cues can augment infants’ looking performance; developmental and individual differences in infants’ attention predict measures of working memory both concurrently and longitudinally; and individual differences in adults’ control of attention correlates with capacity.

One study with infants has shown the effect of exogenous attention on infants’ visual working memory performance. As reviewed in previous chapters, 6-month-old infants show no preference when tested on set size three trials in the change-preference task (Oakes et al., 2006, 2009; Ross-Sheehy et al., 2003). However, Ross-Sheehy, Oakes, and Luck (2010) showed that 5-month-old infants look longer to a set size three change display if attention is drawn to the item that is changing. This version of the paradigm compared looking time to changing streams (comparable to the change display in the current study) with attentional cues that indicated the changed color (valid) versus one of the unchanged colors (invalid) across trials. When the cue was rotation of the colored square, 5-month-old infants looked significantly longer to the valid versus invalid stream, suggesting that this cue supported memory for the items within the display (see Ross-Sheehy et al., 2010, for discussion of a different manipulation that did not augment infants’ performance).

How could this type of manipulation be implemented in the dynamic model used here? Motion presumably serves as a spatial cue to direct the infant’s gaze in the experiment, but the dynamic model does not include a representation of space within each display. Using the kind of architecture described in the previous section on feature binding, where space and color
each provide inputs to a color-space field, a stronger spatial cue (through motion) would increase the overall input to the working memory field at the location of the cued item. This would, in turn, increase the likelihood of encoding the color of the item at that location in the working memory field, which would make the change easier to detect. Thus, exogenous attentional cues could be simulated as additional spatially specific input to the model.

A second notion of attention that is commonly discussed in the infant literature refers to the dynamics of looking behavior. Rose and colleagues quantify attention through switch rates in looking tasks and describe attention as an important component of looking behavior that predicts later cognitive functions (Rose, Feldman, & Jankowski, 2004; Rose et al., 2012). As shown in Chapter 4, the dynamic model architecture can produce different rates of switching through changes in the fixation system dynamics as well as in the underlying memory representations that contribute to looking behavior. The fixation dynamics in the model could be interpreted as attention, specifically, the parameter that influenced the fluctuation in the fixation nodes’ resting levels based on their activation (i.e., $h_{\text{down}}$ in Tables 4 and 5). In the simulations presented in Chapter 4, this parameter varied across age groups, but could also be set to vary between “individuals” (i.e., runs of the model) within an age group.

Simulations by Perone and Spencer (2014) suggests that this level of individual variation in processes may not be necessary, however. They simulated infants’ performance in a visual paired comparison task that included objects that varied in both colors and shapes. Similar to infants’ performance, analyses of their simulation results showed that the model’s looking during familiarization—measured by switch rate, look duration, and peak look length—predicted whether the model would prefer the familiar or novel item at test. Importantly, Perone and Spencer used identical parameter values for each separate run of the model through the task, meaning that the “individual” differences arose through stochastic noise and the history-dependent nature of looking performance in the model (i.e., where the model looked determined the input to the model). These simulations, combined with the analyses of the dynamic model’s performance in Chapter 4, suggest that the individual differences that Rose et al. (2004, 2012) have described as attention may depend on processes of encoding and comparing in visual working memory.

Lastly, studies with adults have shown that individual differences in attentional control correlate with capacity in the change detection task. For example, Vogel, McCollough, and Machizawa (2005) found that participants’ capacity was predicted by their ability to filter out irrelevant stimuli from the memory array. To assess filtering, they used ERPs to measure contra-lateral delay activity, which had previously been shown to scale with the number of
items held in memory (Vogel & Machizawa, 2004). Participants completed a change detection task in which memory arrays could contain two or four items; on some of the set size four trials, participants were instructed to ignore half of the items though a cue at the beginning of each trial. Filtering efficiency was calculated by comparing contra-lateral delay activity on the trials when two items were to be ignored to the other trial types: activity comparable to the set size two trials indicated good filtering, whereas activity comparable to the set size four trials indicated poor filtering. Results showed that filtering efficiency differed between high- and low-capacity individuals (defined by a median split), and that capacity estimates correlated with filtering efficiency. These results suggest that the ability to control what information gets into working memory contributes to individual differences in capacity limits.

Similarly, Fukuda and Vogel (2009) showed that individual differences in resistance to attentional capture (also measured through ERPs) correlated with capacity estimates from change detection. Their measure of attentional capture compared evoked responses to probes at target versus distractor locations in a cued target identification task. In particular, participants were instructed to attend to one of two cues (signaled by corresponding colors in different locations) to encode the identity of a shape presented later at the cued location. Across trials, either the valid or invalid location could be presented; the difference in amplitude of the ERP wave form between these two trial types was the attentional capture (i.e., more similarity corresponds to more capture). Thus, participants with strong evoked responses for the valid locations, and weak evoked responses for the invalid locations, showed the best resistance to attentional capture, which correlated with capacity.

Could this sense of attention be realized in a dynamic field model? In its current form, the model described in Chapters 2 and 4 had no mechanism explicitly serving as attention in the change detection task, and also no individual differences. It is possible that modulation of the existing parameters could simulate individual differences in capacity. For example, Buss and Spencer (2014) used a dynamic neural field architecture similar to the feature-binding model described above to capture children’s performance in the dimensional change card sort task. In this task, children are told to sort cards by one feature (i.e., color or shape) in a first block of trials, then by the other feature in a second block of trials. Buss and Spencer simulated these instructions by boosting the resting level of the field associated with the correct feature (i.e., the color-space field on color trials and the shape-space field on shape trials); using this mechanism, they simulated results from 14 conditions of the card sort task. Exploratory simulations of the adult parameters reported in Chapter 4 showed that this type of boost to the resting levels in the three-layer architecture modulated capacity estimates, with
higher estimates when the perceptual field or inhibitory layer had a higher resting level, but lower estimates when the working memory field had a higher level. It is possible that different combinations of resting levels across “individuals” could produce reliable differences in capacity.

Individual differences in capacity could also be simulated in the dynamic model by modifying the parameters that were strengthened over development in Chapter 4. Note that even without any sense of “individuals” in the model (i.e., each run of the model within a given “age group” included identical parameter values), the quantitative fits presented in Chapter 4 provided a satisfactory fit of variability as well as mean performance (see error bars in Figure 20; MAE for standard deviations ranged from 0.22 to 0.27 trials across age groups). Given the number of ways individual differences could be implemented in the dynamic model architecture, the question of whether the dynamic model could simulate individual differences is somewhat trivial. The more critical question is whether the neural indexes of attention described by Vogel et al. (2005) and Fukuda and Vogel (2009) could be measured analogously in the model. At present, there is no clear answer to this question. Although models within the Dynamic Field Theory framework build from neural population models (Amari, 1977), and other models within the framework have been used to predict ERP differences across conditions (e.g., McDowell, Jeka, Schöner, & Hatfield, 2002), the models used to simulate visual working memory performance are abstractions that do not map directly to neural areas (see Johnson, Spencer, & Schöner, 2009, for discussion). Furthermore, activation within a given field in the model may not be analogous neural activity; rather, activation projected from one field to another might be a more suitable comparison. More work needs to be done before this question can be addressed, and there are currently efforts to connect the Dynamic Field Theory framework more directly to neural measures (e.g., Spencer, Buss, & Magnotta, 2013; see also Buss, Magnotta, Schöner, & Spencer, 2016).

**GENERALIZING THE COGNITIVE DYNAMICS THEORY AND REAL-TIME STABILITY HYPOTHESIS**

Although the cognitive dynamics theory has led to novel insights into behavior and empirical predictions within research on visual working memory development, as described above, one of the primary motivations for this project was to build toward a general theoretical perspective that can be applied across a broad range of domains. This section considers two potential extensions, first to the types of working memory and development that were described in Chapter 1, then to cognitive development research more generally.
Chapter 1 reviewed behavioral phenomena and theoretical accounts from two domains of working memory research: visual working memory during infancy and children’s performance on complex working memory tasks. One of the goals of the current empirical and computational work was to advance a process-based theory of working memory development that could be generalized to these two domains. To this end, the current section reviews the major findings from these domains and address how the cognitive dynamics theory and real-time stability hypothesis align with this part of the literature.

Recall from Chapter 1 that infant working memory has been assessed using tasks that require delayed reaching or looking to different stimuli over time. A number of the developmental changes in infants’ performance in looking paradigms have already been modeled using dynamic neural fields, demonstrating that these effects could arise through increasing real-time stability: faster encoding (Perone & Spencer, 2013a,b, 2014; see also Chapter 4); stronger novelty preferences (including the early familiarity-to-novelty shift; Perone & Spencer, 2013a,b, 2014); more looks and switches between displays (Perone & Spencer, 2013a, 2014; see also Chapter 4). Both looking and reaching paradigms have also shown increases in the durability of memory, with older infants retaining information for longer delays. Some of these effects are driven by long-term memory (e.g., recognizing an object from a previous lab visit; Courage & Howe, 2001), whereas others depend more directly on working memory (e.g., reaching correctly after a longer delay; Diamond, 1985). Although neither of these effects has been modeled directly, the simulations from Chapter 4 can provide some insight. In particular, the developmental changes implemented in the dynamic model led to fewer “forgotten” items (i.e., peaks dropping out during maintenance; see Figures 22 and 23) over development, which would correspond to better durability if memory was probed after varying lengths of delay. The increased durability in long-term memory could arise through increases in strength and durability of working memory representations as well: in the dynamic model presented here, long-term memory is accumulated as a trace of the above-threshold activation from the working memory field. Thus, as items are represented more strongly and held longer in working memory (as shown in Chapter 4), the corresponding long-term memory representations would also be stronger. Together these simulations of various looking paradigms across infancy and early childhood suggest that the major developmental improvements noted during infancy—faster encoding, increased durability, stronger novelty preferences, and more switching—could all arise through a common underlying mechanism.

In addition to these developmental changes during infancy, Chapter 1 reviewed evidence from complex span tasks showing developmental
improvements and correlations with higher cognitive skills. Recall that these tasks either require manipulation of information in memory (e.g., repeating a series of items backward) or switching between storage and processing tasks (e.g., counting sets of dots and storing the results from the series for later recall). Four potential sources of change have been proposed to account for developmental improvements in performance on complex tasks: faster processing speed, increased capacity, reduced decay, and more resistance to interference. Although the dynamic model used here cannot perform complex tasks in its current form due to the very different task structures and demands, simulations in Chapter 4 showed how three of these general characteristics of memory—capacity, decay, and interference—change developmentally. It is possible that the faster encoding speed noted in Chapter 4 could also provide the faster processing speed in complex tasks: due to the differences in task demands, however, this comparison is not straight-forward. More work is needed to adapt this type of modeling framework to encompass all of the processes required to perform complex span tasks.

The dynamic model presented here produced eight characteristics of developmental change in visual working memory—collectively described as increases in real-time stability—through a single underlying mechanism. Comparisons to other working memory tasks within infancy and later childhood suggest that these developmental processes could capture many of the improvements across those tasks and developmental periods as well. This synthesis across areas of research suggests developmentality in the processes underlying working memory performance from infancy through childhood. As noted in Chapter 1, separate studies have shown that cognitive skills are predicted by both infants’ looking behavior and children’s performance on complex span tasks, raising the question of whether the same working memory processes underlie performance between infancy and complex tasks. The empirical and computational work presented here suggests that they overlap at least partially, and perhaps completely. An important next step in this line of research is to understand how the processes in common between infant looking tasks and complex span tasks support higher cognitive functions.

Future Directions: Real-Time Stability as a General Construct

Beyond this proposed synthesis between infant looking paradigms and complex working memory tasks in childhood, a long-term goal of the cognitive dynamics theory is to extend the general concept of real-time stability across broader areas of research. Through dynamic systems approaches to motor development, related concepts of stability have been demonstrated in the development of reaching, locomotion, and spatial
cognition. In the domain of reaching, Thelen and colleagues showed that successfully grasping an object requires bringing together multiple processes in the moment—holding the trunk stable, controlling both the direction and velocity of the arm’s movement, opening and closing the hand at the right moment (Thelen et al., 1993; Thelen, Corbetta, & Spencer, 1996; Thelen & Spencer, 1998). These different components of reaching may develop at different rates, but are not independent and combine to produce mature reaching.

In the domain of locomotion, Adolph and colleagues have shown newly acquired behaviors, like crawling and walking during infancy, are relatively unstable and therefore more likely to fail when task conditions change. For example, a series of studies assessing how infants descend slopes showed that age was not an important predictor of whether infants would attempt to descend steep slopes; rather, infants with more experience with the form of locomotion (crawling vs. walking) were better at judging their ability to descend (Adolph, 1997; Adolph, Eppler, & Gibson, 1993; Adolph, Tamis-LeMonda, Ishak, Karasik, & Lobo, 2008). This difference was evident within individuals over time: infants with little walking experience would resist descending a too-steep slope when in a crawling posture (the more stable form of locomotion) but would attempt descent of the same angle of slope while walking. Thus, the relative stability of each form of locomotion predicted how well infants could adapt to different task demands.

Applications to spatial cognition provide the most direct analogy to the development of visual working memory (see Simmering & Schutte, 2015; Spencer et al., 2007, for reviews). Beginning with infancy, increasing real-time stability can account for developmental changes in the Piagetian A-not-B task. In this task context, unstable representations of the current hiding location leads to more frequent errors on A trials in 8- to 10-month-old infants (Schöner & Dineva, 2007), as well as perseverative responding on B trials in 10- to 12-month-old infants (see Simmering et al., 2008, for details). Later in development, more stable representations can be maintained more robustly over the short delay, supporting correct performance. This notion of stability led to the counter-intuitive prediction that even younger infants would not search perseveratively in the A-not-B task. This is because the accumulation of long-term memory on A trials (the source of such errors) requires sufficient stability in working memory. Clearfield, Diedrich, Smith, and Thelen (2006) showed that 5-month-old infants reach correctly on B trials, indicating that perseveration requires some underlying stability.

Beyond infancy, as connectivity continues to strengthen, peaks may sustain over longer delays and in the absence of continued visual input (i.e., without lids marking the hiding locations; Schutte, Spencer, & Schöner, 2003). Between the ages of 3 and 6 years, as children are able to perform more challenging recall tasks, four types of changes in spatial working memory tasks
that arise through increasing real-time stability. First, the spatial range across which long-term memory influences working memory reduces: when recalling two nearby locations across trials, younger children’s responses are biased toward previous locations at relatively large separations between locations, whereas older children’s and adults only show such bias at small separations (Schutte et al., 2003). Second, children’s memory for locations is biased by the perceived reference frames in the within the task space, with a complex pattern of change over development. This pattern has been quantitatively modeled within the Dynamic Field Theory framework using the spatial precision hypothesis (Schutte & Spencer, 2009). Third, these findings have been generalized to predict performance in a different task context, position discrimination, with children showing improved discrimination (similar to color discrimination results from Simmering & Patterson, 2012) as well as changing influences of perceptual reference frames over development (Simmering & Spencer, 2008). Fourth, Schutte and Spencer (2010) showed that children’s spatial recall could be improved through increased perceptual structure in the task space, which augments real-time stability through stronger input.

Although the types of tasks and developmental effects that have been modeled in the Dynamic Field Theory framework differ between spatial cognition and visual working memory, they still represent a relatively narrow range of cognitive development research. The examples from motor development expand on this account further, but do not yet connect with areas traditionally considered “higher level” cognition. Some research has been used to expand this modeling framework in that direction, including accounts of rule use (as described above in the dimensional change card sort task; Buss & Spencer, 2014) and language development (Samuelson et al., 2009; Samuelson, Smith, Perry, & Spencer, 2011).

In the domain of word learning, Samuelson et al. (2009) showed that changes in the strength of input in a dynamic neural field model could simulate toddlers’ attention to shape versus material in a novel noun generalization task. Furthermore, Samuelson et al. (2011) demonstrated how interactions between features and space provided a foundation for children to associate labels with objects that were separated in time, using a variant of the two-dimensional model architecture described briefly above. In these cases, model simulations provided an explanation for how children’s behavior varied across task contexts—consistent with the cognitive dynamics perspective—but have not yet addressed questions of developmental change.

For the cognitive dynamics perspective and real-time stability to be expanded further, however, conceptual principles from this approach should be tested without depending on formal implementation in a model. This requires finding ways to assess and modify real-time stability empirically, to
evaluate whether it is changing developmentally and how it contributes to task performance. The simulations in Chapter 4 provide a number of potential avenues for this type of exploration. First, a less-stable system builds representations more slowly and loses them more quickly; changing the length of stimulus presentations and delays can evaluate time-based changes in representations. Second, unstable memory representations may be used reliably for relatively simple behaviors, but not for more demanding behaviors. The cognitive development literature is filled with such examples of “decalage,” in which children’s performance differs across multiple tasks that purportedly rely on the same underlying processes. The example tasks from the current monograph showed mechanistically how such a difference can arise through the details of the task structures and behavioral responses; identifying such differences requires careful analysis of the processes necessary to perform each task. Third, manipulating task details to provide more support for memory representations could augment stability and allow younger children to perform more comparably across different tasks. This is exactly how Perone et al. (2015) improved children’s performance in the dimensional change card sort task: providing children with additional color experience increased the stability of their color representations, which in turn supported their ability to sort cards along the color dimension. Increasing real-time stability suggests a developmental cascade of improvements: as working memory representations become more stable, more accurate long-term memory can be formed, and these representations can be used more reliably in service of other tasks.

CONCLUSIONS

The experiments and simulations presented here make two important contributions to the field of cognitive development. First, the cognitive dynamics theory and real-time stability hypothesis provide a specific, mechanistic account of visual working memory capacity over development and across behavioral tasks. This theory predicted that capacity estimates would differ across tasks despite the fact that both tasks rely on the same underlying memory system, and performance correlated across tasks. These results demonstrate the utility of computational models in formalizing theoretical constructs to test how tasks relate over development. Second, in the broader context of working memory research, this theory provides a potential explanation for why both infants’ performance on looking tasks and children’s performance on complex span tasks are related to later cognitive skills—the simulations in Chapter 4 suggest developmental continuity in the mechanisms supporting improvements in working memory. The critical next step for this application of the current theory is to investigate how the
common processes across working memory tasks relate to performance on standardized tests of reasoning and general intelligence.

Future empirical and theoretical work is needed to connect these processes to the broader cognitive skills that are related to working memory functioning, but the theory presented here takes an important step toward providing the specificity needed to make meaningful connections across tasks. Theories must be expanded beyond a single task, domain, and developmental period in order to explain—and predict—how working memory changes within an individual over time (Simmering & Perone, 2013). If we do not understand how the tasks we use in the lab relate to one another, we cannot predict when training in lab tasks will produce lasting improvements in behaviors outside of the lab.


DOI: 10.1111/mono.12254
© 2016 The Society for Research in Child Development, Inc.
REFERENCES


REFERENCES


REFERENCES


ACKNOWLEDGMENTS


Thanks to the families and students who participated in this research. The behavioral data presented here were part of the author’s dissertation, which was funded by a fellowship from the University of Iowa Graduate College. Portions of the data were presented at the 16th International Conference on Infant Studies in Vancouver, BC, the 30th Annual Meeting of the Cognitive Science Society in Washington, DC, and the 68th Biennial Meeting of the Society for Research in Child Development in Denver, CO. Special thanks go to: John P. Spencer for input and guidance during the conceptualization of the project; Shannon Ross-Sheehy for providing materials and analysis details for the looking paradigm; Jeffrey S. Johnson for programming the change detection tasks; Valerie Vorderstrasse and Chelsey M. (Patterson) Wood for assistance with data collection; Nicholas J. Fox and Sammy Perone for help coding and analyzing looking data; Martha W. Alibali, Karl Rosengren, and John P. Spencer for input on previous versions of this manuscript; and Jessica Horst, Heather Kirkorian, Audra Sterling, and Haley Vlach for support and encouragement during the writing and revision process.
This commentary considers Vanessa Simmering’s monograph on a dynamic-systems theoretical approach to understanding working memory development, with reference to the past, present, and future. In the section on the past, I attempt to provide a further historical context for the work, discussing from where it stemmed and how it is unique. In a second section, I contemplate the purpose of the present modeling. The aim of the monograph may be primarily to establish a simple possible account of development based on neural connection strength and dynamic principles; it should not be judged as a proposal of what is necessarily true. Finally, in the section on the future, I suggest some phenomena that dissociate performance levels from stability over time and, therefore, appear to require modifications of the theory. Several suggestions are made as to where further refinement of the modeling effort could lead.

Upon my arrival at the University of Missouri, where I have worked for most of my career, I was assigned to a laboratory recently vacated by Esther

Corresponding author: Nelson Cowan, Department of Psychological Sciences, University of Missouri, McAlester Hall, Columbia, MO 65211; email: cowann@missouri.edu
DOI: 10.1111/mono.12257
© 2016 The Society for Research in Child Development, Inc.
Thelen, a foremother of the present-day interest in dynamic systems theory as applied to child development. When Esther moved out of her infant-stepping lab, she left attached to the door one of my favorite cartoons, which I have since misplaced but have thought of often. It featured a beaver talking to another animal, with an enormous hydroelectric dam in the background. Of this structure the beaver remarked, “Well, I didn’t actually build it, but it was based on my ideas.” The dynamic systems modeling of processes underlying working memory from infancy through childhood in the present monograph by Vanessa Simmering might be viewed as a hydroelectric version of earlier beaver dams. My contribution is to comment on how the dam is related to past work (the history), how well it is operating (the modeling), and where we might go from here (the persistent questions).

THE PAST: A LITTLE HISTORY

As series of important beaver dams, first there are the empirical phenomena. Working memory is a critical process in human cognition, representing the small amount of information that can be held in mind and used in the service of many processes: remembering the early part of a sentence long enough to integrate it with what comes later, carrying a digit when doing mental addition, using mental imagery to rotate puzzle pieces to see which ones might fit together, or in an infant, perhaps comparing a babbled utterance to an adult model or retaining memory of Mom as she disappears behind a door. The study of working memory may be as old as the study of memory generally. Ebbinghaus (1885/1913) is typically credited with initiating the scientific study of memory, in his groundbreaking research in which he repeatedly tested himself until he learned series of nonsense syllables. What is germane here is his finding that, although a list of 12 syllables could be learned only after 16 repetitions, a shorter list of 7 syllables could be learned in a single presentation or, as he put it (p. 33), a “first fleeting grasp” of the items. Studies of the childhood development of immediate memory soon followed in the form of memory span experiments (Bolton, 1892; Jacobs, 1887).

Miller, Galanter, and Pribram (1960) introduced the term “working memory” to describe memory for one’s near and distant future plans, and Sperling (1960) greatly expanded our understanding of temporary memory in general, and specifically in the case of visual stimuli. Baddeley and Hitch (1974) popularized the term working memory and applied it to a multicomponent system with devoted automatic buffers (verbal and by implication nonverbal visual) as well as an attention-demanding central resource comprising executive processes. Tests of limited resources were later extended to infant and child development by numerous investigators, for
example, in tests of a relation between memory and processing speed in children (Case, Kurland, & Goldberg, 1982; Hulme, Thomson, Muir, & Lawrence, 1984). More recent and closely germane strands of the developmental research history are well-covered in the monograph.

A complementary set of key beaver dams are the theoretical explanations of the development of working memory and cognition. As noted in the monograph, many investigators have offered verbal and pictorial explanations for how working memory operates, or how it develops. That is still a far cry from a principled, mathematical model of how working memory develops. One can imagine that certain verbal or pictorial models lead to particular predictions, but sometimes this kind of speculation depends on assumptions that have not been made clear, and sometimes are not fully appreciated even by the investigator doing the speculating. Mathematical modeling leaves less room for unappreciated assumptions because one needs to fill in the assumptions to yield the desired mathematical result.

Work using equations to specify psychological processes seems to have begun with Ernst Weber and Gustav Fechner in the late 1800s. Estes (1950) was perhaps the first to show that mathematical precision could be brought to the task of stating models of learning and memory in a more rigorous fashion, and many related approaches have followed. Investigators sometimes called neo-Piagetian (McLaughlin, 1963; Pascual-Leone & Smith, 1969) used concepts of temporary memory and information processing to explain conceptual development in childhood, including some mathematical specification. Others have pioneered various principles that were incorporated into the present modeling, such as the principle of lateral inhibition applied to cognitive concepts in memory by Walley and Weiden (1973).

THE PRESENT USE OF MATHEMATICAL MODELING

Before discussing the topic of mathematical modeling further, I would note that my qualifications to this topic include being all over the map in terms of my attitude. I have a love–hate relationship with modeling. I can see important pitfalls of mathematical modeling, and I can see enormous benefits. I have ignored some mathematical models that I am expected to know, and I have done mathematical modeling myself (generally with technical help). However one feels when reading or trying to read a mathematically involved work like the present monograph, I am probably sympathetic.

To become a connoisseur of mathematical modeling, one must first appreciate that there are multiple aims of the modeling, and the correct aim must be attributed for a model to be appreciated. The present modeling shows how certain theoretical accounts of working memory development are possible, not necessarily how they are necessary. Is it possible to explain why
children do steadily better in working memory tasks as they get older? Why some materials are remembered better than others? Why what develops may include both the number of items in working memory (e.g., how many colors) and the precision by which those items are represented (the fidelity of the remembered shades)? Can we understand why performance depends to some extent on the individual and to some extent on the task he or she is to carry out? How about dependence on the similarity of items to be remembered and the interference between them (cf., Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012)? Can all of these factors be understood with the same simple principles, tying together development from infancy through childhood? In the present work, these questions are all answered in the affirmative. A working assumption is made that, as infants and children develop, the connectivity in the neural architecture is strengthened. At least with one reasonable architecture, the simple principle of connectivity is said to account for all of these phenomena.

For the modeling results to be useful, their limits must be understood. There need be no claim that the details of the architecture are in fact correct. Perhaps a completely different model would also explain the facts equally well. What can be established, though, is that the model, though itself complex in its mathematical details, is essentially an argument for simplicity. Before, there may have been a perceived need for different principles to explain the infant and child data. Several complex mechanisms might have been thrown in to explain separately the development of capacity in terms of the number of items represented in working memory and the precision of the representations. Instead, these details can be seen as falling out of the same architecture with the principle of increasing connectivity, leading to increasing stability of the representations.

One naturally hopes that what has been proven possible in one’s model eventually proves to be the actual case, or even proves to be necessary to account for the results. Then one could be considered “right.” Even without being right, modeling work is useful as it sharpens up the concepts being discussed, which aids in evaluating them.

What the model does best is to strengthen the plausibility of a dynamic systems approach. As stated in the monograph, “Within dynamic systems theory, the focus is on how behavior emerges from multiple underlying causes, encouraging researchers to explore the various contributions to behavior, and to evaluate the robustness of behavior relative to the circumstances required to support them.” This mundane-sounding theoretical statement actually runs counter to the thinking of Jean Piaget, who tended to believe that once a mental structure was fully acquired, it was robustly demonstrated across task demands. The neo-Piagetians would have disagreed. Once, as a teaching assistant during graduate school in the later 1970s, I confirmed to my own satisfaction the neo-Piagetian stance. By
mid-childhood, children are supposed to understand that water poured from a squat beaker to a thin one does not change in volume, and that clay rolled from a ball into a sausage shape does not change in volume either. With undergraduate students, I upped the level of complexity by asking what would happen if heavy clay in a ball shape versus a sausage shape were totally submerged into identical beakers of water. Many college students incorrectly predicted that the sausage-shaped clay would raise the water level more, a failure of the conservation-of-matter principle in a complex context.

The model also shows that there is room for error in the encoding, maintenance, processing, decision, and response phases of a task. We may choose only one of these as the source of error in a particular simplified model, but no model of behavior across contexts can survive without coming to grips with processing during all of the phases of the task.

FOR THE FUTURE: SOME UNANSWERED QUESTIONS

The first and perhaps foremost unanswered question I have is how long the model can persist before the need to modify it sets in. When the model does have to be modified, how extreme will the modification have to be? How many detailed facts can the model account for and still elegantly show that the increasing-connectivity principle accounts for the development of working memory across tasks?

There is some mystery left for me with regards to the effect of the similarity of the items to be remembered. It was stated that “When multiple peaks form near one another, their related inhibition combines, making it more difficult to form new peaks.” Some of the sources cited, however, indicate that sets of more similar items, such as several shades of green, can be remembered better than sets of less similar items, such as different colors together (e.g., Lin & Luck, 2009). I cannot figure out how to reconcile the principle with the findings.

Although the developmental results that were accounted for are rich and varied, they do generally seem to involve just about every aspect of working memory performance getting better with age. That pattern of development in itself has been referred to as “the dull hypothesis” (Perfect & Maylor, 2000); the dull hypothesis can be rejected when one finds an interaction between tasks and age groups. To some extent, in the present approach, the dull hypothesis is rejected in the fitting of the model to young children in two tasks, the typical infant task and the typical child task.

In the future, however, it may be necessary to reject the dull hypothesis more severely. Among the predictions of the model is that with development, memory representations gain more stability and, therefore, are preserved better across a retention interval; that is, in more mature participants, the
representations decay less than they do in younger ones. In general, though, we have not found that decay difference across age groups. In one procedure (Cowan, Nugent, Elliott, & Saults, 2000), children were tested on memory for lists of digits that were ignored at the time of their presentation and then occasionally were cued for recall 1, 5, or 10 sec after the last digit in the list. With the list length adjusted to each individual’s span, the rates of decay of the list across 10 sec did not change significantly between 7 and 20 years. There was, however, a large age effect restricted to the final serial position, which could be accounted for by age differences in either sensory memory persistence or covert attention-shifting to the end of the list. There is a similar finding of an age difference in the decay of isolated tone information (Keller & Cowan, 1994). Using spatial arrays of unfamiliar characters, followed by a mask to reduce the use of sensory memory, Cowan, Ricker, Clark, Hinrichs, and Glass (2015) found no difference between 7-year-olds and college students in the rate of decay of memory for array characters across 10 sec. Taken together, these results suggest that a key principle of the model may not apply in the same way to sensory and conceptual information.

To understand age differences in the decay of information, perhaps one needs to specify the mechanism that produces stabilization. Camos and Barrouillet (2011) have studied the ability to use spare time to refresh working memory representations. They presented series of animal pictures to be remembered and, between each pair of animals, either one or two colored spots to be named. An important change occurred between 6 and 7 years of age. In 6-year-olds (kindergarteners), performance was better when the total time between animals was shorter. In 7-year-olds (first-graders), performance was better when the proportion of time between animals that was free for refreshing was high, no matter whether the total time was short or long. The results suggested that 6-year-olds do not use the refreshing process and, therefore, are subject to steady decay, whereas decay is counteracted by refreshing in 7-year-olds. This study differs from the ones by Cowan et al. (2000, 2015) in that those studies presented conditions for which refreshing would be difficult or impossible (because the items were unattended digits or attended arrays of unfamiliar characters). In all, the results suggest that age differences in instability of representations might often be attributed to the growing effectiveness of attention-demanding refreshing processes. The idea is that only in tasks in which older participants can refresh items can they stabilize their working memory representations during the maintenance period better than younger, less mature participants.

The notion that attention-based processes are used for refreshing does not contradict the model presented in the present monograph, but it does appear to restrict the scope of the model. The scope may have to be restricted to situations in which there are age differences in a process that can be mapped onto better stability of the traces. For some stimuli, in some age
groups, such as letters in children 5–7 years old, there will be age differences in knowledge that can result in differences in stability of the representations that are already manifest at the time of encoding. For other stimuli, age differences occur at the time of maintenance through refreshing. These sorts of differences would be expected to produce age differences in decay. For stimuli that are neither encoded nor refreshed advantageously by older age groups, there may be no difference in decay.

The finding of no age difference in decay is interesting when it is obtained with the list length adjusted to the participant’s ability (e.g., Cowan et al., 2000) or with age groups at different levels of performance despite no decay differences (e.g., Cowan et al., 2015). The intriguing thing here is that age group effects in performance level and decay are dissociated. It is unclear to me how to modify the present dynamic systems model to produce this dissociation. It seems as if the “self-sustaining” state of working memory may be self-sustaining in some ways based on automatic processes (such as encoding clarity that may often favor more mature participants who have more knowledge) and self-sustaining in other ways only when voluntary strategies can be implemented (such as memory persistence over time, which may favor more mature participants only for stimuli that lend themselves to refreshing or rehearsal processes).

Beyond empirical issues such as this, it is possible to use modeling in a more specific manner than was done in the present monograph. Many researchers attempt to produce models that match the data so closely that they can present a pattern of predicted results in one panel of a figure and a panel of obtained results next to it. In this kind of approach, to get the data and model to resemble one another so closely, there are usually a host of auxiliary assumptions resulting in parameters in the equations with arbitrary values. The strength of the approach is that there can be no doubt that one can get the actual pattern of results from the model; the down side is that one still has to figure out how much of the success comes from parameters that reflect basic principles of the model (e.g., connection strength) and how much of the success depends instead on parameters that were supposed to be incidental or unimportant but actually are doing the heavy lifting, forcing the model to match the data in a manner that has little to do with its stated principles. A related problem is that it is difficult to get these more specific models to account for a variety of circumstances as the present modeling does; for example, I have seen serial recall models that could not be modified in a foreseeable way to account for free recall.

In the more specific modeling approaches, one often compares multiple models that differ in important ways and finds out which model fits the data better according to standard fit statistics. In the present dynamic systems approach, it would be possible to present models with connectivity that develops at different rates for different layers in the model, which may make
the model more complex but may be consistent with evidence that different parts of the brain develop at different rates (Sowell et al., 2003; Thomason et al., 2008). Some fit statistics (such as AIC and BIC) are designed to penalize models for extra complexity and see if the more complex models are worth it.

Finally, modeling is a tricky exercise when one takes the popular approach of comparing two or more possible models to determine which model is more apt. In a recent effort (Cowan et al., 2016), we tried out models of how adults perform in a new task in which they were presented with two arrays of colored spots in succession, and asked to judge how many of the array items changed color between a studied display and a test display. In our first modeling approach, a separate decision was assumed to be made by the participant for each array item. The model fit the pattern of means beautifully, but utterly failed to fit the distribution of responses in each condition, leading to a rejection of the model in favor of other models that would not have been considered, had the first model not failed.

To sum up, modeling is a tricky enterprise. The present work very nicely sets out a simple neural scheme and then follows the implications of that scheme for working memory development in the case of visual arrays of items to be remembered. Perhaps the strongest recommendation for the model is that it served as a motivator to get Simmering to test the same children on both the usual infant and usual child procedures, leading to an elegant set of results that in turn led to further refinement of the model. The model served as a very nice proof that one can go far with basic, elegant neural principles. In a similar manner, in the future, one could imagine that the modeling would help to guide further work on the nature of decay effects, the relation between items in working memory and precision of the representations, and potential differences between the rates of development of these concepts. It might lead to predictions about when and in what way working memory can be trained. In normal individuals, there has been little evidence that working memory training helps to improve tasks other than those similar to the one trained (e.g., Melby-Lervåg & Hulme, 2013; Redick et al., 2013), but there might be more hope in individuals with processing abnormalities. It is possible to impair a model and then see how it might be trained, leading to predictions that could be tested in real individuals.

The work illustrates that further progress depends on close communication between researchers with empirical and modeling orientations, as well as between researchers emphasizing development in infancy versus later childhood.

REFERENCES


CONTRIBUTORS

Vanessa R. Simmering is an Assistant Professor in the Department of Psychology at the University of Wisconsin–Madison. Her research program focuses on understanding cognition and development from a dynamic systems perspective by considering how multiple factors influence behavior across timescales and contexts.

Nelson Cowan is Curators’ Professor at the University of Missouri, Columbia. He has conducted NICHD-funded research on working memory, attention, and their childhood development since 1984. His work is shaped by a philosophical interest in human consciousness combined with the hope that the findings can be useful to educators and neuropsychologists.

DOI: 10.1111/mono.12258
© 2016 The Society for Research in Child Development, Inc.
Page numbers in *italics* represent figures.

**accuracy**, 68–70, 72, 73, 74
- color change detection, 68–70, 69
- shape change detection, 72, 74

- change-preference task, 57–58
- color change detection, 65–66
- shape change detection, 71–72*

**attention**, 18–19, 52, 111, 118–121, 154–155
- attentional model, 13, 14

**binding**, 28–30, 49, 116–118, 120

**brain**
- cortex, 108
- maturation, 17, 18
- myelination, 107
- neural activation, 55n ii, 121
- prefrontal cortex, 17, 107
- synaptic pruning/strengthening, 107
- synchronous neural firing, 26

**capacity estimates**
- change-preference task, 62
- color change detection, 70
- differences, 102
- limits, 95, 99
- shape change detection, 74

**change-detection task**
- capacity estimates, 20, 71, 78, 110, 119–120
change-preference task, 62
color change detection, 63, 64–71, 77, 79, 80, 83, 114
comparisons across tasks, 75–78, 77, 78, 80
in computational model, 33–36
performance in cognitive dynamics model, 36–42
predictions, 46–47, 47, 50–51, 53, 56
shape change detection, 71–75
simulation method, 82, 83, 85, 87, 88–92, 93, 96
theories of development, 26–30
visual working memory, 19–20, 20
change-preference task
comparisons across tasks, 75–80, 77, 110
development of, 21
in computational model, 33–36
performance in cognitive dynamics model, 42–45
predictions, 45–53, 48, 56, 57–64
sample trial, 21
simulation method, 82, 83, 87–88, 92–93, 104, 110
theories of development, 26–30
working memory research, 18
coding, observer, 59–63
cognitive dynamics theory, 25–26
change-detection task, theories of development, 26–30
change-preference task, theories of development, 26–30
real-time stability and —, 30–54, 109–110, 121–126
cognitive processes, 8
dynamics, 9, 30, 51, 53
stability, 9, 33
color change detection. See change-detection task
color-location binding, 28–29
comparison, process of, 18, 21, 23, 28–29, 35–41, 43, 45, 46
certainty intervals
change-preference task, 61–62, 61
color change detection, 69
connectivity
account, 18
changes in, 105–108
cognitive dynamics, 33–34
development, 124
predictions across tasks, 46–47
quantitative simulations, 83–85
strengthening, 29–30, 36, 54–55, 80, 94–102, 110–111, 113, 124, 152, 153, 155
decay, 95, 98, 108n i, 123, 153–155
delayed response, 15
Dynamic Field Theory, 9, 32, 36, 45, 55nn I, iii, 84, 105, 107, 110, 116–117, 121, 125
dynamic model, 31, 52–53
dynamic systems theory, 31

encoding speed, 18–19, 37, 39, 41, 43, 45, 46, 52, 95, 97, 117–118, 122, 123
excitation, 18, 32–36, 46, 53
eye-tracking measure, 111–113

familiarity detection, 16, 18, 42, 53–54, 76, 78–79, 103, 112, 122
fixation system, 42, 43–44, 119, 145–146
gating mechanism, 37

habituation paradigms, 16–18
Hebbian fields, 143–144

inhibition, 9, 11, 18–19, 32–34, 36, 41, 46, 53, 104, 115, 141–142
“integrated object,” 116
interference, reduced, 95, 97–98, 123

locomotion, development, 123
looking tasks/behaviors, 8
   change-preference task, 60, 61, 62, 83, 122
   comparisons across tasks, 80
   infant research, 15–19, 22, 119
   relationship with memory and recognition, 42

maintenance, 37, 39, 41, 43, 45, 46, 97
memory
   array, 19, 29, 37, 96, 115, 120
   correspondence between behavior and —, 95, 99
   delay, 39, 46, 97–98, 114
   long-term, 35–36, 50, 83, 104, 125
   relation with recognition and looking, 42
   See also working memory capacity
multi-component model, 12, 14
multiple features, 111, 116–118

novelty detection, 16, 18, 21, 42–43, 52–54, 76, 78–79, 93, 103, 105, 112, 122, 146
participants, 56
  change-preference task, 57
  color change detection, 64
  comparisons across tasks, 76
  shape change detection, 71
perceptual processing, 34–36, 44–46, 139–141
performance across tasks, 103–105, 110
predictions, empirical tests
  change-preference task, 57–64
  color change detection, 64–71
preference scores, 61, 76
procedure
  change-preference task, 58–59
  color change detection, 66–67
  shape change detection, 72
processed-based computation model, 12
processing speed, 18, 23, 123, 151
quantitative simulations, 83–85
  method, 85–88
reaching tasks, 15–16, 123–124
real-time stability
  cognitive dynamics, 30–54, 109–110, 121–126
  comparisons across tasks, 80, 110
defined, 30
  increasing, 46, 79, 82, 95–103, 110, 117–118, 122–123
terminology, 36, 55n iii
testing hypothesis, 82–108
recognition, relationship with memory and looking, 42
resolution, 113–114
resource-based explanation, 13–14, 27, 30
response process, 36–41, 50, 144–145
shape change detection. See change-detection task
short-term memory
  vs. working memory, 10–12
“slot” models, 26, 30, 38–39, 114
spatial precision, 36, 45, 123–125
stability
  described, 33
  self, 33, 52

165
See also real-time stability
switch rates, 61, 62–63, 76, 79, 80, 104, 110, 119, 122–123

tasks
  comparisons across, 75–80
  context, 95, 99–101, 123
  See also specific task

undifferentiated resource, 13–14

visual paired comparison, 21
visual working memory. See working memory capacity

word learning, domain, 125
working memory capacity, 7–9
  adult, 20
  change-detection and change-preference task theories, 26–30
  children, 20, 20, 21–22
  color, 40
  comparisons across tasks, 80, 110
  described, 8
  empirical tests of predictions, 56–81
  estimation, 37, 51, 70, 74
  field, 142–143
  history, 150–151
  improvements, 15–16
  increase, 49, 82
  infant research, 14–19
  model simulations, 82–108, 146–147
  modeling dynamic processes, 7–24
  vprocess-oriented theory, 8
  vrepresentations, 19
  research, 9–10
  short-term vs. —, 10–12
  theories of, 12–14, 23 (see also attentional model; multi-component model; undifferentiated resource)
  use of term, 150–151
  See also memory