2008

Developing a magic number: the dynamic field theory reveals why visual working memory capacity estimates differ across tasks and development

Vanessa Renée Simmering-Best
University of Iowa

DEVELOPING A MAGIC NUMBER: THE DYNAMIC FIELD THEORY REVEALS WHY VISUAL WORKING MEMORY CAPACITY ESTIMATES DIFFER ACROSS TASKS AND DEVELOPMENT

by

Vanessa Renée Simmering-Best

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

December 2008

Thesis Supervisor: Associate Professor John P. Spencer

ABSTRACT

Many daily activities require the temporary maintenance and manipulation of information in working memory. A hallmark of this system is its limited capacity—research suggests that adults can actively maintain only about 3-4 items at once. A central question in working memory research is the source of such capacity limits. One approach to this question is to study the developmental origins of working memory capacity. Developmental working memory research has shown a general increase in capacity throughout childhood and into adulthood. There have been few investigations, however, into the mechanisms behind these developmental changes, and proposals that have been put forth do not specify the processes underlying changes in capacity. One particularly puzzling, and unexplained, finding is an apparent decrease in visual working memory capacity over development, from an adult-like capacity of 3-4 items at 10 months to 1.5 items at 5 years. One probable source of this developmental discrepancy is that these two capacity estimates were derived from different tasks: preferential looking in infancy and change detection in later childhood. Although these tasks differ in many respects, existing theoretical explanations of the processes underlying performance in these tasks are underspecified, making it difficult to identify the origin of the apparent regression in capacity over development.

To investigate the developmental discrepancy across tasks, I developed a unified model to capture how capacity limits arise in preferential looking and change detection. I used this model to generate three specific behavioral predictions: 1) in preferential looking, children and adults should show higher capacity estimates than infants; 2) capacity estimates should be higher in preferential looking than in change detection when tested in the same individuals; and 3) although capacity estimates differ, performance should be correlated across tasks because both rely on the same underlying working memory system. In addition, I proposed a fourth prediction regarding the unified model:

developmental changes in both tasks should be captured by a specific developmental mechanism, the Spatial Precision Hypothesis. Results from three experiments and model simulations confirmed all four predictions, providing strong support for a new Dynamic Field Theory of visual working memory.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

DEVELOPING A MAGIC NUMBER: THE DYNAMIC FIELD THEORY REVEALS WHY VISUAL WORKING MEMORY CAPACITY ESTIMATES DIFFER ACROSS TASKS AND DEVELOPMENT

by

Vanessa Renée Simmering-Best

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

December 2008

Thesis Supervisor: Associate Professor John P. Spencer

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Vanessa Renée Simmering-Best

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Psychology at the December 2008 graduation.

Thesis Committee: _____
John P. Spencer, Thesis Supervisor

_____
Larissa K. Samuelson

_____
Jodie M. Plumert

_____
Bob McMurray

_____
Rodica Curtu

ACKNOWLEDGMENTS

This thesis, along with my entire education, would not have been possible without innumerable people who have helped and supported me throughout this long process. I will do my best to name the most important of them here, starting (where else) at the beginning.

First and foremost, I must thank my family. My parents, Bob and Dorothy, instilled in me from the very beginning a hunger for knowledge and the perseverance to pursue my dreams. Thank you for supporting me through everything I have done, no matter how crazy an idea it may have seemed at the time. I am also grateful to my siblings, Marcia, Nathan, and Erin, for being such interesting people that it inspired me to spend my life figuring out what makes people the way they are. You all have been a great source of support as well, listening to my endless introspections and thoughts about life (and not a little bit of complaining along the way). Special thanks are owed to Marcia, my fellow academic, for understanding the things that only those of us "on the inside" can.

Next I want to thank a few key teachers I had through my early years, who formed me into the student I had to be to achieve everything I have. First was Mary Haaf at Mulberry Elementary School in Muscatine, Iowa, who recognized that I was different from her other students, and took great initiative, time, and effort to ensure that I got the best education I could. Without her early encouragement and support, I don't know if I could have made it as far as I have. At Muscatine High School, I was fortunate to take English courses from Beth Hetzler and Kathy Brooker, both of whom equipped me with invaluable skills in critical reading and writing. In addition, Dennis Inman taught me the beauty of calculus, how math could be used to describe anything in the world. Little did I know at the time, but someday this training would form the foundation of my research. The fundamental skills he taught me will continue to serve me for the rest of my career.

Over my eleven years (yikes!) at the University of Iowa, I have learned from a great many faculty and fellow students, more than I can name here. Beginning with only my second psychology course as an undergrad, I had the good fortune to meet my advisor, John Spencer. I simply cannot overstate the importance of this experience—I doubt I would even be in this line of work had it not been for John's great inspiration and guidance. Through his superb teaching, both in and out of the classroom, John showed me that human development is the most interesting process in the world. He taught me that psychology could (and should!) be a science as rigorous as biology and physics, and that as good scientists we can revolutionize the way people think about development. It is no exaggeration to say I owe the path of my career to his ideas and encouragement.

In addition, John was an indispensable advisor on the practical aspects of academia, taking the time to teach me anything and everything he could to prepare me for this career. Along side him in this process was my unofficial co-advisor, Larissa Samuelson. She has provided the perfect complement to John, reinforcing what he has taught me as well as providing a vital female point of view. Beyond their intellectual support, I am also eternally grateful for the friendship John and Larissa have provided over the years. They have achieved the perfect balance between mentorship and camaraderie, and they will certainly be what I miss most from my years at Iowa.

The additional core developmental faculty at Iowa have also played a fundamental role in my decision to pursue this line of work: Jodie Plumert, Lisa Oakes, Scott Robinson, Mark Blumberg, and Bob McMurray. Each has, in their own way, shaped who I am as a scientist and supported me through the decisions I have made along the way. Other faculty at Iowa that have had a particular impact on my education and career over the years include Shaun Vecera, Steve Luck, Gregg Oden, Prahlad Gupta, Andrew Hollingworth, Eliot Hazeltine, and Inah Lee. Thank you all for everything you have taught me along the way, and a special thanks to those who helped me decide that Iowa was the best place for my graduate training (which it certainly was!).

Although he is not on the faculty at Iowa, Gregor Schöner has been an essential advisor to me as well. His skills in teaching complicated mathematical modeling are unparalleled, and his patience for taking non-mathematicians under his wing is admirable. He has also provided me with great friendship through the years, and I am looking forward to many years of collaboration to come. Thanks are also owed to him for introducing me to "the Germany crew" of modelers, Evelina Dineva (now a colleague at Iowa), Claudia Wilimzig, Christian Faubel, Sebastian Schneegans, and Yulia Sandamirskaya. You have served as great teachers as well, and I continue to look forward to our collaborations.

In addition to the faculty at Iowa, my fellow graduate students have played a much bigger role in my education than I could have anticipated. First were Anne Schutte and Alycia Hund, who served as mentors during my undergraduate years and helped my transition into graduate school. Special thanks go to Anne for her work introducing me to modeling, and managing the lab so well in its early years. Within the lab, I thank Phillippe Taborga, Brandi Dobbertin, Valerie Vorderstrasse, and Chelsey Patterson for their work as lab coordinators over the years. In a category all her own is Wendy Troob, who uniquely shares with me the path from undergrad assistant to lab coordinator to grad student. It's been quite the journey we have taken together over the years, and I am grateful for your friendship and support through it all.

Very special thanks go to the "modeling guys" in the lab—John Lipinski, Jeff Johnson, and Sammy Perone—who have been along side me as we have learned the skill and art of computational modeling. You have each been a great source of collaboration, inspiration, education, and companionship to me over the years. I look forward to sharing many beers with each of you at conferences to come! In addition, the newest member of our lab, Aaron Buss, though only overlapping with me for a short time, has also helped me make it where I am today. He provided me the unique opportunity to teach and learn at the same time, and is proving to be a great addition to our team.

Outside of my lab, special thanks go to those I consider my closest cohort: Kara Recker, Kristine Kovack-Lesh, and Jessica Horst Kavlie. You were great friends to me throughout the years, as well as great classmates and colleagues. Thanks especially for your support and guidance in the crazy transition from student to faculty: to Jessica and Kristine for leading the way on the job market, and to Kara for being along side me as we went through the same insanity at the same time. I am looking forward to following your careers in the years to come, and finding time to catch up at all of those fun developmental conferences!

In addition, I owe thanks to the other developmental students who have been there with me at one time or another: Shannon Ross-Sheehy, Penny Nichols-Whitehead, Jennifer Lee, Valerie Mendez-Gallardo, Christine Ziemer, Lynn Perry, Katie Haggerty, and Marcus Galle. I have enjoyed learning from you in classes and meetings, and watching the progress you have made. Thanks for your words of support, especially in these last few months as my level of stress and insanity has reached its pinnacle.

I must also acknowledge the many sources of funding that allowed me to pursue this degree. First, as an undergraduate, I was supported by the Undergraduate Scholar Assistantship – Iowa Advantage Program (through the University of Iowa Career Center) as I completed my honors thesis. Throughout my first five years of graduate school, I received great financial support through a Presidential Fellowship from the University of Iowa Graduate College. Fortunately, when and where this fellowship did not cover my studies, John was able to provide me with support from his federal grants over the years: National Institute of Mental Health Grant R01 MH62480, National Science Foundation Grants BCS 00-91757 and HSD 0527698. In addition, although the monetary amounts are small in comparison, my graduate career was made much easier and more enjoyable due to the various travel awards I received to present my work at conferences, both through the University—including the Psychology Department, Student Government, Graduate Student Senate, and Women in Science and Engineering—as well as through

national sources—the Computational Cognitive Neuroscience Conference Student Travel Fellowship, and the American Psychological Foundation Elizabeth Munsterberg Koppitz Travel Stipend.

Now at the end of my graduate training, I owe a repeated extraordinary thank you to my committee members—John Spencer, Larissa Samuelson, Jodie Plumert, Bob McMurray, and Rodica Curtu—for their unending patience and support, and their willingness to read the absurdly long document that follows. As I'm sure you appreciate, I literally could not have done this without you.

Last and forever, that brings me to my dearest husband, Andrew Best. Words cannot describe how much your love and support have meant to me over the last six years, and even more so in these last few months when the disarray in our lives hit an all-time high. Thank you for keeping me grounded, keeping me laughing, and keeping me sane. I could not have asked for a better companion through it all—you truly make all things in life better.

ABSTRACT

Many daily activities require the temporary maintenance and manipulation of information in working memory. A hallmark of this system is its limited capacity—research suggests that adults can actively maintain only about 3-4 items at once. A central question in working memory research is the source of such capacity limits. One approach to this question is to study the developmental origins of working memory capacity. Developmental working memory research has shown a general increase in capacity throughout childhood and into adulthood. There have been few investigations, however, into the mechanisms behind these developmental changes, and proposals that have been put forth do not specify the processes underlying changes in capacity. One particularly puzzling, and unexplained, finding is an apparent decrease in visual working memory capacity over development, from an adult-like capacity of 3-4 items at 10 months to 1.5 items at 5 years. One probable source of this developmental discrepancy is that these two capacity estimates were derived from different tasks: preferential looking in infancy and change detection in later childhood. Although these tasks differ in many respects, existing theoretical explanations of the processes underlying performance in these tasks are underspecified, making it difficult to identify the origin of the apparent regression in capacity over development.

To investigate the developmental discrepancy across tasks, I developed a unified model to capture how capacity limits arise in preferential looking and change detection. I used this model to generate three specific behavioral predictions: 1) in preferential looking, children and adults should show higher capacity estimates than infants; 2) capacity estimates should be higher in preferential looking than in change detection when tested in the same individuals; and 3) although capacity estimates differ, performance should be correlated across tasks because both rely on the same underlying working memory system. In addition, I proposed a fourth prediction regarding the unified model:

developmental changes in both tasks should be captured by a specific developmental mechanism, the Spatial Precision Hypothesis. Results from three experiments and model simulations confirmed all four predictions, providing strong support for a new Dynamic Field Theory of visual working memory.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

WORKING MEMORY CAPACITY

In the effort to understand how human memory works, researchers and theoreticians within the psychological sciences have divided memory in a variety of ways, for example, sensory versus short-term versus long-term (e.g., Atkinson & Shiffrin, 1968), episodic versus semantic (e.g., Tulving, 1972), declarative versus procedural (e.g., Squire, 1986), implicit versus explicit (e.g., Schacter, 1987). One particular sub-division of memory has been studied extensively since it was proposed by Baddeley and Hitch (1974) more than three decades ago: a three-component working memory system comprised of a verbal store (the articulatory loop), a visual store (the visuo-spatial sketchpad), and a central executive.[1] In this model, the central executive selects and controls processes operating on the information being held in the verbal and visual stores. These "slave" systems serve to temporarily store and manipulate information, but each has a limited capacity. One common example of a task that depends on working memory is mentally adding a series of numbers—to arrive at the correct answer, one must hold intermediate values in mind before adding the next number. This type of process underlies many facets of everyday human behavior, from sentence processing to integration of a visual scene across saccades to planning sequences of actions to move through the world or reach for objects. The ability to store and manipulate information is vital for successful behavior.

Although Baddeley and Hitch (1974) were the first to propose a conceptual model of working memory that posited separate stores for verbal and visual information, they were not the first to suggest that memory has a limited capacity. Miller (1956) proposed a

---

[1] Since the original paper, a fourth component has been added by Baddeley (2000)—a multimodal episodic buffer that can store information in a multi-dimensional code.

"magical number seven, plus or minus two" as the specific limit for what was then termed "immediate" memory. Outside of the field of psychology, Miller's proposal of a seven-item capacity has largely survived this past half-century; within the field, however, it took almost 20 years before other researchers challenged this conclusion (see Cowan, Morey, & Chen, in press, for review) . The first challenge came from research by Baddeley and colleagues (Baddeley, Thomson, & Buchanan, 1975) showing that the capacity of verbal working memory depended on the details of the items being remembered, for example, that capacity tends to be lower for long words than for short words. In the same year, a chapter by Broadbent (1975) refined the idea of capacity by proposing that near-ceiling performance should be expected for list-lengths within one's capacity, as opposed to the 50% correct threshold previously used. With this modification, the capacity of verbal working memory seemed to be only three to four items. Broadbent noted, however, that there are cases in which people can remember more than four items reliably, but only when the items being stored have been learned as units or "chunks" (see also Broadbent, 1971).

Although the field has debated the details of capacity over the years, there is little disagreement that working memory has a limited capacity or that working memory should be divided into separate verbal and visuo-spatial systems (see Cowan, 2001, and commentaries for in depth discussion; see also Miyake & Shah, 1999). A central question in working memory research is the source of such capacity limits. One approach to answering this question is to study the developmental origins of working memory capacity. The first step is to map out developmental changes in working memory tasks, finding the age at which children first show working memory abilities, how these abilities change throughout childhood, and when children finally reach adult-like levels of performance. Mapping these changes has been the focus of most research on the development of working memory capacity. The next step should be to look for the causes behind these developmental changes; however, only a few researchers have made this

question a focus of their work (but see Case, 1995). In the sections that follow, I review research in verbal and visuo-spatial working memory, focusing specifically on proposals regarding the nature and developmental origins of capacity limits. Then, I describe recent changes in the study of visual working memory that have revealed a discrepancy in capacity estimates over development, as well as the limitations of our theoretical understanding of how different tasks assess capacity. The goal of this project is to use a computational modeling approach to explore the mechanisms underlying the development of visual working memory capacity, thereby bridging a critical empirical and theoretical gap in the literature between infancy and early childhood.

Before describing research on capacity and working memory, however, it is important to note that there has been much debate regarding how to define working memory and which laboratory tasks test working memory versus, for instance, short-term memory. Those researchers that espouse the short-term memory view assert that a task must involve manipulation of information, not just storage, in order to qualify as a working memory task. From this perspective, a secondary task must be used—for example, solving equations between each word in a list-memory task (e.g., Anderson, Redere, & Lebiere, 1996). Others take a broader view of working memory, suggesting that information held in memory must simply be used in service of a specific task (e.g., reciting an ordered list). Within this perspective, researchers often choose to measure the "span" of working memory.

In this thesis, I will adopt a third alternative that focuses on the active maintenance of information once input has been removed. A central component of this view of working memory is that using information in service of other cognitive operations requires *stability*. That is, the cortical system actively maintaining the required information must keep that neural representation stable to support operations on this information conducted by, for instance, some other cortical system. This requires stability in the face of neural noise, stability in the face of potentially "distracting" inputs, and so

on (see Chapter 2 for further discussion, see also Spencer, Perone, & Johnson, in press). Given that this view differs from the common definitions of working memory in the literature, I will review results from the various perspectives above; note, however, that "working memory" performance varies considerably between tasks with different demands. This variation across studies highlights the need for specificity regardless of the particular view of working memory that is adopted. One way to achieve such specificity is to organize concepts using formal models (see Miyake & Shah, 1999, for steps in this direction). Formal models require theorists to be explicit about the mechanisms they propose to underlie performance, and the assumptions for how these mechanisms interact in service of performance. As such, they serve as a test of the internal consistency of the theory, as well as allowing direct examination of how changes in these mechanisms may affect performance (see Mareschal & Thomas, 2007, for discussion).

As I will show below, moving to the level of a formal theory of the processes that underlie working memory can clarify empirical discrepancies in estimates of capacity across tasks in the literature. Formal theory can also reveal critical conceptual gaps in common terms such as "capacity" and can shed new light on the mechanisms that give rise to changes in working memory abilities over development. This is accomplished by reuniting cognition and behavior across multiple timescales, explicitly linking real-time memory processes with the behavior produced in working memory tasks and tying changes in these processes to mechanisms of learning and development. It is useful in this context to note that the formal theory adopted in this thesis derives from dynamic systems theory, which emphasizes the context-specific nature of behavior. This approach is particularly well-suited to the question of how working memory capacity changes over development because capacity has been estimated in the literature using different tasks at different ages. From this perspective, it is important to move away from thinking about capacity has an underlying competence tapped in specific paradigms. Rather, capacity is

intricately tied to the second-to-second processes that give rise to working memory representations, and how such representations are probed in the task setting.

Verbal Working Memory

In much of the early research following Baddeley and Hitch's (1974) proposed working memory system, the focus was on the capacity of the verbal component of working memory in adults (Baddeley, 1981). Verbal working memory tasks are typically tests of memory span—a series of words are presented to the participant, who must then recite the words back, in the same order, to the experimenter. The number of words in the span increases one at a time until the participant can no longer repeat them correctly, and the longest span that was reliably recalled[2] is taken as the capacity of verbal working memory (e.g., Dempster, 1981). In his review, Dempster (1981) reports adults' span as approximately 6 items, averaged across digit, word, and letter span tasks from a number of studies; he also cites individual differences ranging from 4 to 10 items in these tasks.

Although span tasks are one of the most common approaches to studying the capacity of verbal working memory, they are not without limitations. First, the capacity of verbal working memory depends on the details of the items being remembered: capacity tends to be lower for long words than for short words (e.g., Baddeley et al., 1975). One particularly striking demonstration of this point was provided by Ellis & Hennelley (1980) who found that bilingual participants would show higher capacity estimates when tested with English digits (about 6.5 items) versus Welsh digits (about 5.75 items), since Welsh number-words are longer than English number-words.

---

[2] Note that different researchers adopt different criteria for what constitutes reliable recall; as mentioned above, some require participants to be correct on only 50% of trials, whereas others determine span from the proportion of items recalled from long (e.g., "supraspan") lists. See Dempster (1981) for discussion.

A second difficulty in the assessment of the capacity of verbal working memory is that people may group or "chunk"[3] items in memory to increase their retrieval (see Dempster, 1981, for review). For example, in a digit span task, participants may group 3, 4, or 5 digits together, as they commonly do when remembering a phone number or ZIP code. Once digits are grouped in this way, the groups become the units of memory; what was previously a list of 10 digits may be remembered as 3 groups, thus making capacity appear larger than if the items were not chunked in memory. In an empirical test of this hypothesis, Ryan (1969) presented adults with sequences of 9 digits—beyond the typical adult capacity—but grouped these items into sets of 3 (resulting in 3 groups) during presentation in some conditions. Adults in the grouping conditions performed up to 23% better than adults in the ungrouped control condition (although different grouping techniques led to different levels of improvement). Thus, grouping the items during presentation did indeed help adults recall more items at test.

These details of performance provide important caveats for interpreting measures of the capacity of verbal working memory—the particulars of the task and participant strategies may alter the number of items that appear to be held in working memory. Nevertheless, verbal working memory capacity still seems to measure some valid psychological construct, as it often correlates positively with other measures of verbal intelligence or performance. For example, Daneman & Carpenter (1980) showed that word span and reading span both correlated with reading comprehension measures, including performance on the verbal section of the Scholastic Aptitude Test (see, e.g., Cowan et al., 2005, for similar correlations over development).

---

[3] Dempster (1981) uses the term "chunking" to refer only to cases in which grouping results in a familiar sequence, such as chunking the letters F-B-I.

More recent work on verbal working memory has explored the developmental origin of capacity limits. A summary of research in this area is shown in Table 1. Research has demonstrated that children tend to have lower verbal working memory capacity (in span tasks) than adults, and that capacity increases from about 2.5 items at 2 years to about 5.25 items at 12 years of age (Dempster, 1981). Note, however, that estimates vary by task; for example, these averages reported by Dempster were derived from digit span, word span, and reading span tasks. Using only a digit span task, Isaacs and Vargha-Khadem (1989) reported spans of 5.2 at 7 years to 6.7 at 15 years. In addition, developmental change through this period is not necessarily linear; the rate at which span increases is relatively rapid in early childhood and slows after the age of 7 years (see, e.g., Dempster, 1981, for review). Interestingly, limited capacity might have adaptive consequences early in development (see Chapter 5 for further discussion). For instance, in language development, the "less-is-more" hypothesis (e.g., Newport, 1990) suggests that children's limited memory of variable input actually helps them learn a regularized grammar.

As with studies on adults' verbal working memory capacity, developmental findings are not entirely straight-forward. The same caveats from adult research apply to children: the length of words affect span, and children may adopt strategies to increase how many words they remember at once. Indeed, some researchers suggest that it is the development of such strategies that drives developmental changes in verbal working memory capacity (see Dempster, 1981, for review). Research suggests that children younger than 7 years of age do not spontaneously rehearse items being held in working memory, as evidenced by similar findings in conditions with and without verbal load. Once children employ rehearsal in memory, there are still important developmental changes to achieve adult-like performance. First, when children do begin rehearsing items in memory, their speed of rehearsal is not as fast as adults'. As a result, more time passes between the rehearsal of an individual item, and items are more likely to decay in

memory. Second, after children begin rehearsing somewhat automatically, they are still not likely to spontaneously group items in memory to increase span (for review, see Dempster, 1981).

These types of differences between children's and adults' verbal working memory performance led Dempster (1981) to summarize 10 factors that have been proposed to contribute to developmental and individual differences in performance.[4] He classified these factors as comprising two categories: four were strategic variables, and six were non-strategic variables. Although many researchers propose that strategic variables (e.g., rehearsal, chunking) are a primary source of developmental improvement in working memory tasks, there is little direct evidence regarding how or why strategies might change over development. For non-strategic variables (e.g., speed of processing, resistance to interference), on the other hand, theorists have put forth some mechanisms that may produce developmental change. In particular, Pasual-Leone (1970) and Case (1995) propose that mental resources, referred to as M-space or M-power, are changing over development. Some researchers suggest that M-space is increasing over development, whereas others believe that the efficiency with which M-space is used is the source of developmental improvement. In either case, the primary mechanism proposed to cause this developmental change is an increase in myelination. According to Case (1995), this is a likely source because myelination not only increases the speed of neural conductivity, but also reduces interference through electrical insulation. Although it seems intuitive that these changes would produce either an increase in resources available or improve the efficiency of the use of these resources, no specific process has

---

[4] Interestingly, Dempster (1981) concludes that only one of these ten proposed mechanisms—the speed and accuracy of item identification—has shown strong empirical support as a factor in developmental changes in span tasks.

been proposed by which increased neural conductivity would improve performance in span tasks.

To summarize, research on verbal working memory has shown an increase in capacity over development, with dramatic increases in early childhood that slow down after about 7 years of age. For children and adults, a number of factors other than capacity may affect performance, including details of the stimuli as well as encoding or rehearsal strategies. Changes in these types of strategies are one proposal to account for developmental improvement, although no specific mechanism has been offered to account for why children first adopt and then improve in their strategy use. In addition, other theorists have suggested a more "maturational" explanation—that myelination improves the speed and efficiency with which children can perform these tasks. In these cases, however, the specific process by which myelination leads to changes in performance is still undefined.

Visuo-Spatial Working Memory

Researchers have recently begun to explore the capacity of the visuo-spatial sub-system of Baddeley & Hitch's (1974) working memory model, with more of a focus on developmental change. A summary of results from studies of visuo-spatial working memory is presented in Table 1. Early tests of this system took two general forms. First, researchers used visual tasks with items that were very likely to be verbally recoded and rehearsed—like digits, letters, or easily-named objects—and required verbal report (e.g., Hitch, Halliday, Schaafstal, & Schraagen, 1988; Hitch, Woodin, & Baker, 1989). Second, researchers used spatial tasks that presented a sequence of locations, as in the Corsi block task (Corsi, 1972; Milner, 1971) in which participants must reproduce a series of spatial locations. Note that both of these approaches utilize span measures, in which items must be recalled in the correct sequence.

Not surprisingly, performance in the first type of task correlated highly with verbal working memory measures (Hitch et al., 1989). When children were tested in these tasks, however, an interesting developmental change was uncovered: young children (typically younger than 7 years) did not spontaneously recode the stimuli verbally. This was determined by including a verbal load in the task (e.g., repeating a nonsense phrase like "blah blah blah" throughout the duration of each trial), which did not disrupt performance in young children, but did impair older children's performance (Hitch et al., 1989). This developmental change corresponds to the age at which children begin to show spontaneous rehearsal in verbal tasks (as described above), so perhaps these findings reflect the same phenomenon—even if children did recode items verbally, if they do not rehearse, they would gain no benefit from recoding.

In the Corsi block task (Corsi, 1972; Milner, 1971), the experimenter and participant sit facing each other with a display of 9 asymmetrically arranged blocks on a table between them. The experimenter taps a sequence of blocks in the display, and the participant then taps the same blocks in the same order. Testing begins with a single block (span length 1), then increases in span length over trials until the participant's performance falls below a pre-determined threshold (e.g., correct performance on two out of three trials at a given span, Logie & Pearson, 1997). In this task, children show an increase in span over development from about 2.5 blocks at 5 years of age (Logie & Pearson, 1997) to about 5.6 blocks at 15 years of age (Isaacs & Vargha-Khadem, 1989[5]); note that estimates were higher in a recognition version of the task, (Logie & Pearson, 1997; see Pickering, 2001, for review). Adults show spans of about 6 blocks in the standard task (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999).

---

[5] The children in this study were tested in both digit span and the Corsi block task; at every span length, children recalled more digits than blocks when recalling items in the order they were presented. Interestingly, when required to recall items backward, children recalled more blocks than digits.

The Corsi block task requires remembering which of the blocks has been tapped (location memory) as well as the order (sequence memory), and may also depend on motor memory—on each trial, when the participant responds, he or she is executing a specific motor pattern that may form the foundation of future responses. This raises the question of whether memory depends on actions, locations, or both and, correspondingly, what cognitive systems this task measures. For instance, Logie (1995) has argued that the visuo-spatial sketchpad should be further divided into the inner scribe and the visual cache. According to Logie, the inner scribe represents spatial and movement information, as is required in the Corsi block task. The visual cache, on the other hand, stores information about the visual form and color of objects. This suggests that there may be a component of visuo-spatial memory that is not measured in the Corsi block task.

To try to avoid the problems of verbal recoding and sequencing or motor memory, a third type of task—the visual patterns task—was developed to measure visual working memory (Della Sala, Gray, Baddeley, & Wilson, 1997; Wilson, Scott, & Power, 1987). In this task, participants view a grid of blocks, half of which are unfilled and half of which are filled with a single color (e.g., all black, all red). The grid is briefly presented to participants, then removed for a short delay, and finally a blank grid is presented. For the standard recall response, participants must indicate which of the blocks in the grid were previously filled. Some studies have used a partial recall or recognition responses. In the partial recall version of the task (called recognition by Logie & Pearson, 1997), the grid is presented with one of the blocks changed (either from filled to unfilled or unfilled to filled), and the participant must indicate which of the blocks changed. For both response types, memory capacity is determined by the largest number of filled blocks (i.e., half the grid size) on which the participant could perform reliably—as in verbal working memory research, each researcher determines what constitutes reliable. The capacity of visuo-spatial working memory has been estimated at approximately 2.5 blocks at 5 years of age, increasing to about 8 blocks at 12 years of age (again, estimates were higher in a

recognition version of the task; Logie & Pearson, 1997). Adults show a capacity of about 9 blocks in the standard recall task (Della Sala et al., 1999).

The visual patterns task was developed to differ from previous tasks in three important ways: the stimuli are not easily verbally recoded; performance does not depend on sequential ordering; and the stimuli and response are not strictly spatial. Although the first two of these goals were accomplished, the third is more uncertain. This task does successfully remove the spatial *movement* component of the Corsi task; however, the task may still be considered a spatial task because it depends on memory for locations—where the filled blocks were in the grid. However, some evidence suggests that the Corsi block and visual patterns tasks depend on at least partially separable memory systems. Specifically, Logie and Pearson (1997) showed that the two tasks showed different developmental trajectories, with better performance and earlier proficiency in the visual patterns task, especially if a recognition response was required rather than recall. Moreover, patients with visual or spatial deficits showed specific impairments on the task that corresponds to their deficit (Della Sala et al., 1999). Finally, Della Sala et al. (1999) showed selective impairment of performance on the Corsi block task from a spatial interference task, and impairment of performance on the visual pattern tasks from a visual interference task.

Although the Corsi block and visual patterns tasks show some dissociation, both depend on spatial components of memory, suggesting they are better suited to test the inner scribe rather than the visual cache proposed by Logie (1995). One additional task has been developed to test memory for object features (i.e., the visual cache) rather than spatial positions (i.e., the inner scribe). Vicari and colleagues (Vicari, Bellucci, & Carlesimo, 2003) compared the Corsi block task with recognition of complex objects. In the visual recognition task, a complex object was presented for 2 s on a computer screen, then after a short delay nine similar objects were presented at the bottom of the screen and the participant indicated which of the objects matched the target object. As with other

visuo-spatial task, the number of items to be remembered was increased across trials to arrive at a capacity estimate. In both the Corsi block and object recognition tasks, the authors used the number of items in the last correct trial as the measure of span or capacity.

These authors found that performance on the Corsi block task was more advanced over development, with higher span and faster improvement with age. In addition, they showed that a population with known spatial deficits (Williams syndrome) was deficient in the Corsi block task but not the visual recognition task, relative to mental-age matched controls. Together these experiments support the separable contributions of visual and spatial components within visuo-spatial working memory, as suggested by the distinction of the visual cache and inner scribe by Logie (1995). Interestingly, however, results from Vicari and colleagues (2003) and Logie and Pearson (1997) show opposite patterns of change over development in spatial versus visual tasks: Vicari et al. found the Corsi block task to be easier and improve more quickly over development than performance in the object recognition task, whereas Logie and Pearson found the Corsi block task to be more difficult and improve more slowly than performance in the visual patterns task. This difference across tasks is likely due to differences in the tasks chosen to measure the non-spatial component of visual working memory: the object recognition task may be more difficult than the visual patterns task, and the two tasks may diverge over development due to the spatial nature of the pattern task.

Pickering (2001) summarized the general developmental findings in visuo-spatial working memory and reviewed proposed developmental mechanisms. After ruling out contributions from verbal recoding, she identified four potential causes for developmental change: knowledge (with specific regard for familiar chunks that could be formed), processing strategies, processing speed, and attentional capacity. As in the verbal domain, no specific proposals have been put forth for how and why the use of specific strategies, including chunking, might develop. Both processing speed and attentional capacity, on

the other hand, are suggested to be linked to neurological maturation. Again, however, no specific processes for how maturation would affect either factor—and then how these factors would improve capacity—have been proposed.

A New Approach to the Study of Visual Working Memory

The previous sections highlighted some limitations in our understanding of the nature and developmental origins of capacity limits. In verbal working memory research, one challenge is the dependence of performance on the details of the stimuli—results seem to depend more on speed of rehearsal or processing than on storage, especially early in development when children do not spontaneously rehearse. In addition, verbal span tasks depend on sequencing, which may be a separable component of working memory. In studies of visuo-spatial working memory, early tasks relied heavily on the verbal system, showing the same developmental transition when children begin spontaneously rehearsing. Tasks that were developed to be strictly spatial, like the Corsi block task, also rely on sequencing as the primary measure of span. One task that avoids the sequential nature of the verbal and spatial tasks is the visual patterns task. Although performance on this task seems to be at least partially separable from the Corsi block task, both involve the use of spatial locations and, therefore, depend critically on developmental changes in children's ability to remember locations in a spatial frame of reference. Thus, nearly all previous research on working memory capacity depends on rehearsal (including strategies like chunking), sequencing, or spatial cognitive abilities.

Recently, however, researchers studying visual working memory (VWM) have adopted a new technique that minimizes the influence of these factors. Beginning with Luck & Vogel's (1997) seminal paper, *change detection* has been the method of choice for testing memory capacity for individual visual features, as well as for multiple features bound into object representations (see Vogel et al., 2001, for discussion of how this task was developed). Figure 1 shows a standard change detection trial. First, a memory array

is shown briefly (100-500 ms) to the participant, followed by a short delay (250-1000 ms). Then, a test array appears in which either all of the objects match the memory array, or the feature(s) of one object has changed to a new value. The participant reports whether there was a change in the display. This task avoids the influence of the three primary limitations of other visual tasks. First, the brief presentation of the items and short delay reduces the likelihood of verbal recoding and rehearsal (see Vogel, Woodman, & Luck, 2001). Second, the items are presented simultaneously for both encoding and testing, so memory for sequence is not a factor. Third, location is not a relevant dimension (i.e., it is not part of the response and does not change from the memory array to the test array), so the influence of spatial memory is minimized. Thus, the change detection task may provide a more direct test of the storage of non-spatial visual information than previous tasks.

Findings across a host of studies have shown that adults have a capacity of approximately 3-4 simple items (e.g., colored squares or oriented bars) in change detection. As in research on verbal working memory, VWM research has shown that visual information can be grouped or chunked by combining multiple features of an item (e.g., color and shape) into a single integrated object. It is these integrated objects, rather than the individual features, that serve as the units of representation in VWM. Specifically, if adults must remember multi-featured objects (e.g., colored oriented bars), they show no significant decrement in their performance relative to when they must remember only colors or only line orientations (although this depends on the details of the task; see Johnson, Hollingworth, & Luck, 2008 for disucssion). While these findings have been replicated widely, they are not without controversy. For example, recent studies have shown that the details of the object features held in working memory can limit the capacity of this memory system, suggesting that working memory utilizes a pool of limited resources that can modify the number of “slots” available in a given task

(Alvarez & Cavanagh, 2004). I will return to this issue for further discussion in Chapters 2 and 5.

Although the exact nature of the cognitive representations that underlie VWM is under debate, researchers agree that VWM is a capacity-limited system. As in the verbal working memory literature, the next step for research in this area was to explore the developmental origins of these capacity limits. This required finding tasks that can be used across a broad range of ages. Some researchers adapted the adult procedure for use with children (the standard approach in the study of verbal working memory). This has proved challenging, particularly with young children because change detection generally requires adult subjects to complete dozens of trials to derive robust capacity estimates. Other researchers have designed a more child-appropriate method believed to measure the same construct. At present, however, it is unclear whether these new methods assess the same underlying cognitive processes tested in the adult change detection task.

The standard change detection task has been used, with some modification, with children between the ages of 5 and 12 years of age (Cowan et al., 2005; Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; Riggs, McTaggart, Simpson, & Freeman, 2006). These studies have shown that, as in other developmental working memory studies, children tend to have lower capacity than adults (see Table 1). Moreover, capacity increases with over this period of development, reaching adult-like levels around 10 years of age. In particular, Cowan et al. (2005) estimated capacity as the smallest set size at which children responded correctly on at least 75% of trials. Using this method, they computed capacity estimates of approximately 3.5 items for 8-year-olds, improving up to around 4.8 items for 12-year-olds. Cowan et al. (2006) also tested children (ages 8 to 12 years) in change detection, but did not compute capacity in these studies. These authors did report developmental improvements in children's overall percent correct, which would correspond to a developmental increase in capacity. In a study including younger children, Riggs et al. (2006) used the Pashler (1988) formula to estimate VWM capacity

of approximately 1.5 items for 5-year-olds, 2.9 items for 7-year-olds, and 3.8 items for 10-year-olds. Compared to the measures discussed in the previous sections, these capacity estimates are relatively low for these age groups (see Table 1).

Other research on the development of VWM has used a different type of change detection task, known as a flicker or change blindness task (Shore, Burack, Miller, Joseph, & Enns, 2006).[6] In this task, two photos or drawings of an object are presented side by side on two computer screens for 250 ms, followed by blank screens for either 50 or 250 ms, then the re-presentation of the images for another 250 ms, and so on. As the images flick on and off, some detail of the image on one screen will change back and forth which each flicker. The participant is asked to indicate which of the two displays contains the changing image. Performance is measured as the proportion of trials on which the correct side is chosen, as well as the reaction time to make the correct choice. Results showed that both accuracy and reaction times improve between 6 and 10 years of age and into adulthood (Shore et al., 2006). Note that the capacity of VWM cannot be inferred from this task since each image only contains a single complex object; however, percent correct suggests a generally linear improvement between 6 and 10 years (for color changes, performance increased from about 95% to 98% for photos and from about 86% to 97% for drawings during this period of development).

A variant of this task has also been developed to test infants in a preferential looking paradigm; note that this is the first study to probe developmental changes in

---

[6] One study has used a flicker paradigm to test capacity in adults (Rensink, 2000). This study tested memory for multiple items with either two simple shapes (rectangles oriented vertically or horizontally) or two colors (black and white). In this task, only one display was presented in which an item always changed, and capacity was estimated from search time to identify which item was changing. Capacity estimates were quite high, around 5.5 items for rectangles and at least 8 items for colors. These findings are somewhat controversial, though, because all of the items belonged to two highly distinct categories; this may have encouraged grouping by participants, which could inflate capacity estimates.

working memory capacity during this period of development, and therefore is the first study to be able to address the developmental origins of capacity in infancy.[7] As in the flicker paradigm described above, images were presented on two side-by-side displays and blinked on and off over the course of the trial. To adapt this task to test the capacity of VWM, Ross-Sheehy, Oakes, & Luck (2003) presented arrays of colored squares, like those used in adult change detection tasks, on each display. A schematic of this task is shown in Figure 2; in one display (randomly determined across trials), one of the items in the array would change after each delay. The amount of time the infant spends looking at each display is tabulated across each 20-second trial. Preference scores are then calculated as the proportion of time on each trial that the infant spent looking at the changing display.

The rationale behind this method is that if the number of the items in the display is within the infant's VWM capacity, the infant will remember the items over the brief delay, will notice that the items are changing in one display, and will prefer to look at that display (given infants' general bias to look at novelty in preferential looking task, Fagan, 1970, 1973). If the number of items is beyond the infant's capacity, however, the infant will not be able to detect the difference across sides and should show no preference. Using this method, Ross-Sheehy et al. (2003) found that 6-month-old infants only showed reliable change preferences for displays with a set size of 1 item. By the age of

---

[7] Rose, Feldman, and Jankowski (2001) also developed a span type task to measure visual short-term memory span (as they refer to it) in infants. Although the design of this task was adapted from span tasks used in traditional studies of verbal and visuo-spatial working memory, the adaptations required to make the task appropriate for infants make the results difficult to interpret. In particular, the stimuli were real, complex objects (and therefore may be remembered as multiple "items"; see discussion of Alvarez & Cavanagh, 2004, below) and were presented individually to infants for up to 10 s per object with inter-stimulus intervals of approximately 4 s each. Thus, by the time all of the items (up to four per trial) had been presented, nearly a minute may have elapsed since the first item was presented. As such, long-term memory likely played a significant role in performance, and the contributions of working memory cannot be clearly evaluated.

10 months, however, infants showed preferences when the set size (i.e., number of objects in the array) was as high as 4 items. These findings suggest that between the ages of 6 and 10 months, infants VWM capacity is rapidly increasing to adult-like levels.[8]

Comparisons across these developmental studies raises a fundamental question: why is there a discrepancy in capacity estimates over development, with 10-month-olds showing a capacity of 4 items in preferential looking, and 5-year-olds showing a capacity of only 1.5 items in change detection? What developmental change could account for this pattern? Riggs et al. (2006) proposed that this divergence between tasks reflects a difference in what each task measures. Specifically, they contend that the preferential looking task measures something other than visual working memory capacity—possibly a more passive type of memory—or provides an index of working memory that does not effectively scale with capacity. This 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, and Chapter 2 for further discussion).

To understand why performance might differ across these two tasks, it is important to consider how they differ from one another, not only in the behavioral demands, but also in the structure of the task. First, these tasks require different behaviors, looking back and forth over 10 to 20 second trials versus generating an

---

[8] Recent data suggests that the developmental transition from capacity of 1 to 3 items occurs between 6 and 7.5 months of age (Oakes, Messenger, Ross-Sheehy, & Luck, in press). Note, however, that only 10-month-old infants have been tested at set sizes of 4 items or higher. Older infants (13 months) have only been tested on set sizes of 1-3 items (Ross-Sheehy et al., 2003, Experiment 1), therefore it is unknown whether capacity remains the same or increases after 10 months.

explicit decision after the presentation of the test array on each trial. In addition, they use different measures to infer capacity, above-chance change preference scores versus the pattern of decisions across dozens of trails. Moreover, the different behaviors and task structures call into question the relationship between VWM capacity and the behaviors being measured. Thus, it is important to consider both tasks in detail to understand why they produce paradoxical findings over development.

Change Detection – Task Analysis

A single trial in change detection requires a participant to (1) quickly encode the memory display, (2) maintain that information over a brief delay, (3) compare the items in memory to the items in the test array, and (4) decide whether the items match. Research suggests that adults encode items in an all-or-none fashion (e.g., Zhang & Luck, 2008): if the number of items in the array exceeds capacity, some items are not encoded (rather than encoding partial information about all items). It is assumed that all items have an equal likelihood of being encoded, that is, if one has a capacity of 4 items and 5 items are presented in the array, there is an 80% chance that any given item will be encoded. Assuming these items are maintained accurately over the delay, when the test array is presented, the 4 items that are being held in memory will be compared to the corresponding items in the array. Given that each item has the same likelihood of changing at test, there is a 20% chance on any given trial that the item that is <u>not</u> being held in memory will change. Therefore, performance is expected drop sharply when the array size surpasses capacity.

Pashler (1988) developed a formula to translate a participant's hit rate (*H*; the proportion of correct responses on *change* trials) into an estimate of capacity. This formula begins with the assumption that a participant will hold a discrete number of objects in memory (*k*) and that memory of those items remains accurate over brief delays. When these items are compared to the test array, the participant will respond "change"

only if the change occurred in one of the item being held in memory. If the array size (S) is higher than capacity, then there is a k / S likelihood of the change item being held in memory. In addition, a participant may correctly guess that there was a change (g) on a small number of trials in which the change was not actually detected (i.e., because the change item was not being held in memory). This guessing rate is inferred from "false alarm" trials—*no change* trials on which the participant incorrectly responded "change". Because guesses would only occur on the trials in which the change item was not being held in memory ($1 - [k / S]$), the contribution of guesses is $[1 - (k / S)] * g$. Thus, the resulting formula is:

$$H = (k / S) + [1 - (k / S)] * g, \tag{1.1}$$

where the first component captures "genuinely" correct trials, and the second component captures "lucky guesses". To calculate capacity, then, the formula rearranges as:

$$k = S * (H - g) / (1 - g). \tag{1.2}$$

This formula has been generally accepted as an appropriate estimate of capacity, although Cowan (2001) later updated the formula to also take into account correct performance on *no change* trials (although this formula only applies when one item is cued at test).

In summary, each change detection trial requires the encoding items from the sample array, maintaining them over delay, comparing them to the test array, and then generating a decision based on this comparison (and, perhaps, some guessing strategy). Capacity is computed under the assumption that performance will be near ceiling for array sizes within a person's capacity, and that performance will drop for larger arrays.

Preferential Looking – Task Analysis

Preferential looking differs from change detection in three important ways. First, although each array of items is presented only briefly (500 ms, similar to change detection), the arrays are presented repeatedly over the 20-second trial (i.e., for a total of 26 presentations). Thus, rather than having a single opportunity to encode the items on

each trial, the infant is given repeated opportunities to encode the arrays. As a result, forming a memory representation of the array may be completed over multiple presentations. Moreover, repeated presentation of the items may allow support from long-term memory to contribute to the formation of working memory representations (for discussion, see Perone, Spencer, & Schöner, 2007).

Between the stimulus presentations are brief (250 ms) delays in which both displays are blank. The assumption is that infants maintain their memory for the array during this delay, similar to the maintenance of memory for the sample array in change detection. Evidence suggests that infants' maintenance of information over this delay is an important component of capacity; specifically, if the "blank" delays are removed from the task (i.e., one display shows a static image of colored squares, while the second display has a different color change every 500 ms), then 6-month-old infants will show a preference for the changing display at set size 3 (Ross-Sheehy et al., 2003, Experiment 4). Thus, removing the delay between the stimulus presentations allows young infants to detect changes at higher set sizes—note, however, that removing the memory delays is confounded with a difference in onsets across displays in this design.

The second important difference between preferential looking and change detection is that the stimuli in the change detection task appear on a single display, whereas preferential looking includes arrays presented on two side-by-side displays. As a result, at all times in the preferential looking tasks, the infant is deciding which display to fixate, as well as whether to fixate a display at all; Ross-Sheehy et al. (2003) report average looking times of only 9-12 s per 20 s trial. Although looking time is often used to infer underlying cognitive processes, little is known about the real-time processes underlying infant looking behavior in these contexts, that is, the processes that drive looking from one fixation to the next (for discussion, see Goldberg, 2008; Perone et al., 2007).

This highlights the third important difference between the two tasks—it is not clear how to map looking behavior over the course of each 20 s trial to measures like percent correct (or percent hits). Indeed, inferring capacity from a looking time measure is quite complicated. The assumption is that infants' preference scores—whether they look at the Change display longer than the No-Change display—is an indication of whether they were able to hold the items in a single display in memory. It has been argued, however, that this may not be a valid assumption (Riggs et al., 2006).

To illustrate this point, note that infants typically fixate either display approximately 50% of the time in the preferential looking task (Ross-Sheehy et al., 2003). This means that only about 12-14 stimulus presentations are fixated in the infant preferential looking task. Since stimuli are presented on both screens, however, there is an approximately 50% chance that the infant will be fixating the Change display during any particular fixation. This reduces the number of stimulus presentations that can drive a change preference to about 6 or 7. If an infant has a capacity of 1 item and is being tested in set size 2, there is a 50% chance of remembering the item that will change when fixating the Change display. Thus, infants might notice a change on about 3 stimulus presentations out of the 12-14 that are viewed. The question is whether this enough to drive a change preference. This question is further complicated by the fact that infants' fixations of the displays are not necessarily synchronized with the stimulus presentations—they may look back and forth at any time within a presentation, or fixate a single display for multiple presentations. Thus, without analyzing infants' performance on a fixation-by-fixation basis, we cannot know how many changes were actually viewed by each infant. This example highlights that the generalization from looking time to VWM capacity is not straight-forward.

Summary

Research from a variety of tasks suggests that working memory—both verbal and visual—has a limited capacity. In addition, there are clear increases in capacity over development (see Table 1). Although results in verbal and visuo-spatial working memory may be affected by factors other than capacity, including strategies such as rehearsal and chunking, new studies in VWM have minimized these contributions and still show developmental changes. Across domains, however, there have been few attempts to understand the mechanism driving developmental changes in working memory capacity. Explanations have typically pointed to the development of strategy use or neural myelination, although neither type of explanation has been linked to a specific process to capture how these changes produce changes in capacity.

In addition, a particularly puzzling developmental result has emerged in the VWM literature: there is a discrepancy between capacity estimates at 10 months (3-4 items in preferential looking) and 5 years (1.5 items in change detection). A detailed analysis of the two tasks used with infants and young children suggests that these different capacity estimates may be driven by differences across tasks. Nevertheless, as is the case with the broader working memory literature, it is not clear what process actually leads to these task-specific differences in capacity estimates. To understand the relationship between these tasks and estimates of capacity over development, we need a theory that can address how estimates of capacity are linked to the behaviors in each of these tasks, and how this system might be changing over development. The goal of this project is to move toward such a unified theory.

To this end, the next chapter reviews limitations of the current theoretical approaches to these tasks, and proposes that an alternative approach grounded in a dynamic neural field model can both explain previous findings using these tasks as well as generate novel predictions for how these tasks should be related within-subjects over development. I will describe how this type of model has previously captured performance

in the adult change detection and infant preferential looking tasks. Next, I will show model simulations demonstrating how a unified model can perform both tasks, yet produce different capacity estimates across the two tasks. From this model, I generate a set of three specific predictions which are tested empirically in Chapter 3; first, by adapting the change detection task for use with younger children (3-, 4-, and 5-year-olds), and, second, by testing the same participants in the infant preferential looking task. In Chapter 4, I test a fourth prediction—that the unified model can quantitatively capture these data and changes in capacity across tasks and over development using a previously proposed developmental hypothesis, the Spatial Precision Hypothesis (Schutte, Spencer, & Schöner, 2003; Simmering, Schutte, & Spencer, 2008; Spencer, Simmering, Schutte, & Schöner, 2007). Chapter 5 concludes by discussing these findings within the context of the broader literature on the development of working memory.

Figure 1. Events in a single change-detection trial.

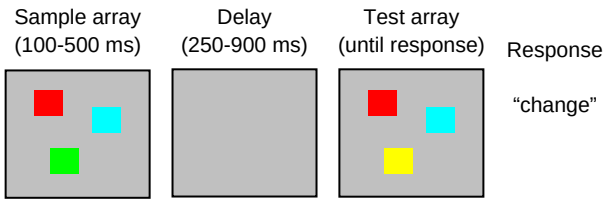


Figure 2. Events in a single preferential looking trial.

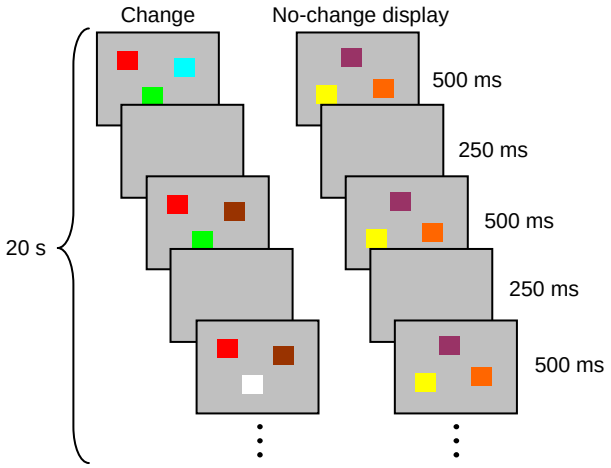


Table 1

Sample of Results from Developmental Studies of Working Memory Capacity

	Age (years)													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Verbal														
Cowan et al. (2005)														
auditory							2.8		3.3		4.0			
counting[a]							2.8		3.6		4.1			
digits[a]							4.7		4.9		5.4			
ignored speech							2.0		2.0					
listening[a]							1.8		2.4		2.7			
running[a]							2.0		2.6		2.7			
Dempster (1981)[b]														
average	2.6			3.9		4.5		4.9			5.3			
digits	2.25			4.35		5		9			6.5			
letters				3.5		4.5		4.9			5.3			
words	3			4		4.2		4.3			4.5			
Isaacs & Vargha-Khadem (1989)														
digits						5.2	5.7	5.8	6.1	6.1	5.8	6.2	6.3	6.7
Visuo-spatial														
Hitch et al. (1989)														
pictures				1.9						3.6				
Isaacs & Vargha-Khadem (1989)														
blocks (Corsi)						4.1	4.7	4.7	5.3	5.3	5.4	5.4	5.4	5.6
Logie & Pearson (1997)														
blocks (Corsi), recall					2.8			4.5			5.2			
blocks (pattern), recall					3.0			6.2			8.0			
blocks (Corsi), recognition					3.7			7.1			8.0			
blocks (pattern), recognition					4.0			10.9			14.8			
Vicari et al. (2003)														
blocks (Corsi)				3.1	3.8	4.3	4.6	4.7	4.8					
complex figures				2.4	2.7	2.8	3.0	3.3	3.4					
Wilson et al. (1987)														
blocks (pattern), recognition				3.7		8.2				14.1				
Visual (change detection)														
Cowan et al. (2005)														
colors[a]							3.6		4.4		4.8			
Riggs et al. (2006)														
colors				1.5		2.9		3.8						

Note. Results reported here were from standard versions of the tasks only; conditions designed to alter performance (e.g., interference added, additional manipulation required, extended delays) are not included.

[a] These results are averaged across two experiments presented in the same paper.

[b] These values are estimated from reviews of multiple other studies; see reference for details.

CHAPTER 2
DYNAMIC FIELD THEORY AND VISUAL WORKING MEMORY CAPACITY

To date, the dominant theoretical perspective for understanding change detection and preferential looking has been Information Processing (IP). Although this perspective has yielded solid empirical progress over the past decade, there remain two specific limitations of the IP explanation of the processes underlying these tasks. First, this approach has yet to resolve whether VWM should be thought of as a discrete form of memory with fixed-resolution "slots" or as a flexible pool of resources in which memory representations vary according to the type of stimuli. Although a number of studies have been designed to address this debate (e.g., Alvarez & Cavanagh, 2004; Zhang & Luck, 2008), neither perspective has fully explained the data that have been generated. Moreover, these two characterizations of VWM are only loosely tied to neural processes and have yet to be rigorously examined in this context.

The second limitation of the IP approach to these tasks is its incomplete account of development. Indeed, some prominent developmental theorists contend that IP is not, in fact, a developmental theory (see Thelen & Bates, 2003). Three specific developmental changes have yet to be captured in the IP account of these tasks: the increase in capacity in preferential looking in the first year of life; the increase in capacity in change detection between the ages of 5 and 10 years; and the discrepancy between findings at 10 months in preferential looking and at 5 years in change detection. In addition to these limitations, even though change detection and preferential looking ostensibly measure the same cognitive system, there has been no explicit attempt in the literature to connect these tasks using a single, specific process-based account.

This project seeks to overcome these limitations by unifying these tasks in a theoretical framework that has already shown promise for capturing the real-time details

of working memory over development—the Dynamic Field Theory (DFT, Spencer, Simmering, Schutte, & Schöner, 2007b; see also, Simmering, Schutte, & Spencer, 2008b). As described in Chapter 1, I am adopting a definition of working memory that focuses on the active maintenance of information after input has been removed. This definition reflects my desire to ground working memory in known neural processes and puts particular emphasis on the real-time stability of working memory states, both of which are captured in the dynamic neural field model described in this chapter. First, I will review the general foundations of this model and how it has been applied to working memory. Next, I will describe the specific application of the model to the maintenance of non-spatial visual information in working memory in both the change detection and preferential looking tasks. Lastly, I will generate a set of specific predictions derived from a unified model that captures performance in both tasks. These predictions are then tested in experiments and simulations in subsequent chapters.

Dynamic Field Theory

The DFT is in a class of theoretical models that fall within the broader theoretical framework of dynamic systems theory. A primary emphasis of dynamic systems theory is that behavior is not rigidly pre-programmed; rather, behavior is emergent, "softly assembled" in the moment in response to the demands of the task, the details of the environment, and the history of the organism. The approach brings a new perspective to the question of the development of working memory because many previous theories attributed differences across tasks and ages to issues of competence versus performance (e.g., Baillargeon & Graber, 1988, regarding the development of the A-not-B error; see also Diamond, 1990a; Diamond, 1990b). Such arguments contend that children may have a particular competence early in development (i.e., the ability to remember where an object is hidden), but their performance only reflects this underlying competence in certain situations (i.e., in looking versus reaching tasks). Although these types of theories

have been popular in developmental psychology, they are essentially unfalsifiable because competencies are not directly testable—any failure in performance can be attributed to task demands that mask competence (for discussion, see Thelen & Smith, 1994).

Within dynamic systems theory, on the other hand, behavior is viewed as multiply-determined, encouraging researchers to explore the various contributions to behavior and to evaluate the robustness of behavior relative to the circumstances required to support them. Just as with competence-performance explanations, this perspective highlights the influence of task details, but also proposes specific processes and mechanisms by which such details should affect behavior. Consistent with this dynamic systems view, a primary focus of the DFT is to unite cognition and behavior by explicitly modeling the connection between underlying memory representations and the overt behavioral responses generated in specific memory tasks.

The DFT arose from a class of models known as neural field models. Amari (1977) first proposed this class of models to capture the real-time dynamics of neural activation in visual cortex. This framework has been broadly applied, accounting for the processes that underlie saccadic eye movements (Kopecz & Schöner, 1995; Wilimzig, Schneider, & Schöner, 2006), motor planning (Erlhagen & Schöner, 2002; Schutte & Spencer, 2007), infants' performance in Piaget's A-not-B task (Thelen, Schöner, Scheier, & Smith, 2001), the dynamics of neural activation in motor and premotor cortex (Bastian, Riehle, Erlhagen, & Schöner, 1998; Bastian, Schöner, & Riehle, 2003), the behavior of autonomous robots (Bicho, Mallet, & Schöner, 2000; Iossifidis & Schöner, 2006; Steinhage & Schöner, 1998), and the development of spatial cognition (see Spencer et al., 2007b, for review).

Recently, the DFT has also been extended to working memory for non-spatial features (Johnson, Spencer, & Schöner, 2008b; Perone, Spencer, & Schöner, 2007). This extension of the DFT can address the first limitation of the IP approach to visual working

memory capacity, that is, the nature of working memory representations. In the DFT, specific values along metric dimensions are represented as sustained peaks or "bumps" of activation in neural fields (see Johnson, 2008, for discussion). Figure 3A shows a simple two-layer neural field model for illustration. In this model, the top layer consists of a single population of excitatory neurons, which are reciprocally coupled to the bottom layer of inhibitory neurons. Along the *x*-dimension in this figure is the continuous dimension (e.g., color) to which both layers of neurons are "tuned". Specifically, neurons coding for similar values along this dimension in the excitatory layer engage in locally excitatory interactions (see solid curved arrow in Figure 3A), as well as projecting excitation to similarly-tuned neurons in the inhibitory layer. These inhibitory neurons, in turn, project broad lateral inhibition back to the excitatory layer (see dashed arrows in Figure 3A). Through this local excitation/lateral inhibition interaction profile, activation in the excitatory layer is able to form sustained peaks of activation.

Patterns of activations within neural fields exhibit different attractor states, that is, stable states around which activation patterns cohere. For instance, in the absence of input, the activation level across the population remains at a stable baseline rate (i.e., the neurons' negative resting potential), indicating that no relevant features are currently "represented". However, when presented with input—for example, the appearance of an object with a particular color in the environment—activation increases for the neurons selectively tuned to this feature value. If input is sufficiently strong, the stable resting state is destabilized and the field moves into an "on" state, characterized by the formation of a localized, above-threshold peak of activation in the field. The location of this peak along the represented dimension (i.e., color) reflects the field's estimation of the metric value present in the task space—that is, the detection of a particular color. Multiple peaks along the dimension would reflect the detection of multiple colors in the environment, with the level of activation of each peak providing an estimate of the relative certainty or salience of each informational source.

These peaks of activation have characteristics that can be specifically applied to capture different cognitive processes. In particular, peaks may exist in a *self-stabilizing* state, in which peaks remain active as long as the input is present; once input is removed, however, activation falls back to the resting state (see Figure 3B, dashed curve). This state can capture perceptual processes that detect items in the environment. Note that this state commonly emerges when local excitation is weak relative to lateral inhibition. If local excitation is strengthened, fields can move into a *self-sustaining* state, in which peaks of activation are maintained above-threshold (i.e., above zero) even after input has been removed (see Figure 3C, dashed curve). 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 neural field—accomplishes two key goals needed for a theory of VWM. First, peaks capture the notion of stability, which is necessary for memory representations to be used in service of behavior or other cognitive tasks. As discussed below, this type of stability plays an important role in the DFT account of development. Second, this approach provides neural grounding to the concept of working memory by illustrating how multiple items could be represented in layers of cortex. The application of neural fields to the representation of stimulus dimensions like color builds off of work by Georgopoulos and colleagues (Georgopoulos, Lurito, Petrides, Schwartz, & Massey, 1989) showing that one can construct neural fields based on functional topography. Specifically, neurons in many areas of cortex are not arranged topographically: two neurons next to one another in cortex might "prefer" very different values along the dimension to which they are tuned. Nevertheless, neural activation at the population level actually obeys topographic relations; that is, even though two neurons might be far apart on the cortical surface, they can still "prefer" similar stimulus values (e.g., similar colors) and therefore be mutually excitatory. By connecting neural field models to functional topography, direct connections can be made between model predictions and neurophysiological measures

(e.g., Bastian et al., 1998; Bastian et al., 2003; Erlhagen, Bastian, Jancke, Riehle, & Schöner, 1999; Jancke et al., 1999; Jancke, Erlhagen, Schöner, & Dinse, 2004).

Change Detection in the Dynamic Field Theory

As described above, the 2-layer model shown in Figure 3A can enter a type of perceptual state when a self-stabilizing peak forms; this type of architecture can also enter a working memory state when a self-sustaining peak forms. Taken separately, these states capture important components of the VWM tasks described above: detecting inputs in the environment and maintaining specific information over the course of brief delays. In order for a single model to perform a change detection task, however, these processes need to be integrated. This integration is achieved by combining two 2-layer models—one tuned to act in a more perceptual state and one tuned to act in a working memory state—into a single three-layer model.

This three-layer architecture, proposed by Johnson, Spencer, and Schöner (2008b) to capture change detection performance, is shown in Figure 4. The model consists of an excitatory perceptual field (PF(u); Figure 4A), a layer of inhibitory interneurons (Inhib(v); Figure 4B), and an excitatory visual working memory field (WM(v); Figure 4C). In each field, the x-axis consists of a collection of neurons tuned to particular colors, and the y-axis shows each neuron's activation level. These layers pass excitation and inhibition as indicated by green and red arrows, respectively, to capture the locally excitatory/laterally inhibitory interactions described above. In addition, to capture the *change/no change* decisions required in the change detection task, a response system is also shown that contains two self-excitatory, mutually-inhibitory nodes (Figure 4). One neuron receives summed excitatory activation from PF(u) to generate *change* responses, while the other receives summed activation from WM(w) to generate *no change* responses (see Simmering et al., 2008b; and Simmering & Spencer, 2008 for a similar process in position discrimination). Activation projects to these nodes when a decision is

required in the task (i.e., at a "go" signal during the presentation of the test display), and they are coupled in a "winner-take-all" fashion, such that only one node will sustain activation above-threshold, thereby generating a response. Thus, the model's response is the result of competition between activation projected from the two excitatory layers, PF(u) and WM(w).

Model equations. Activation in the perceptual field, PF(u), is captured by:

$$\tau \dot{u}(x,t) = -u(x,t) + h_u + \int c_{uu}(x - x')\Lambda_{uu}(u(x',t))\,dx' - \int c_{uv}(x - x')\Lambda_{uv}(v(x',t))\,dx' + S(x,t) + noise \qquad (2.1)$$

where $\dot{u}(x,t)$ is the rate of change of the activation level for each neuron across the spatial dimension, x, as a function of time, t. The constant τ determines the time scale of the dynamics (Erlhagen & Schöner, 2002). The first factor that contributes to the rate of change of activation in PF(u) is the current activation in the field, $-u(x,t)$, at each site x. This component is negative so that activation changes in the direction of the resting level h_u.

Next, activation in PF(u) is influenced by the local excitation / lateral inhibition interaction profile, defined by self-excitatory projections, $\int c_{uu}(x - x')\Lambda_{uu}(u(x',t))\,dx'$, and inhibitory projections from Inhib(v), $\int c_{uv}(x - x')\Lambda_{uv}(v(x',t))\,dx'$. These projections are defined by the convolution of a Gaussian kernel with a sigmoidal threshold function. In particular, the Gaussian kernel is specified by:

$$c(x - x') = c\exp\left[-\frac{(x - x')^2}{2\sigma^2}\right] - k, \qquad (2.2)$$

with strength, c, width, σ, and resting level, k. The sigmoidal function is given by:

$$\Lambda(u) = \frac{1}{1 + \exp[-\beta u]}, \qquad (2.3)$$

where β is the slope of the sigmoid, that is, the degree to which neurons close to threshold (i.e., 0) contribute to the activation dynamics. Lower slope values permit graded activation near threshold to influence performance, while higher slope values ensure that only above-threshold activation contributes to the activation dynamics. At extreme slope values, the sigmoid function approaches a step function. For all simulations presented here, $\beta = 0.5$.

Inputs to the model take the form of a Gaussian:

$$S(x,t) = c\exp\left[-\frac{(x - x_{center})^2}{2\sigma^2}\right]\chi^{(t)}, \tag{2.4}$$

centered at x_{center}, with width, σ, and strength, c. These inputs can be turned on and off through time (e.g., as items appear and then disappear). This time interval is specified by the pulse function $\chi^{(t)}$.

Lastly, activation within the field is influenced by the addition of a stochastic component consisting of spatially correlated noise:

$$noise = q\int dx'\, g_{noise}(x - x')\xi(x',t). \tag{2.5}$$

Noise was added to the simulations by convolving a noise field composed of independent noise sources with a Gaussian kernel specified by:

$$g_{noise}(x - x') = \frac{1}{\sqrt{2\sigma_{noise}}}\exp\left[-\frac{(x - x')^2}{2\sigma_{noise}{}^2}\right], \tag{2.6}$$

where σ_{noise} is the spatial spread of the noise kernel (set to 10 in all simulations). (For discussion of the differences between spatially correlated noise and Gaussian white noise, see Schutte, Spencer, & Schöner, 2003).

The second layer of the model, Inhib(v), is specified by the following equation:

$$\begin{aligned}\tau\dot{v}(x,t) = -v(x,t) + h_v + \int c_{vu}(x - x')\Lambda_{vu}(u(x',t))\,dx' \\ + \int c_{vw}(x - x')\Lambda_{vw}(w(x',t))\,dx' + noise \quad .\end{aligned} \tag{2.7}$$

As before, $\dot{v}\ (x,t)$ specifies the rate of change of activation across the population of feature-selective neurons, *x*, as a function of time, *t*; the constant τ sets the time scale (note that the time scale for inhibition is faster than for the excitatory layers, i.e., $\tau_v < \tau_u$); *v(x, t)* captures the current activation of the field; and h_v sets the resting level of neurons in the field. Inhib(v) receives activation from two projections: one from PF(u), $\int c_{vu}\,(x - x')\Lambda_{vu}\,(u(x',t))\,dx'$; and one from WM(w), $\int c_{vw}\,(x - x')\Lambda_{vw}\,(w(x',t))\,dx'$.

As described above, these projections are defined by the convolution of a Gaussian kernel (Equation 2.2) with a sigmoidal threshold function (Equation 2.3). Finally, this field also receives spatially-correlated noise, as described above. This noise is independent from the noise sources in the other layers of the model.

The third layer of the model, WM(*w*), is governed by the following equation:

$$\begin{aligned}\tau\,\dot{w}(x,t) = -w(x,t) + h_w + \int c_{ww}\,(x - x')\Lambda_{ww}\,(w(x',t))\,dx' \\ - \int c_{wv}\,(x - x')\Lambda_{wv}\,(v(x',t))\,dx' + \int c_{wu}\,(x - x')\Lambda_{wu}\,(u(x',t))\,dx' \\ + c_s S(x,t) + noise \quad . \end{aligned} \tag{2.8}$$

Again, $\dot{w}\ (x,t)$ is the rate of change of activation across the population of spatially-tune neurons, *x*, as a function of time, *t*; the constant τ sets the time scale; *w(x,t)* captures the current activation of the field; and h_w sets the resting level. WM(*w*) receives self excitation, $\int c_{ww}\,(x - x')\Lambda_{ww}\,(w(x',t))\,dx'$, lateral inhibition from Inhib(v), $\int c_{wv}\,(x - x')\Lambda_{wv}\,(v(x',t))\,dx'$, and excitatory input from PF(u), $\int c_{wu}\,(x - x')\Lambda_{wu}\,(u(x',t))\,dx'$.

This field also receives direct target inputs, *S(x,t),* scaled by c_s to be weaker (i.e., $c_s < 1$), and includes spatially-correlated noise as described above. Again, this noise is independent from the noise sources in the other layers of the model.

The nodes in the response layer are governed by the following equations:

$$\tau\,\dot{r}_c(t) = -r_c + h_r + c_{cc}\Lambda_c(r_c) - c_{nc}\Lambda_n(r_n) + \int c_{cu}\Lambda_u(u(x',t))dx' + noise \tag{2.9}$$

$$\tau \dot{r}_n(t) = -r_n + h_r + c_{nn}\Lambda_n(r_n) - c_{nc}\Lambda_c(r_c) + \int c_{nw}\Lambda_w(w(x',t))dx' + noise \tag{2.10}$$

The rate of change of each node's activation, $\dot{r}$ (where the constant τ determines the timescale and the subscripts *c* and *n* denote the *change* and *no change* nodes, respectively), is determined by the current activation level, -*r*, and the resting level of the node activation, h_r. Each node has a self-excitatory connection, $c_{cc}\Lambda_c(r_c)$ or $c_{nn}\Lambda_n(r_n)$, and receives inhibition from the other node, $c_{cn}\Lambda_{cn}(r_n)$ or $c_{nc}\Lambda_{nc}(r_c)$. Additionally, the *change* node receives summed excitatory input from PF(u), $\int c_{cu}\Lambda_u(u(x',t))\,dx'$, and the *no change* node receives summed excitatory input from WM(w), $\int c_{nw}\Lambda_w(w(x',t))\,dx'$. Lastly, *noise* represents white noise added to the activation of the node at each time step. The noise sources are independent for each node.

Model simulations. The three-layer architecture described above is the first model to capture each of the processing steps in a change detection trial (as described in Chapter 1): encoding the items, maintaining these items through the short delay, comparing the items held in working memory to the test array, and then generating a decision. To illustrate how the model works, I describe a set of sample simulations below. Critically, the DFT is able to make both *change* and *no change* decisions, and responds correctly on the majority of trials with one to four items in the array (as adults do in this task; see, e.g., Vogel et al., 2001). In addition, the model also makes errors that match the performance of research participants as the number of items in the sample array increases and exceeds the model's capacity. Table 2 shows the parameter values for these simulations.

Figure 5 shows the DFT performing a set size 3 (SS3) "correct rejection" trial—responding *no change* when no items in the display changed. This figure shows time-slices through all three layers at critical points in the trial: (B) when the sample array is presented; (C) when the memory items are being maintained during the delay; and (D) when the test array is presented and the decision is generated. To show the decision process, activation of the decision nodes (Figure 5A) is shown across time in the trial. The trial begins with the sample array presented for 500 ms; these inputs are projected

strongly into PF and weakly into WM. During the stimulus presentation (Figure 5B), self-sustaining peaks form in WM, which lead to bumps of inhibitory activation in Inhib. These project inhibition back to PF and suppress the input-driven peaks there. During the 900 ms delay (Figure 5C), the peaks in WM sustain and the associated patterns of inhibitory activation project to PF, leading to "troughs" of inhibition at these stimulus values in PF. When the test array is presented with the same three colors as in the sample array, these values are inhibited in PF as they are maintained in WM (see Figure 5D). Critically, the input is not strong enough to generate above-threshold (i.e., above 0) peaks in PF. As a consequence, when the decision nodes are engaged by boosting the resting level of these neurons after the test array is presented (see arrow in Figure 5A), the three peaks being held in WM send strong activation to the *no change* node, whereas the inhibitory troughs in PF prevent this layer from sending activation to the *change* node. The *no change* node easily "wins" the competition between nodes and generates a *no change* response.

Figure 6 shows the DFT performing a SS3 "hit" trial—responding *change* when one item has changed. This trial begins as the previous trial did, with input from the three colors projecting to PF and WM, and quickly forming self-sustaining peaks in WM (Figure 6B). These peaks are maintained through the delay (Figure 6C), and the shared Inhib layer produces troughs of inhibition at the remembered values in PF. When the test array appears this time (Figure 6D), though, one of the colors has changed. This input projects to a relatively uninhibited region of PF, that is, at the new color value, and an input-driven peak forms in PF (see red circle in Figure 6D). As a consequence, when the decision nodes are boosted (Figure 6A), PF sends strong activation to the *change* node. Although activation is also sent from WM to the *no change* node, the activation to the *change* node is stronger (through a stronger projection from PF to the decision nodes than from WM; see Table 2). The *change* node is able to win the competition, rising above threshold and generating a *change* decision.

These two simulations illustrate how the comparison and decision mechanisms work for trials within the model's capacity, that is, when the model is able to stably form a WM peak associated with each color value. These two response types, correct rejections and hits, are the most frequent in adults' change detection performance. As the number of items increases and exceeds participants' capacity, however, performance declines. In particular, the proportion of hits begins to decrease as participants make "miss" errors—*no change* responses on Change trials. These are the most common errors, which occur about three times more frequently than "false alarm" errors—*change* responses on No Change trials (e.g., Vogel, Woodman, & Luck, 2001). This primary decrease in hits along with the small increase in false alarms figure most prominently into the calculation of capacity using Pashler's (1988) formula (see discussion above).

This raises a central question for the DFT: can the mechanism proposed for change detection performance capture this capacity limitation? This is not a trivial issue—it is not enough for performance to simply decline as the set size increases, but the pattern across the four response types must be approximated to produce similar capacity estimates. Recently, we tested the capacity limits of the model described above (Simmering, Johnson, & Spencer, 2008a). Interestingly, these simulations revealed that capacity limits are not strictly defined by the number of items being held in WM in the DFT, which could be expected from typical explanations of change detection performance. Rather, three constraints in the model contribute to capacity estimates: the number of peaks held in WM, the details of the task, and the decision process.

First, as might be expected, the number of peaks in WM is limited; this limit arises from the balance between excitation and inhibition in this layer. When inputs are projected into the model, self-sustaining peaks begin to form in WM. As excitation builds in WM, activation is projected to Inhib, leading to a bump of inhibition associated with each peak. Because the projection from Inhib back to WM is quite broad (a constraint we derive from neural data showing broad surround inhibition in ventral stream cortical areas

such as inferior temporal cortex, Desimone & Duncan, 1995), multiple bumps in Inhib will project broad global inhibition back to WM. This, in turn, slows the building of peaks in WM. At some point, the number of peaks in WM is high enough that the balance tips in favor of inhibition, and one or more WM peaks are either squashed or fail to build in the presence of input.

Figure 7 illustrates this point with a simulation of the DFT performing a SS4 false alarm trial. This trial begins with four inputs projecting to PF and WM. As in the SS3 trials, the model forms peaks in WM at these color values (see Figure 7B). During the delay (Figure 7C), however, these peaks are not all maintained: the peak at 80° dies out (see arrow). As a result, there is no corresponding trough of inhibition in PF at this color value (see red circle). When the same four colors are presented in the test array, therefore, an input-driven peak forms at 80° in PF (in fact, activation in PF did not fully return to baseline during the delay because there was no active inhibition from a peak in WM to squash it). When activation projects from PF and WM to the decision nodes, the input-driven peak in PF is enough to boost the *change* node above threshold, producing a *change* response. Essentially, the model has identified this color as new because it was never "consolidated" in WM.

This simulation also demonstrates that the details of the task will influence performance. Specifically, because the four colors that happened to be presented on this trial were near each other in color space, inhibition associated with these peaks grows quite strong and feeds back into WM. As a result, it is more difficult to maintain all of the peaks throughout the delay, and one of the peaks is squashed. In this example, then, capacity is determined in part by the details of the stimuli.

In conjunction with the limit on the number of peaks in WM, the decision process also influences performance. Figure 8 illustrates this with a simulation of the DFT performing a SS4 miss trial. The trial begins with an array of four items, which project inputs at four color values into PF and WM. Again, the model forms peaks in WM at

these color values (see Figure 8B), and these peaks sustain throughout the delay (Figure 8C). At test, one of these colors changes (i.e., yellow/40° changes to cyan/120°); as such, this new color value projects into a relatively uninhibited region of PF and WM. Instead of generating a correct *change* response, however, this input quickly builds a peak in WM, which in turn inhibits the activation in PF. Although the activation is slightly above-threshold in PF, this activation is not strong enough to generate a stable decision. Rather, the five peaks now being held in WM send activation to the *no change* node, which generates the (erroneous) response on this trial.

Taken together, these simulations provide important insights into the nature of capacity in the change detection task. According to traditional IP accounts of this task, misses result from a failure to store all of the items in memory, while false alarms are often attributed to either "guessing" or a decision bias. In these DFT simulations, on the other hand, the exact opposite was true: the false alarm resulted from remembering only three of the four items, and the miss resulted from the decision process (i.e., the *change* signal being too weak relative to the *no change* signal). By analyzing performance in a real-time process model, the mechanisms behind performance can be directly examined. *In this task, the building of activation in PF at test—the neural event that captures detection of a change—supports the generation of a change response.* Therefore, in the model, "capacity" is not simply about the number of items being held in WM. Rather, performance depends on the amount of inhibition projected to PF, which is related to but not strictly defined by the number of peaks in WM.

This is shown more directly in Figure 9. Figure 9A shows quantitative simulations of the DFT to capture adults' performance from a set of 3000 simulations (50 "participants" completing 12 trials at each of 5 set sizes) using the same parameters as in Figures 5-8 (see Simmering et al., 2008a). Note that, in this figure, correct rejections and false alarms sum to 1.0 (all No Change trials) and hits and misses sum to 1.0 (all Change trials). Figure 9B shows the average number of peaks held in WM over the delays on

these trials. Note that even though the model can hold at least four and sometimes five peaks in WM at the higher set sizes, the performance shown in Figure 9A yielded a capacity estimate of 3.98 items using the standard formula (Pashler, 1998). *Thus, the number of peaks held in WM is not equal to capacity as it is traditionally estimated.* This highlights the importance of moving concepts like "capacity" beyond previously-accepted IP definitions, to the level of neurally-grounded process. Results of these simulations show that there is more to capacity estimates than appears at first blush.

Given that the DFT reveals some hidden complexity in the concept of capacity and how this construct is measured in change detection, it is important to demonstrate the utility of this theoretical framework. In particular, does this theory generate any novel predictions? Johnson, Spencer, Luck, and Schöner (in press-a) have generated and tested several novel predictions about the nature of the memory representations underlying change detection performance. For instance, they showed that as peaks of activation are held in memory over the short delay, metric interactions can subtly change the accuracy of memory in predictable ways. In a series of experiments, participants were presented with sample arrays of three colored squares, two of which were similar to one another (e.g., two shades of purple), and one of which was distinctive. Based on findings in verbal working memory (e.g., Conrad, 1964), one might expect that the two similar colors will interfere with one another. The DFT predicts, however, that people should be *better* at detecting changes in similar colors. This occurs because two close peaks in the model will share laterally inhibitory interactions which effectively "sharpen" the representation of the two similar items. These two sharper peaks can then support a more precise detection of changes in the display. This is precisely what adults showed: change detection performance was enhanced when one of the close items was probed relative to when a distinctive color was probed. Thus, the DFT is able to capture established patterns of performance in change detection, as well as generate novel predictions about performance.

Preferential Looking in the Dynamic Field Theory

The previous section described a new model of visual working memory and change detection. In this section, I describe how this work has been recently adapted to capture the processes that operate in the infant preferential looking task.

Spencer and colleagues (Spencer & Perone, 2008; Spencer, Simmering, Perone, & Johnson, 2007a) proposed a five-layer dynamic neural field model that builds on the model of adult change detection described above. Two modifications were made to the three-layer architecture to develop the five-layer model described here. First, a type of Hebbian learning was added to the model through the addition of two long-term memory (LTM) fields, one coupled to each of the excitatory fields. As activation builds in PF or WM, traces of activation are laid down at associated sites in the LTM fields. Such traces feed back onto PF and WM, biasing the model to build peaks of activation at feature values that were activated earlier in the trial, leading to a form of Hebbian learning. Second, rather than a decision system to generate *change/no change* responses, the model was equipped with a simple fixation system to capture autonomous looking behavior. This system consists of three self-excitatory and mutually competitive nodes: one associated with the left display, one for the right display, and one for looks away from the testing apparatus. When one of the nodes "wins" this mutual competition by piercing threshold (i.e., above zero activation), the system "fixates" the corresponding location (i.e., the left display, right display, or neither). This system is described in more detail below.

Model equations. The equations for the three-layer architecture were given previously. Thus, the present section describes the model equations for the preferential looking model which add LTM fields and an autonomous looking system to the basic three-layer architecture.

The LTM fields, $LTM_{PF}(u_{ltm})$ and $LTM_{WM}(w_{ltm})$, are governed by the active states on the associated excitatory field (note that, for simplicity, only the $LTM_{PF}(u_{ltm})$ equation is shown):

$$\tau_{ltm}\,\dot{u}_{ltm}(x,t) = \dot{u}_{ltmbuild}(x,t) + \dot{u}_{ltmdecay}(x,t) \tag{2.11}$$

The rate of change of activation, $\dot{u}_{ltm}(x,t)$, for each neuron in the long-term memory layer across the spatial dimension, *x*, as a function of time, *t*, is specified by two components: the build-up of activation, $\dot{u}_{ltmbuild}$, and the decay of activation, $\dot{u}_{ltmdecay}$. These components were specified as follows:

$$\tau_{ltmbuild}\dot{u}_{ltmbuild}(x,t) = \left[u_{ltm}(x,t) + \Lambda(u(x,t)\right]\cdot\theta(u(x,t))\,; \tag{2.12}$$

$$\tau_{ltmdecay}\dot{u}_{ltmdecay}(x,t) = u_{ltm}(x,t)\cdot\left[1-\theta(u(x,t))\right]. \tag{2.13}$$

The shunting function θ ($\theta = 1$ if $u(x,t) > 0$ and $\theta = 0$ otherwise) determines where the activation is built up and maintained in u_{ltm}. The build-up rate is relatively fast ($\tau_{ltmbuild}$ = 10 for all simulations) while the decay rate is much slower ($\tau_{ltmdecay}$ = 100,000 for all simulations); this allows previous memory to be maintained over longer periods. Separating the build-up and decay mechanisms approximates accumulation and depression of synaptic change (Deco & Rolls, 2005).

Nodes in the fixation system are specified by the following equations:

$$\tau\,\dot{f}_l(t) = -f_l + h_f + c_{ll}\Lambda_l(f_l) - c_{lr}\Lambda_r(f_r) - c_{la}\Lambda_a(f_a) + c_{trans} + c_{const} + \int c_{lu}\Lambda_u(u(x',t))dx' \tag{2.14}$$

$$\tau\,\dot{f}_r(t) = -f_r + h_f + c_{rr}\Lambda_r(f_r) - c_{rl}\Lambda_l(f_l) - c_{ra}\Lambda_a(f_a) + c_{trans} + c_{const} + \int c_{ru}\Lambda_u(u(x',t))dx' \tag{2.15}$$

$$\tau\,\dot{f}_a(t) = -f_a + h_f + c_{aa}\Lambda_a(f_a) - c_{ar}\Lambda_r(f_r) - c_{al}\Lambda_l(f_l) + c_{trans} \tag{2.16}$$

The rate of change of each node's activation, $\dot{f}$ (where the constant τ denotes the time scale, and subscripts *l*, *r*, and *a* indicate the *left*, *right*, and *away* nodes, respectively) is determined by the current level of activation, *f*, and resting level of the node, *h*. The resting level of each node is dynamic and is specified by:

$$\tau_h\,\dot{h}_f\;(t) = -ah_f + h_{rest} + \Lambda_f(h_f)\cdot h_{down}, \tag{2.17}$$

with a slope (a). Note that the dynamic resting level of the node decreases toward the attractor specified by h_{down} when activation of the node (f) surpasses threshold (i.e., 0). When activation of the node drops below threshold, the resting level dynamically returns to the attractor specified by h_{rest}.

The nodes also receive non-specific input to draw fixation to a stimulus. In particular, to indicate the onset of a stimulus, a transient attentional capture (c_{trans}) is sent to the *left* and *right* nodes. Moreover, all nodes received constant input, c_{const}, corresponding to the salience of the stimulus being presented. For the *left* and *right* nodes, this reflected that something has been presented on one of the displays; for the away node, the input was much weaker to reflect the possibility that visual features other than the displays might capture attention and draw fixation away from the experimental apparatus. At each time step, independent white noise was applied to the input c_{const}.

The fixation nodes are coupled to PF(u) such that fixation to one display (e.g., activation of the left node) gates input of the feature values at the fixated location into PF(u); in turn, above-threshold activation in PF(u) supports activation of the currently active looking node. Thus, the equation for PF(u) becomes:

$$\tau \dot{u}(x,t) = -u(x,t) + h_u + \int c_{uu}(x - x')\Lambda_{uu}(u(x',t))\,dx' \qquad (2.15)$$
$$- \int c_{uv}(x - x')\Lambda_{uv}(v(x',t))\,dx' + c_{ul} f_l(x,t) + c_{ur} f_r(x,t) + c_{ltm} u_{ltm} + noise \quad ,$$

such that the terms c_{ul} and c_{ur} are zero when the node is not active, and one when the node is active. In addition, all simulations presented here used the same strength of the projection from the LTM fields, that is, $c_{ultm} = c_{wltm} = 0.1$.

Model simulation. Figure 10 shows a sample simulation in which the "infant" version of this model performs a single SS3 preferential looking trial. The model first fixates the No Change display (blue line in Figure 10A), which gates input from the three colors presented in that display and projects input strongly to PF and weakly to WM (as in the model of change detection described previously). This input produces three self-

stabilized peaks in PF initially (see circle Figure 10B), which help to build three self-sustaining peaks in WM (see circle in Figure 10C; note that in this "overhead" view of the fields, fuchsia "hot spots" indicate peaks). As with the change detection model, the peaks sustaining in WM projects inhibition back into PF via Inhib (not shown); this destabilizes the self-stabilized peaks in PF and drops activation back below threshold. Because activation dropped below threshold, PF is no longer supporting fixation, and the node drops back below threshold.

Next, the model switches to look at the Change display when activation of this node rises above threshold (see "switch" arrow in Figure 10A). During this fixation, the model builds peaks associated with the three colors; as these peaks begin to suppress the self-stabilized peaks in PF, the input "blinks" off. When the input comes back on, one of the colors has changed (see dashed circle in Figure 10C). This leads to a new self-stabilizing peak in PF, which continues to support activation of the fixation node to continue looking at the Change display. Before the next "blink" to change to another new color, however, support for this node has dropped enough that noise can cause a switch (see "switch" arrow in Figure 10A).

Now the model fixates the No Change side again. At this point, I have added a time-slice through the fields to show how activation is evolving (see vertical arrow across fields; Figure 10D-E). As Figure 10E shows, there are two peaks sustaining in WM (at 40° and -80°); these correspond to two of the three colors in the No Change display. One of these peaks (40°) has been maintained since the last fixation; as a result, it is already inhibiting the associated region of PF and preventing a self-stabilizing peak (see arrow in Figure 10D-E). For the second color, activation at one value (-80°) was near threshold at the beginning of this fixation, and therefore quickly builds a peak (although it is not yet strong enough to suppress the self-stabilizing peak in PF). The third color (-120°) is slower to build; as a result, activation in PF is still strong.

At this point in the trial, the model is beginning to show a change preference: fixation of the Change display continued through multiple stimulus presentations, while fixation of the No Change display was quickly released. How does this behavior relate to the underlying memory representation? The model tends to look away from items that are familiar—but in this architecture and this task, familiarity has two timescales. First are the items that are being actively maintained, that is, the self-sustaining peaks in WM. These also form the basis for change detection performance, as described above. But recall that this architecture has the addition of LTM fields. Each time a peak builds, it lays down a trace in the associated LTM field (e.g., WM peaks leave traces in LTM_{WM}). At each subsequent time step, this activation in LTM_{WM} feeds back into WM. When input is then presented at these familiar color values, peaks will build more easily and quickly, and will also be more resistant to decay or noise. Thus, even if the model is not actively holding an item in WM (i.e., sustaining a peak) at the time it is presented, it will tend to "look away" (i.e., release fixation) more quickly than if the item had no trace in LTM.

This point is illustrated with the next time slice (second vertical line across fields; Figure 10F-G). This slice was taken at the beginning of the model's next fixation of the No Change display. As can be seen in Figure 10G, activation is above threshold for two of the colors on this display (-80° and 120°; note that two colors from the previous display, -40° and 160°, are also above threshold) and about to pierce threshold for the third color (40°). These peaks then suppress activation in PF and inhibit support of the fixation node. Over the course of many fixations, this quick building and stronger maintenance of familiar colors will produce a preference to look at the Change display.

The model's performance in this simulation highlights how this task taps VWM: if peaks that correspond to the colors in the No Change display build quickly in WM (with support from LTM_{WM}), this inhibits PF and removes support of the fixated node. Essentially, the model prefers not to look at items that are held in memory; rather, it looks to items that are new or, in this case, changing. The model shows a change preference

when it is able to successfully build WM peaks for items on the No Change display; only then is the fixation system freed to look at the other display. This is where capacity limits emerge in the infant model. If, for example, neural interactions in WM allow only two peaks on the No Change side to build simultaneously, then one of the colors in PF would not be suppressed because there was no corresponding WM peak. Consequently, the model would continue to fixate the No Change side and fail to show a robust change preference score. *Change preferences in the preferential looking task, therefore, arise from detection of sameness, that is, successful formation of a WM for items on the No Change display.* Thus, although performance in this task depends on the same general processes from the adult change detection task—encoding, maintenance, and comparison—the different structure of the task—inputs blinking on and off rapidly across two displays throughout the trial, with no explicit decision required—taps the same system in a very different way.

The simulation in Figure 10 shows that the DFT can capture the processes that underlie performance in the preferential looking paradigm. But can this model capture the details of infants' performance including developmental changes in capacity between 6 and 10 months? Figure 11 shows that the answer is 'yes'. Spencer and colleagues (2007a; Spencer & Perone, 2008) were able to capture the general pattern of data reported by Ross-Sheehy et al. (2003, Experiment 1). In particular, the model provided good fits of the data for 6- and 10-month-old infants at set sizes one and three. To capture the developmental change in capacity between 6 and 10 months, only one change was necessary in the tuning of the model: the "older" model had slightly stronger inhibitory projections back to the excitatory layers (see also Simmering et al., 2008b). This helped build more stable peaks of activation in both layers, enabling the model to support multiple peaks simultaneously.

The DFT account of preferential looking moves beyond the standard IP approach in two important ways. First, it provides a specific developmental mechanism—change in

the strength of neural interactions—that underlies the increase in capacity between 6 and 10 months. Second, it provides an account for how the only behavioral measure in this task—looking behavior—relates to the underlying memory system. Importantly, this link has led to novel predictions about infants' performance in preferential looking tasks. In particular, Perone and Spencer (2008; see also Perone et al., 2007) have demonstrated that the frequency of switches between the two displays is related to the formation of working memories for the objects at each fixated location. WM peaks that build stronger and faster over repeated presentations lead to stronger inhibition projected back to PF. This, in turn, results in faster release of fixation and an increase in the tendency to look back and forth between the two displays. Thus, more frequent switching between displays corresponds to more robust memory peaks in the model. This has been verified experimentally, showing that changes in the rate of switching over familiarization trials predicts whether infants will show familiarity or novelty preferences at test (Perone & Spencer, 2008).

Toward a Unified Theory of Visual Working Memory and Changes in VWM Capacity over Development

In the previous sections, I reviewed how the DFT has been applied to adults' change detection performance and infants' preferential looking performance. Both of these models have generated novel predictions in their respective tasks (for change detection, see Johnson, 2008; Johnson et al., in press-a; for preferential looking, see Perone & Spencer, 2008). These models provide an account of the end points of the developmental trajectory probed in the VWM literature and provide the foundation for my work addressing the question raised in Chapter 1: what is the source of the discrepancy between capacity in preferential looking at 10 months and capacity in change detection at 5 years? To this end, the current section describes a unified model I

developed that combines the mechanisms described above to perform both tasks within the same model.

To examine whether a unified model can perform both tasks, I combined the adult change detection model from Simmering, Johnson, & Spencer (2008a), using the parameters listed in Table 2, with the LTM fields and fixation system from Spencer and Perone (2008). I first examined whether the addition of the LTM fields disrupted the change detection performance reported in Simmering et al. In particular, I ran a set of change detection simulations using the five-layer model, which produced comparable results (see Figure 8 above). Next, I examined how the model performed in the preferential looking task. In Ross-Sheehy et al.'s (2003) Experiment 3, they tested infants in two trials each at set sizes 2, 4, and 6. Using these task details, I ran 50 simulations of the model. The results of these simulations, revealed significant change-preference scores at all three set sizes (0.64 in SS2, $t = 8.12$, $p < .001$; 0.61 in SS4, $t = 5.47$, $p < .05$; 0.58 in SS6, $t = 1.66$, $p = .05$). Thus, the same model which showed a capacity of 3.95 items in change detection shows a capacity estimate of at least 6 items when put into the preferential looking task. This leads to two behavioral predictions:

(P1) adults should produce higher capacity estimates than infants in the preferential looking task;

(P2) these capacity estimates from preferential looking should be higher than estimates from change detection *when the same individuals are tested in both tasks.*

These predictions are note-worthy for two reasons. First, adults have never been tested in the preferential looking task. Second, I am predicting that adults will show a capacity of at least 6 items in this task, which is higher than the well-replicated 3-4 items generally accepted in the VWM literature (e.g., Vogel et al., 2001).

Although these simulations provide a strong foundation for these two predictions, they also raise an important question: does the preferential looking task provide any

index of the capacity estimated in change detection? Put differently, is performance directly related across these two tasks? The results described above suggest one possibility: change preference scores should drop as set size surpasses the capacity as estimated in change detection (i.e., capacity was estimated at 3.95 items in change detection, and preference scores for SS4 and 6 were lower than for SS2). A second possibility is also suggested by previous work by Perone and Spencer (2008) using the preferential looking model described above. Their simulations suggest that switch rate—how often the model looks back and forth between displays—provides an indication of the robustness of the underlying working memory representation. In particular, stronger peaks in WM (i.e., a more robust memory representation) produce stronger inhibition that feeds back to PF; as a result, the self-stabilized peaks in PF are suppressed more quickly, which causes the model to look away more quickly. Therefore, strong memory representations lead to fast rates of switching.

This prediction of the model has been verified experimentally (Perone & Spencer, 2008): in a paired-comparison version of the preferential looking task, infants with faster switching early in familiarization showed novelty preferences at test (indicating a robust memory for the familiarization stimulus), whereas infants with slower rates of switching show familiarity preferences at test. To generalize these findings to the version of the preferential looking task used here, switching should be higher for set sizes within capacity (as estimated in change detection) and lower for set sizes beyond capacity. In my simulations of the adult model in the preferential looking task, this would correspond to higher switching in SS2 and SS4, and lower in SS6. This relationship between capacity in change detection and performance preferential looking leads to a third prediction:

(P3) although capacity estimates should differ across tasks when testing the same individuals in both tasks (see P2), capacity in change detection should be related to change preference scores and switching across set sizes.

These first three predictions test key components of the DFT account of performance in these two tasks. However, they primarily concern adults' performance rather than why VWM capacity estimates differ between 10 months and 5 years. Can this model also provide insight into the development of VWM capacity? The second prediction above hints at one component of this discrepancy: when testing the same adults in both tasks, capacity estimates should be higher in preferential looking than in change detection. Generalizing this prediction over development, then, one might expect that children will also show higher capacity estimates in preferential looking than in change detection. Moreover, children should also show the predicted relationship across tasks if both tasks do, in fact, tap the same VWM system.

The question of developmental change points to the final question I am asking with this project: what is the developmental mechanism behind developmental changes in VWM capacity? As described in Chapter 1, previous IP accounts have failed to propose a mechanism for the increases in capacity between 6 and 10 months in preferential looking or between 5 and 10 years in change detection. More generally, developmental changes in capacity in other domains have pointed to two sources of change: increases in strategy use and neural myelination. Neither of these sources, however, has been linked to specific processes that capture how developmental changes in capacity occur.

Within the framework of the DFT, the Spatial Precision Hypothesis (SPH) provides an established account of developmental change from infancy through middle childhood and into adulthood in spatial cognition (see Chapters 4 and 5 for further discussion, see also Spencer et al., 2007b). According to the SPH, development occurs through the strengthening of neural interactions within cortical fields, specifically within the locally excitatory and laterally inhibitory connections among neurons. With weaker interactions early in development, the relative stability of the self-stabilizing and self-sustaining attractor states changes. In particular, weak interactions make it harder for a neural field to move into a self-sustaining state; practically, this means that peaks may

require stronger input or more time to build. In addition, less stable peaks can be de-stabilized by relatively small perturbations (e.g., noise or interference)—that is, peaks will "die out" more easily. This hints at how the relative stability of self-sustaining peaks can have important consequences for behavioral development: by this view, children should have a harder time actively maintaining information in working memory, a claim which is consist with empirical observations in early development (see, e.g., Dempster, 1981).

Could the Spatial Precision Hypothesis produce developmental changes in VWM capacity? A sample simulation of a "child" change detection trial is shown in Figure 12; for this simulation, I used the parameters from Table 2 and simply scaled the local excitation (*c_uu* and *c_ww*) and lateral inhibition (*c_uv* and *c_uw*) by 0.5. As this figure shows, even though four items were presented, only three peaks are formed and sustained in WM (see arrow in Figure 12C); thus, when the same four items are presented at test, the model makes a "false alarm" error, indicating that an item changed (i.e., detecting a peak in PF) even though it did not. Note that, although this peak in PF was quite weak, the peaks in WM were also weak, therefore reducing the input to the *no change* node and allowing the *change* node to dominate. This sample simulation provides initial support for the fourth prediction I tested, in this case, using simulations of the unified model (see Chapter 4):

(P4) implementing the SPH in the unified model will capture developmental changes in VWM capacity, thereby providing support for the hypothesis that changes in the strength of neural interactions over development underlie changes in capacity.

To summarize, I will test four predictions with experiments and simulations in subsequent chapters. First, a series of experiments will test children (ages 3, 4, and 5 years) and adults in both change detection and preferential looking. These experiments will test the first three predictions:

(1) adults, and possibly children, will show a higher capacity than infants when tested in preferential looking—at least 6 items, which is also higher than the commonly accepted adult-like capacity of 3-4 items;

(2) estimates of capacity measured *from the same individuals* in the change detection task will be substantially lower than 6 items; and

(3) because these two tasks depend on the same underlying VWM system, there should be a relationship between capacity measured in one task and capacity measured in the other, in particular, change preference scores and switch rate in preferential looking should relate to capacity in change detection.

Finally, in Chapter 4, I will use these empirical data to test my fourth prediction, that:

(4) implementing the Spatial Precision Hypothesis in the unified model will capture developmental changes in capacity in these tasks.

In particular, I will present quantitative simulations capturing data from both tasks across all four points in development probed experimentally.

If these four predictions are supported, the current project will provide strong support for the first formal model to address the processes that underlie performance in both change detection and preferential looking. Moreover, this project will provide strong tests of the first formal model to capture developmental changes in either task using a specific mechanism that can be directly linked to developmental changes in capacity.

Figure 3. Two-layer neural field model of the type proposed by Amari (1977). (A) The model consists of a single layer of feature-selective excitatory neurons reciprocally coupled to a similarly-tuned layer of inhibitory interneurons; excitatory and inhibitory projections between layers are shown with green and red arrows (respectively). Neurons 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. Neurons in the inhibitory layer transmit broad lateral inhibition back to the excitatory field (dashed arrows). This architecture can exhibit self-stabilizing (B) and self-sustaining (C) attractor states. See text for additional details.

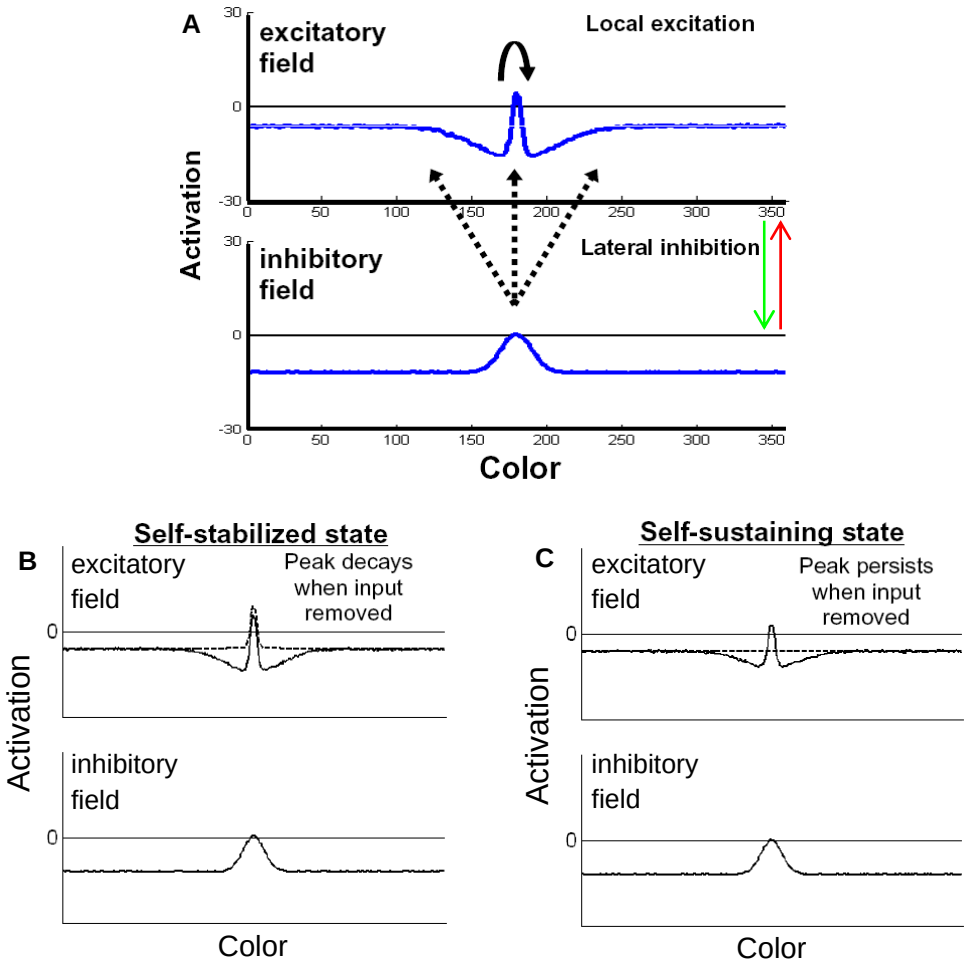


Figure 4. The Dynamic Field Theory (DFT) consisting of three layers: (A) a perceptual field (PF), (B) an inhibitory layer (Inhib) and (C) a working memory field (WM). In each panel, time is across the x-axis, activation on the y-axis, and color value on the z-axis. Arrows indicate excitatory (green) and inhibitory (red) projections between layers. These fields are coupled to two self-excitatory, mutually-inhibitory nodes: one for *change* decisions receives activation projected from PF; one for *no change* decisions receives activation projected from WM.

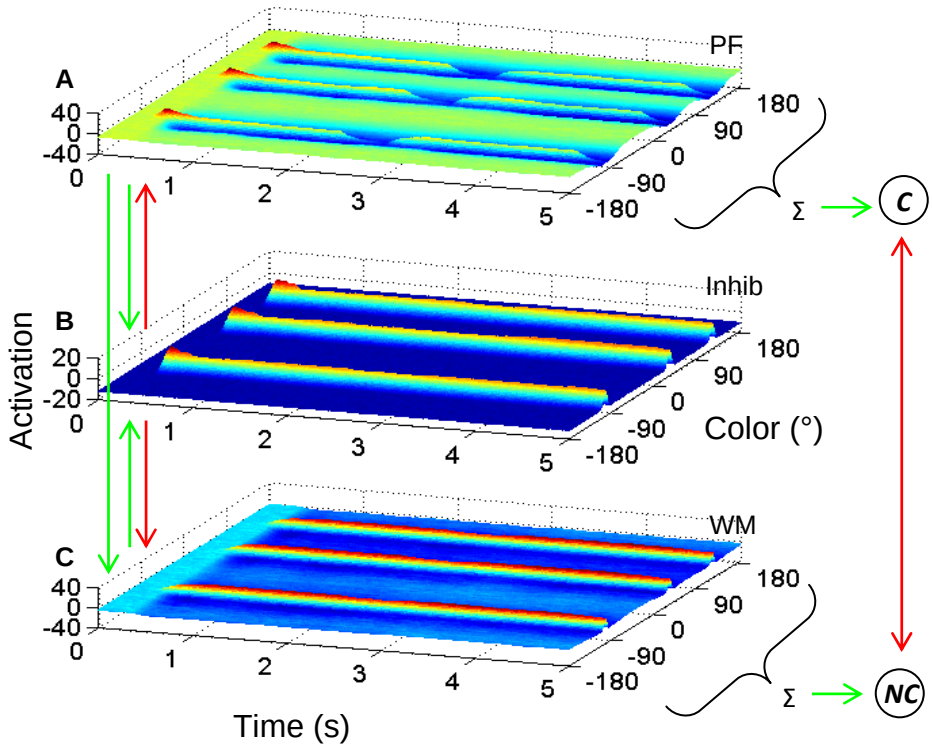


Figure 5. A simulation of the DFT performing a SS3 "correct rejection" trial in the change detection task. This trial consisted of three colors presented for 500 ms, followed by a 900 ms delay, then presentation of the original three colors. (A) Activation (*y*-axis) of the *change* (blue) and *no change* (red) decision nodes over time (*x*-axis). (B-D) Time-slices through PF, Inhib, and WM show activation (*y*-axis) across color values (*x*-axis) at three critical points during the trial.

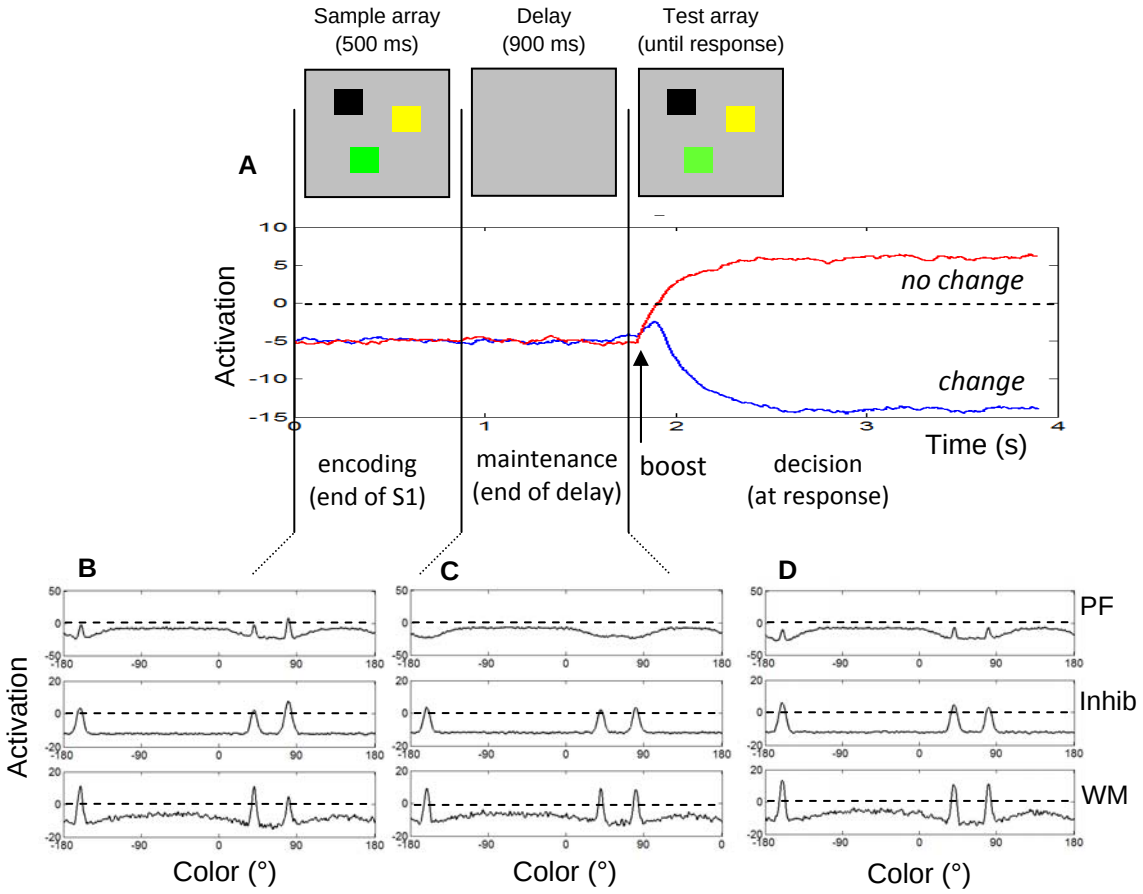


Figure 6. A simulation of the DFT performing a SS3 "hit" trial in the change detection task. Panels and axes are as in Figure 5.

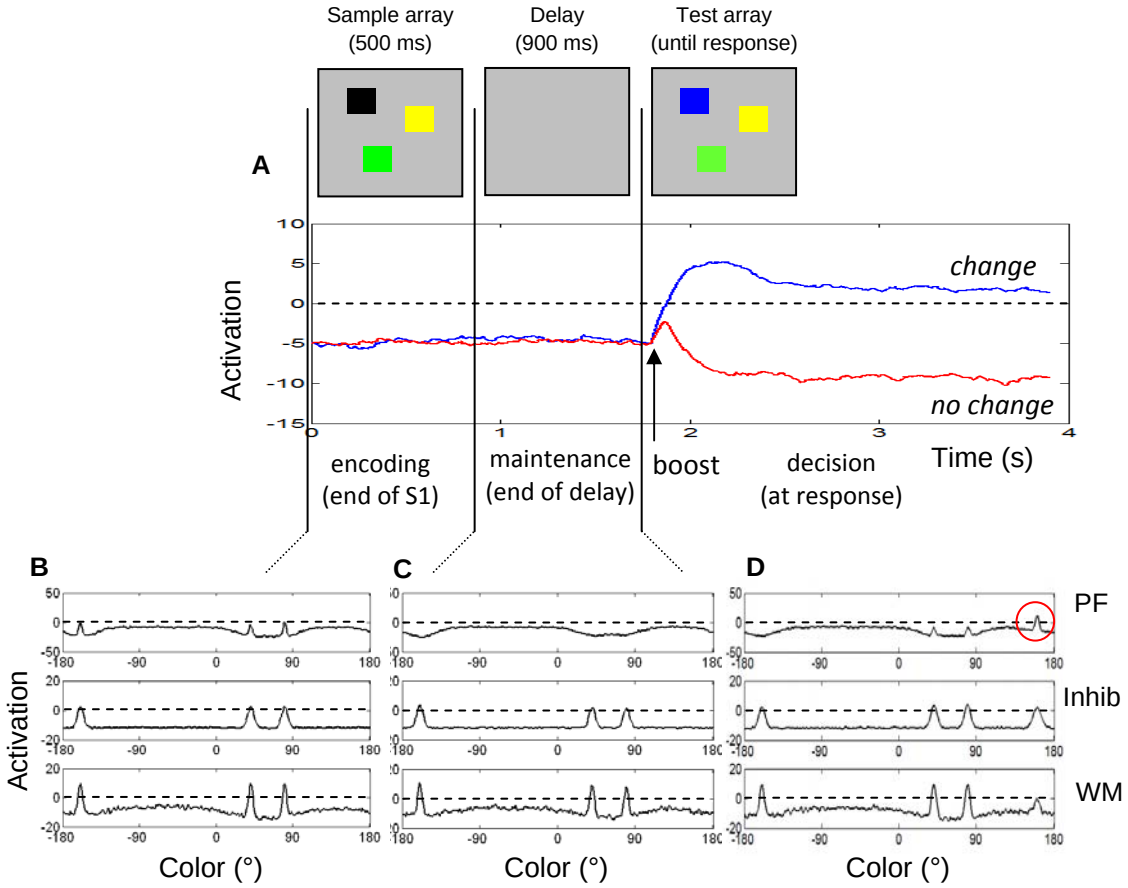


Figure 7. A simulation of the DFT performing a SS4 "false alarm" trial in the change detection task. Panels and axes are as in Figure 5.

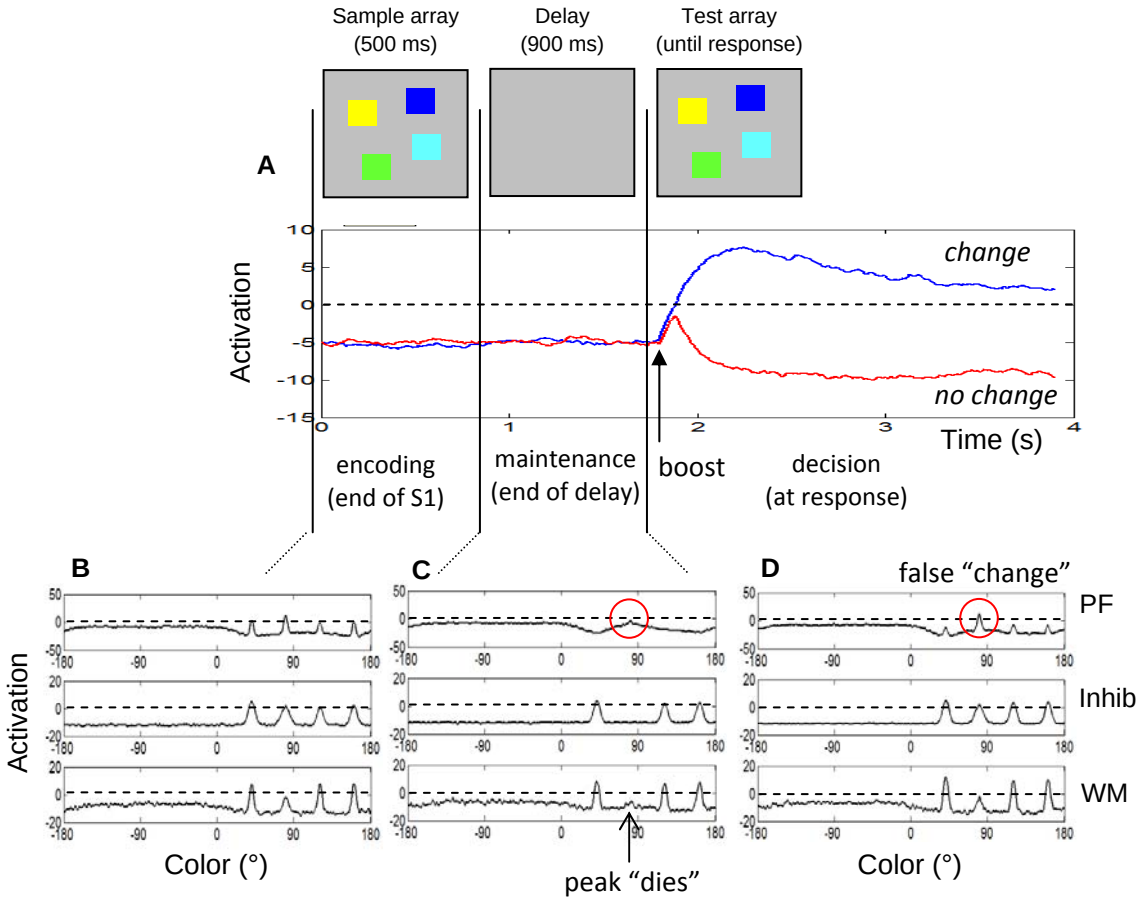


Figure 8. A simulation of the DFT performing a SS4 "miss" trial in the change detection task. Panels and axes are as in Figure 5.

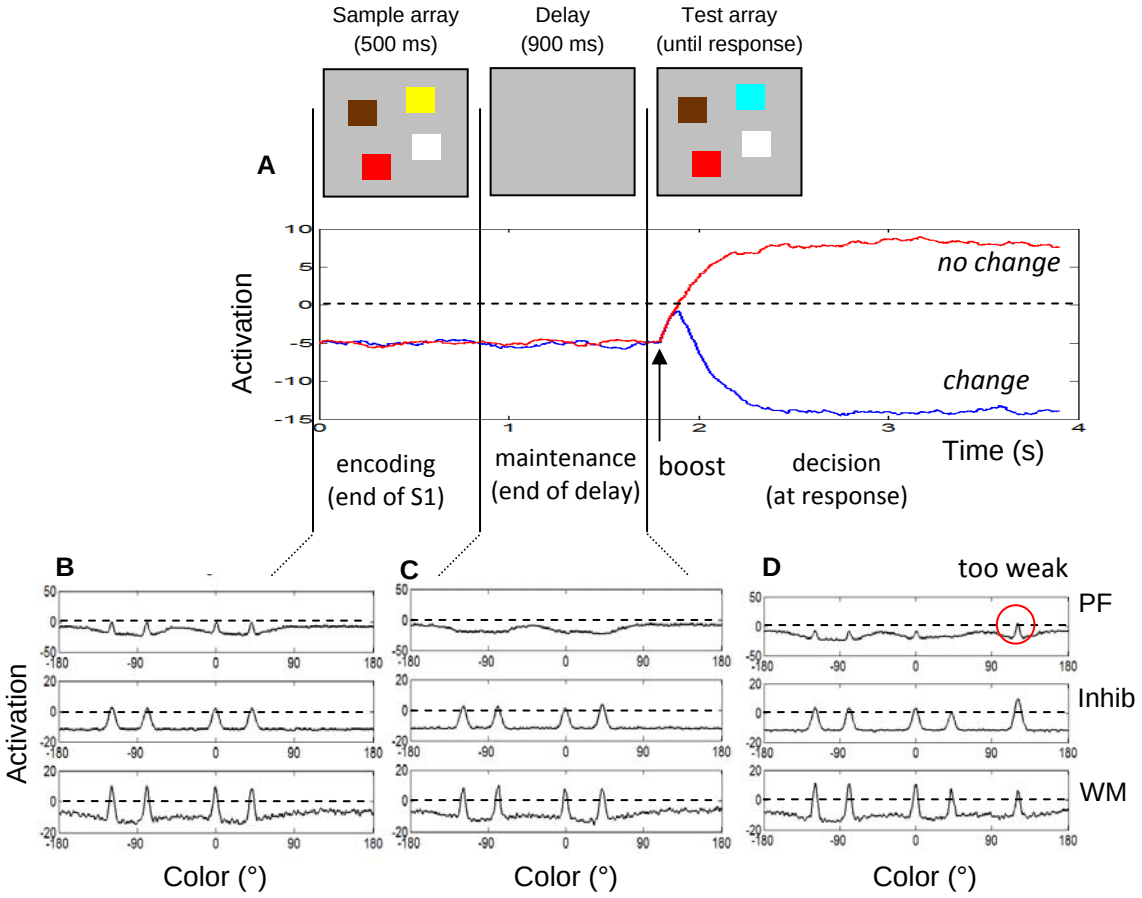


Figure 9. Simulation results for change detection set sizes one through five from Simmering et al. (2008). (A) Portion of trials with each response type. Note that proportions are shown relative to the number of Change and No Change trials; therefore correct rejections (CR) and false alarms (FA), as well as hits (H) and misses (M), sum to 1.0. (B) Mean number of peaks in WM at the end of the delay. Error bars show standard error.

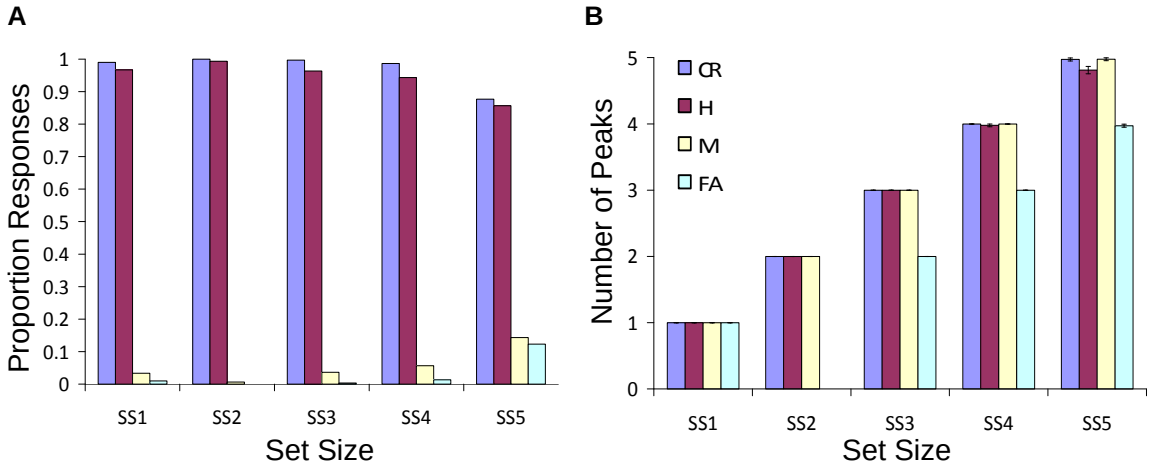


Figure 10. A simulation of the five-layer DFT performing a SS3 preferential looking trial. (A) Activation (*y*-axis) of the fixation nodes over time (*x*-axis). Two of the five layers, PF (B) and WM (C), are shown with time along the *x*-axis and color along the *y*-axis. (D-G) Time slices though these layers are also shown, with color along the *x*-axis and activation along the *y*-axis.

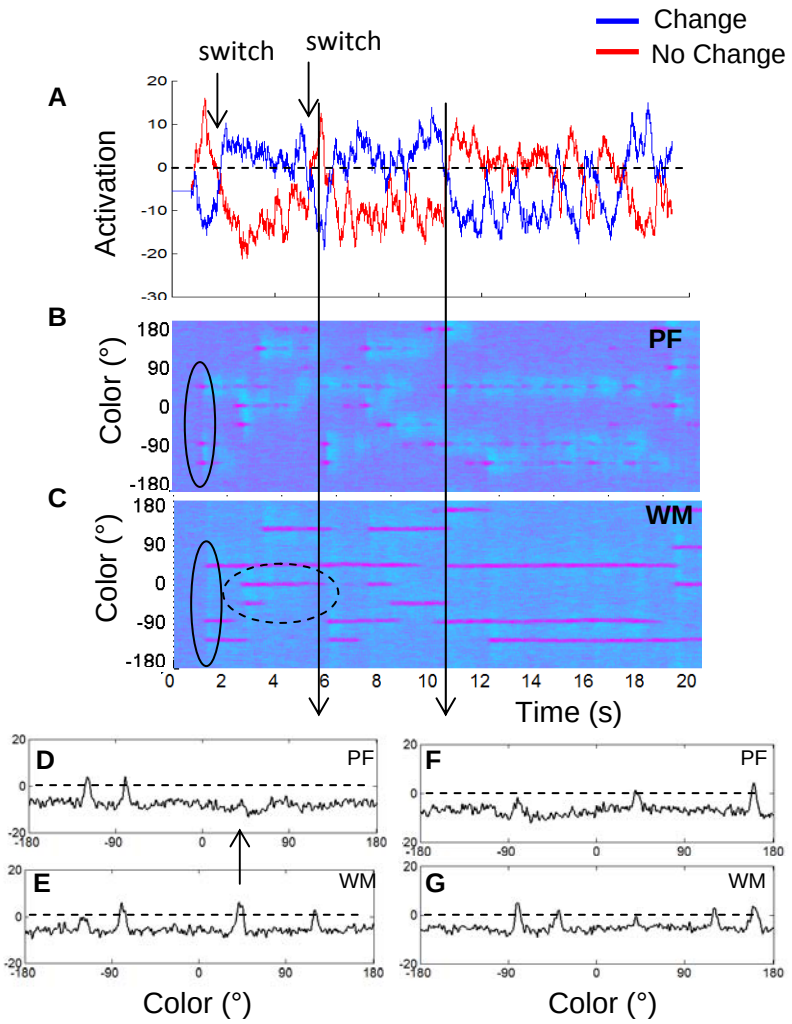


Figure 11. Approximate infant data from Ross-Sheehy et al. (2003) and DFT simulations from Spencer et al. (2007).

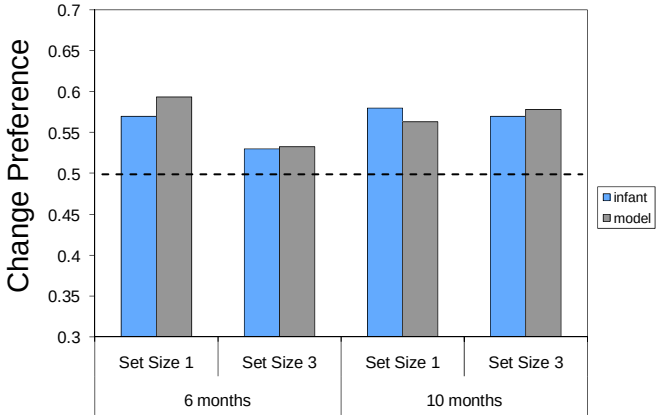


Figure 12. A simulation of the DFT performing a change detection trial with parameters scaled to capture early development. Panels and axes are as in Figure 5.

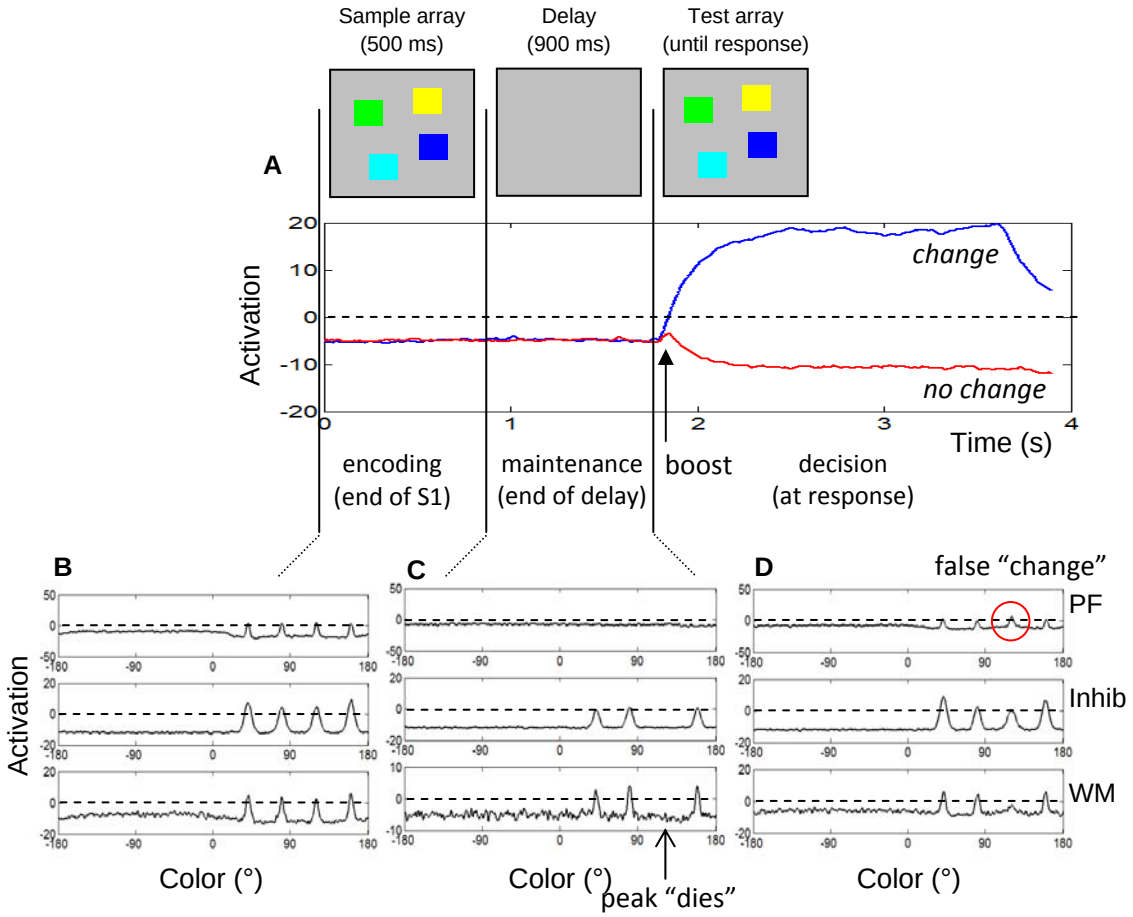


Table 2

Parameter Values from Change Detection Simulations

Layer/Node	τ	h	self-excitation	excitatory projection(s)	inhibitory projection(s)	input	noise
PF(u)	80	-7	$c_{uu} = 3.10$ $\sigma_{uu} = 3$		$c_{uv} = 1.85$ $\sigma_{uv} = 26$ $k_{uv} = 0.05$	$c_{tar} = 18$ $\sigma_{tar} = 3$	$q_u = 0.1$ $\sigma_q = 10$
Inhib(v)	10	-12		$c_{vu} = 2$ $\sigma_{vu} = 7$ $c_{vw} = 1.95$ $\sigma_{vw} = 5$			$q_v = 0.1$ $\sigma_q = 10$
WM(w)	80	-5	$c_{ww} = 3.15$ $\sigma_{ww} = 3$	$c_{wu} = 1.5$ $\sigma_{wu} = 5$	$c_{wv} = 0.6$ $\sigma_{wv} = 42$ $k_{wv} = 0.08$	$c_{tar} = 4$ $\sigma_{tar} = 3$	$q_w = 0.1$ $\sigma_q = 10$
Change(r_c)	80	-5	$c_{cc} = 1.85$	$c_{uc} = 1$	$c_{cn} = 14$	$c_{cu} = 1.005$	$q_c = 0.075$
No Change(r_n)	80	-5	$c_{nn} = 1.85$	$c_{wn} = 1$	$c_{nc} = 14$	$c_{nw} = 0.14$	$q_n = 0.075$

CHAPTER 3

EMPIRICAL TEST OF PREDICTIONS

The previous chapter described the processes in the DFT that underlie performance in the adult change detection task and the infant preferential looking task, with the goal of reconciling the discrepancy in capacity estimates in the literature between 10 months (3-4 items in preferential looking) and 5 years (1.5 items in change detection). By implementing both tasks in the same model, I was able to explain the processes at work in each task, showing that these tasks do, in fact, tap the same visual working memory system, but in different ways. The change detection task depends on the maintenance of peaks in WM and an explicit comparison at test, leading to a single decision on each trial. The preferential looking task, on the other hand, depends on the initial building of peaks for the No Change display over repeated presentations, which can support the reliable detection of sameness over the course of the many looking decisions that occur on each trial. The differences across tasks led to a set of specific predictions by the DFT: capacity estimates for older children and adults should be greater than four items in the preferential looking task; capacity estimates for the same participants should be substantially lower in change detection; and there should be correlations between specific measures of capacity from these tasks over development because they both tap the same VWM system.

The literature currently includes a large empirical gap between 13 months (the oldest age tested in preferential looking) and 5 years (the youngest age tested in change detection), with no data from either task in this intermediate period of development. To try and bridge this gap, the change detection task must be modified to allow children younger than 5 years to make it through the task. Thus, I conducted a pilot study to probe the youngest age that could reliably perform a more child-friendly version of the task.

To modify the change detection task for young children, I used the study by Riggs et al. (2006) as a starting point. In this study, the authors modified the standard task in two ways relative to the task used with adults. First, the total number of trials required of each participant was reduced to 12 (6 Change, 6 No Change) at each set size of one to five items (compared to, e.g., 40 trials per set size; Vogel et al., 2001). Second, the trials were blocked by set size and ordered such that the easier set sizes (one and two) were first, followed by the higher set sizes.

Pilot data with 3- to 4-year-old children revealed two specific problems young children experienced with this task that prompted additional modifications. First, it was difficult for young children to understand the required comparison between the sample and test array. Some children, for instance, would compare the items within a single array; other children compared the items in the test array on one trial to the sample array on the subsequent trial. To help children understand the appropriate comparison between arrays, I changed the description to that of a card matching game, and then presented items in each array on a rectangular background to identify them as a "card". To reduce interference across trials, the spatial location of the cards alternated between the left and right sides of the screen from trial to trial. The second problem for children was viewing and encoding all of the stimuli during the short presentation of the sample array. Children tended to look around the experimental room between trials, so extra display time was needed to ensure they oriented to the computer screen and viewed each display. This was addressed first by increasing the duration of the sample array from 500 ms to 2 s. In addition, rather than presenting the items in randomly-selected locations on each trial, they were clustered near the center of the card at fixed locations on each trial. This reduced the need for children to scan the screen to find all of the items during encoding.

With these modifications, children as young as 3.5 years of age can reliably perform the change detection task. This defines the youngest age that may be tested in this task. I then selected three age groups to participate: 3.5 years, 4.25 years, and 5.0

years. In addition, a group of adults was included for comparison. All participants completed both the change detection task (using the modifications described above) and the preferential looking task. Note that participants completed the preferential looking task first so as not to bias them to look for changes in the displays based on their experience in the change detection task where explicit verbal decisions are required on each trial.

In the remainder of this chapter, I describe participants' performance in the preferential looking task in Experiment 1. A subset of these participants also completed one of two conditions of change detection, color or shape. I included these two versions of the change detection task to test the generality of changes in VWM capacity, that is, whether developmental increases in capacity for colors are specific to color memory, or apply to visual memory for other features as well. The results of these change detection tasks are described in Experiments 2 and 3. Then, I describe comparisons across these two tasks to provide a critical test of the DFT framework. Even though these two tasks are predicted to generate different capacity estimates, the DFT asserts that performance in each task depends on the same underlying VWM system. Thus, although the results may diverge across experiments, there should be correlations in performance across the two tasks that highlight the unified system underlying performance.

Experiment 1: Preferential Looking

Method

Participants. A total of 155 children participated in this study: 56 three-year-olds (*M* age = 3 years, 4.46 months; *SD* = 2.52 months), 44 four-year-olds (*M* age = 4 years, 2.87; *SD* = 1.93 months), and 55 five-year-olds (*M* age = 5 years, 0.81 months; *SD* = 1.74 months). An additional 7 children participated but were not included in analyses due to incomplete data (1 five-year-old; see below for details), 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* age = 21 years, 5.35 months; *SD* = 3.53 years). An additional 2 adults participated but were not included in analyses because they wore glasses; the glare from the computer monitors made it impossible to code their looking. All participants 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, Oakes, & Luck, 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 (note that exact 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 2, a solid gray background was continuously present on both monitors, with arrays of colored squares blinking on and off together; one monitor contained the changing display and the other monitor contained the non-changing display. Each display contained two, four, or six colored squares that measured approximately 5 cm by 5 cm. The set size (i.e., the number of squares on each monitor) was identical for 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 (see Table 3 for RGB values). The colors within a display were always different from each other, but colors 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 constant throughout the 20 s trial. For the Change display, one color would change to a new color at each onset; 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 previous 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 was seated on a chair (or, for some of the younger children, in a parent's lap), facing the two computer monitors. If a parent was seated with the child in the testing room, s/he was given occluding glasses to wear 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 audible 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 was completed, 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. Additionally, about 70% of adults reported this difference across displays.

Each session included two trials at each of the three set sizes—one of each 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

randomly determined for each participant. The experimenter was seated behind the curtain to initiate each trial and record the duration of participants' looking to each of the two monitors using a program developed for the Macintosh (Habit: Cohen, Atkinson, & Chaput, 2000). For some sessions, a trained observer was not available to code looking during the session. These sessions were coded offline using the video recording. In all cases, the coder was unaware of which images were being presented on the monitors.

Each trial began with the attention-getter (the black box with a beeping red light), and the experimenter pressed a key once she judged that the participant was attending to the light. For the duration of the 20 s trial, she then coded the child'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 a 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 minutes to complete.

Results

Before analyzing the results, I checked the data for trials on which the participant fixated only one of the displays. I chose to exclude 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 16 trials were contributed by one child (a 5-year-old), leading to exclusion of all but one trial of this participant's data; as a result, this participant was excluded from analysis.

Total Looking Time. Total looking time was computed as the time (in s) that the participant spent looking 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. As

this figure shows, total looking time varied little across set size and age, although adults' looking time was higher than children's. I analyzed these data in a two-way ANOVA with Set Size (2, 4, 6) as a within-subjects factor and Age (3 years, 4 years, 5 years, adult) as a between-subjects factor. This analysis revealed only a significant main effect of Age ($F_{3, 191} = 17.74$, $p < .001$). Follow-up Tukey HSD tests ($p < .05$) showed that adults looked longer on average than 3-, 4-, or 5-year-olds looked; children's looking time did not differ across age groups. There were no differences in looking time across set sizes.

Change-Preference Scores. Change-preference scores were computed as the time (in s) the participant spent looking to the Change display divided by the total looking time to both displays (i.e., Change / [Change + No Change]) for each trial. Scores were then 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 SS2, with adults showing the highest scores. These scores were analyzed in an ANOVA with Set Size (2, 4, 6) as a within-subjects factor and Age (3 years, 4 years, 5 years, adult) as a between-subjects factor. This analysis revealed only a significant main effect of Set Size ($F_{2, 382} = 5.91$, $p < .01$). Follow-up Tukey HSD tests ($p < .05$) showed that change-preference scores for SS2 were significantly higher than for SS4 and SS6; change-preference scores for SS4 and SS6 did not differ from each other. There were no differences in change-preference scores across ages. Note that the difference in scores over set sizes may be related to capacity estimates in change detection (as suggested in Chapter 2); this possible relationship will be explored further in the comparisons across tasks at the end of this chapter.

Mean change-preference scores were also compared to chance (0.5) using two-tailed t-tests. These analyses revealed that change-preference scores were significantly above chance for all set sizes for all ages (see Table 4). These results suggest that 3-, 4-, and 5-year-old children and adults have a VWM capacity of at least 6 items, although

capacity estimates may have been higher if trials with higher set size had been included. The estimates of capacity from this task show an increase relative to the performance of 10-month-olds, but also suggest a higher capacity than the typical 3-4 items shown by adults in change detection studies (e.g., Vogel et al., 2001), confirming my first prediction.

Switches between Displays. Switches between displays were counted as the number of times the participant looked from one display to the other (with or without an intervening look away) during each of the six trials. The number of switches was averaged across the two trials to the same set size. Mean switch counts are shown in Figure 13C.[9] As this figure shows, adults tended to switch much more than children; in addition, children tended to switch more in SS2 than SS4 and SS6, whereas adults tended to switch more in SS2 and SS4 than in SS6. I analyzed these data in an ANOVA with Set Size (2, 4, 6) as a within-subjects factor and Age (3 years, 4 years, 5 years, adult) as a between-subjects factor. This analysis revealed significant main effects of Set Size ($F_{2, 382} = 19.59$, $p < .001$) and Age ($F_{3, 191} = 44.86$, $p < .001$). Follow-up Tukey HSD tests ($p < .05$) showed that the number of switches were higher in SS2 than in SS4 and SS6; switches in SS4 and SS6 did not differ from each other. As can be seen in Figure 13C, this difference across set sizes is driven primarily by children's pattern of performance; for adults, switches were comparable across SS2 and SS4, but lower in SS6. Additional follow-up tests showed that adults switched between displays more than 3-, 4-, or 5-year-olds did, and the number of switches did not differ across the child age groups. As with change-preference scores, the difference in switches across set sizes may be related to

[9] To ensure that the different number of switches across set sizes and ages was not driven by the total amount of looking time, I also normalized these values by dividing the number of switches on each trial by the looking time on that trial. Analyses of these results were identical to those reported here, suggesting that the number of switches and the rate of switching are comparable.

capacity estimates in change detection; this possibility is explored further in the comparisons across tasks at the end of this chapter.

Discussion

Children and adults showed change-preference scores significantly above chance for all set sizes tested. This suggests that, in this task, participants have a capacity of at least 6 items. Analyses showed no developmental difference in preference scores, although adults looked longer at the displays and switched between displays more often. Interestingly, the only significant differences found across set sizes were between SS2 and SS4—roughly what one might expect capacity in change detection to be based on data from Riggs et al. (2006). Both change-preference and the number of switches were highest in SS2, suggesting a possible link between capacity in change detection and performance in this task. I return to this issue in the final section of this chapter when comparing performance across tasks.

One particularly important finding from this experiment is the comparison between children and the infants tested by Ross-Sheehy et al. (2003). In that study, 10-month-old infants showed a preference at SS4 but not at SS6, suggesting an adult-like capacity of approximately 4 items. In the current experiment, though, children and adults showed a higher capacity of at least 6 items. This suggests that capacity as measured by preferential looking does increase between 10 months and 3.5 years of age. It remains for future studies to explore this developmental transition further—that is, how performance changes for intermediate ages—as well as testing the upper limits of capacity in this task with older children and adults. The current study could not differentiate capacity higher than 6 items because larger set sizes were not tested here.

Higher set sizes were not tested in this experiment for three reasons. First, higher set sizes have not been tested in infancy, and thus there are no existing results to serve as comparison. Second, only nine colors have been tested in this task; if set size was

increased to, for example, eight items, there would always be at least seven colors that matched across the two displays. It seems likely that this much similarity across displays would 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 different colors (e.g., if the colors were more similar to one another, see Johnson, Spencer, Luck, & Schöner, in press-a). Again, since other colors have not been tested with infants, there would be no existing results for comparison. Lastly, increasing the number of items in each display would require decreasing the sizes of each square and/or increasing the spatial distribution of the stimuli. The effect of shrinking the stimuli is untested, and therefore may lead to unexpected changes in results. 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 children could have a harder time integrating across fixations and associating the stimuli with the correct display. Future research may address some of these limitations in order to extend this task to higher set sizes.

Experiment 2: Color Change Detection

To allow for comparisons across tasks, a subset of the participants from Experiment 1 also participated in a modified version of the color change detection task. By testing the same participants in both tasks, I can directly compare capacity estimates across the preferential looking and change detection tasks to look for correlations in performance.

Method

Participants. A subset of the participants from Experiment 1 also participated in this experiment; Table 5 shows the distribution of participants across experiments. Forty-two children completed this task: 14 three-year-olds (*M* age = 3 years, 4.68 months; *SD* = 2.35 months), 14 four-year-olds (*M* age = 4 years, 2.56 months; *SD* = 1.77 months), 14

five-year-olds[10] (*M* age = 5 years, 0.36 months; *SD* = 0.78 months). An additional 19 children were excluded from analyses for the following reasons: 8 children did not understand the task (7 three-year-olds, 1 five-year-old); 9 chose to end early (4 three-year-olds, 5 five-year-olds); 1 experimenter error (three-year-old); and 1 equipment failure (three-year-old). Fourteen adults from Experiment 1 also participated in this experiment (*M* age = 24 years, 0.91 months; *SD* = 4.67 years). All participants reported normal or corrected-to-normal vision and no history of color-blindness.

Apparatus. The change detection task was explained to children using flashcards (3" x 3") that showed SS1, SS2, and SS3. The task was then completed on an 18" Macintosh computer monitor, with the child seated approximately 24-28" from the screen. The experimenter entered the child's response on a keyboard out of the view of the child. In contrast with the standard version of this task, 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 x 4.75" wide gray rectangle. The position of the "cards" on each trial alternated from left to right. This helped reduce interference across trials. Stimuli were the same 1" x 1" colored squares used by Vogel et al. (2001); colors included red, green, blue, yellow, cyan, violet, black, and white (see Table 3 for RGB values). Colors were selected randomly on each trial with the constraint that a color could not appear twice within a single array. Stimuli could appear at any of five equally-spaced positions in a 3" diameter circle around the center of the frame. For SS2-4, the stimuli appeared in neighboring positions; all five positions were filled for SS5. For each set size, the positions were

[10] An additional group of 5-year-olds participated in a control condition to replicate the exact method used by Riggs et al. (2006); see Appendix for details.

chosen randomly for the first trial, but remained constant across the 12 trials to that set size.

Procedure. After completing the preferential looking task, children took a short break before starting the change detection task to ensure that fatigue would not impair performance. Once they were seated in the change detection room, 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 either the right or left side (alternating across trials) of a large sheet of cardstock in the experimenter's lap. The first 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 card for the trial was shown in the same location. The experimenter then asked the child if the two cards matched. After the child responded, the experimenter placed both cards side-by-side and praised or corrected the child as needed. For all children, the flashcard trials were presented in the same order: SS1 No Change, SS1 Change, SS2 Change, SS2 No Change, SS3 Change, SS3 No Change.

Once the participant understood the task, the experimenter began the computerized version of the task. The first block of trials were practice, with the sample array presented for 2 s and a delay of 900 ms. The test array remained visible until the response was entered. Children responded verbally, and the experimenter entered the response on a keyboard. When the 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.

Adults completed the same task as children, with three exceptions. First, rather than completing the flashcard trials, adults were simply shown examples of the trials during the descriptions. Second, because the longer sample array may encourage verbal recoding and/or rehearsal in adults, a verbal load was added. At the beginning of each

block, a 3-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 (Pickering, 2001). Third, adults entered their response on the keyboard themselves, either holding the keyboard in their lap or placing it on the desk below the computer monitor in front of them. The *a* key corresponded to a *no change* response and the *l* key to a *change* response; these keys were marked with stickers to help participants remember the correspondence.

The practice block included eight trials in random order: four trials to SS1 and four trials to SS2 (half of the trials were Change trials, half were No Change trials). Between each block of trials, the child was offered a short break. The order of test blocks was semi-randomly assigned (see Riggs, McTaggart, Simpson, & Freeman, 2006), 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, SS3-5 were randomly ordered across blocks 3-5. A different order was used for younger children because they often chose not to complete SS4 and 5; these high-SS 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 sample array on test trials was shortened to 500 ms, to be comparable with previous studies (Riggs et al., 2006 see Appendix for direct comparison). The total duration of the task was approximately 20-30 minutes.

Results

Participants' responses were classified as correct rejections, hits, misses, and false alarms. Figure 15 shows the distributions of these response types for each age group separately. Note that, as in Figure 9A, correct rejections and false alarms sum to 1.0 (all No Change trials) and hits and misses sum to 1.0 (all Change trials). As this figure shows,

correct rejections and hits (i.e., correct responses) were the most common responses across ages and set sizes. In addition, performance generally decreased (i.e., errors increased) with set size, especially for children.

For each set size, three measures were computed from participants' responses: accuracy (A'), response criterion bias (β), and capacity. Note that, although most 5-year-olds and all adults completed SS4 and SS5, many 3- and 4-year-old did not due to difficulty keeping children motivated to complete blocks with harder trials; therefore, I analyzed performance across SS1-3 only. 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 as follows (Aaronson & Watts, 1987):

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

$$\text{If } H < FA\text{: } A' = \tfrac{1}{2} - (FA - H) * (1 + FA - H) / 4FA * (1 - H)$$

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). Accuracy scores were computed separately for each set size block for each participant. Mean accuracy scores across set sizes for each age are shown in Figure 16A; for comparison, percent correct is shown in Figure 16B. As can be seen, accuracy followed similar patterns to percent correct; analyses reported below did not differ for accuracy versus percent correct, so only accuracy results are reported.

As Figure 16A shows, accuracy was highest for adults and lowest for 3-year-olds, and decreased across set size for children. An ANOVA with Set Size (1, 2, 3) as a within-subjects factor and Age (3 years, 4 years, 5 years, adult) as a between-subjects factor

revealed significant main effects of Set Size ($F_{2, 104} = 41.35$, $p < .001$) and Age ($F_{3, 52} = 17.71$, $p < .001$), which were subsumed by a significant Set Size x Age interaction ($F_{6, 104} = 5.75$, $p < .001$). Tests of simple effects for each Age group separately revealed significant Set Size main effects for 3-year-olds ($F_{2, 26} = 12.43$, $p < .001$), 4-year-olds ($F_{2, 26} = 21.59$, $p < .001$), and 5-year-olds ($F_{2, 26} = 14.69$, $p < .001$), but not for adults. 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. Additional tests of simple effects separately for each Set Size showed significant Age main effects at all set sizes (SS1, $F_{3, 52} = 3.90$, $p< .05$; SS2, $F_{3, 52} = 9.07$, $p < .001$; SS3, $F_{3, 52} = 28.92$, $p < .001$). Follow-up Tukey HSD tests showed that adults' accuracy was higher than only 3-year-olds' in SS1, but higher than all children's accuracy in SS2 and SS3; accuracy did not differ across children in any set size. Thus, the Set Size by Age interaction is largely driven by a different pattern over Set Sizes for children than adults, as well as a divergence of children from adults at higher Set Sizes.

Response Criterion Bias (β). Response criterion bias was calculated as follows (e.g., Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; see also, Snodgrass & Corwin, 1988):

$$\beta = -.5 * [z(\text{proportion hits}) + z(\text{proportion false alarms})]$$

Scores range from -2.33 to 2.33, with zero reflecting no bias (including perfect performance), negative scores indicating a tendency to respond *change* and positive scores indicate a tendency to respond *no change*. Response criterion bias scores were computed separately for each set size block for each participant.[11] These scores are shown in Figure 17A; as this figure shows, bias tended to increase over set size, and only

[11] Because *z*-scores can not be computed on proportion values of 0 or 1, these scores were replaced with 0.01 and 0.99, respectively.

differed across ages at higher set sizes. An ANOVA with Set Size (1, 2, 3) as a within-subjects factor and Age (3 years, 4 years, 5 years, adult) as a between-subjects factor revealed no significant effects. Because the figure suggests that bias differed between 5-year-olds and adults in SS4 and SS5, I ran an additional ANOVA including only these age groups across all five set sizes; this analysis also revealed no significant effects. Thus, the apparent differences in bias across set sizes were not statistically robust.

In general, when these scores are reported in change detection tasks, they are averaged across set sizes and participants to arrive at mean scores for each experimental group. Inspection of children's data, however, showed notable individual differences in response criterion bias, as well as a tendency for bias to change over set sizes. Because the bias score can take on negative or positive values, it is possible that a mean bias of zero may result from two different patterns of bias across blocks that averaged out to zero. To achieve a more complete picture of response criterion bias across set sizes and participants, I chose to categorize participants' scores for each set size into one of three bias groups: "*change* bias" ($\beta < -0.7$); "*no change* bias" ($\beta > 0.7$); or "unbiased" ($-0.7 < \beta < 0.7$). Note that I chose these category boundaries such that the "unbiased" category includes participants who had the same number of *change* and *no change* responses, or that the number of each response differed by only two (e.g., five *change* responses and seven *no change* responses; this would be the case in a block where a participant missed only one trial). The numbers of participants falling into these categories are shown in Figure 17B, separated by age and set size. As this figure shows, all adults fell into the no bias category. In addition, children tended to show no bias in the smaller set sizes. As set size increased, however, more children showed bias, especially *no change* biases. Thus, although the magnitude of biases were too small to cause large enough differences to yield significant effects in the ANOVA, it is clear that there are changes in response biases over development and set sizes.

Capacity Estimates. Figure 18 shows capacity estimates across set sizes (bars) separately for each age group; note that capacity can, at maximum, equal the set size in each block. Because of this limit on capacity estimates, it is not reasonable 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 (see Olsson & Poom, 2005). Mean "maximum" capacity estimates, separated by age group, are also shown in Figure 18 (line). As can be seen in this figure, capacity estimates tend to be similar across set sizes that are near or beyond capacity for that age. For example, 3-year-olds show comparable average capacity estimates for SS2 and SS3, and the maximum estimate for this age group is approximately 2 items; similarly, average capacity estimates for 5-year-olds seem to peak around SS3, which corresponds to the maximum estimate of just under 3 items. Adults, on the other hand, showed near-ceiling performance in SS1-SS4, with average capacity estimates of about 1-4 items (respectively); in SS5, however, the average estimate drops below ceiling and is just over 4 items, which corresponds to the maximum estimate of approximately 4.5 items.

I analyzed mean maximum capacity estimates in a one-way ANOVA with Age (3 years, 4 years, 5 years, adult) as a between-subjects factor, revealing a significant main effect ($F_{2, 30} = 54.81$, $p < .001$). Follow-up Tukey HSD tests ($p < .05$) showed capacity did not differ across 3- and 4-year-olds, but increased significantly for 5-year-olds and again for adults. Thus, as Figure 18 shows, capacity generally increased between 3 and 5 years and into adulthood.

Discussion

Accuracy scores and capacity estimates showed that change detection performance improved significantly over development. In addition, response biases suggest that most participants were unbiased in their responses, but that children were

more likely to show bias than adults, especially at higher set sizes. A key result from this experiment was the increase in maximum capacity estimates over development: children's capacity increased from 2 to 3 items between the ages of 3 and 5 years and adults showed capacity of approximately 4.5 items.[12] In all of these cases, capacity estimates from change detection were lower than the 6 items estimated from preferential looking, supporting the difference across tasks predicted by the DFT. Note that this difference across tasks cannot be driven by underestimation of capacity from the change detection task—even adults' performance dropped from ceiling when set size exceeded capacity (i.e., in SS5). Moreover, by using the maximum estimate of capacity across set sizes, the capacity estimate reflects each participant's best performance across blocks.

Experiment 3: Shape Change Detection

A second subset of the participants from Experiment 1 (see Table 5) participated in a shape change detection task. The goal of this experiment was to provide a second test of the predicted differences across preferential looking and change detection, and to see if the capacity limits found in Experiment 2 are specific to color, or if similar patterns will be found 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. In addition, comparisons across tasks between this experiment and capacity for preferential looking with colors (Experiment 1) will test whether performance is separable across tasks, or if there is a common source of variance across visual feature dimensions.

---

[12] The estimates derived here for 5-year-olds are higher than those reported by Riggs et al. (2006); I discuss possible sources of this difference in the control condition for this experiment in the Appendix.

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), 14 four-year-olds (*M* age = 4 years, 3.93 months; *SD* = 0.61 months), 14 five-year-olds[13] (*M* age = 5 years, 2.21 months; *SD* = 2.09 months). An additional 10 children were excluded from analyses for the following reasons: 7 children did not understand the task (4 three-year-olds, 3 four-year-olds); 2 chose to end early (three-year-olds); and 1 experimenter error (three-year-old). Fifteen adults from Experiment 1 also participated in this experiment (*M* age = 21 years, 4.30 months; *SD* = 2.90 years). All participants reported normal or corrected-to-normal vision and reported no history of color-blindness.

Apparatus. The apparatus was identical to that used in Experiment 2 with one exception: rather than colored squares, the stimuli were white shapes (Figure 19).

Procedure. The procedure was identical to that used in Experiment 2 with one exception: the sample array was presented for 2 s for all children (rather than only 500 ms for 5-year-olds). The increased presentation length was to allow children time to encode the more complex feature of shape (compared to color).

Results

As in Experiment 2, participants' responses were classified as correct rejections, hits, misses, and false alarms. Figure 20 shows the distributions of these response types for each age group separately. Note that, as in Figure 9A, correct rejections and false alarms sum to 1.0 (all No Change trials) and hits and misses sum to 1.0 (all Change trials). As this figure shows, correct rejections and hits (i.e., correct responses) were the

---

[13] A second group of 5-year-olds who did not participate in Experiment 1 also participated in this task, to ensure that completing the preferential looking task before the change detection task did not impair change-detection performance. Details of this condition can be found in the Appendix.

most common responses across ages and set sizes. In addition, performance generally decreased (i.e., errors increased) with set size, especially for children. Relative to Experiment 2, performance is slightly lower, as has been suggested by adult studies comparing memory for shapes and colors (e.g., Wheeler & Treisman, 2002). One particularly striking difference across experiments is in 5-year-olds' performance in SS3: in Experiment 2, their performance was quite high and fairly similar across trial types (i.e., correct rejections compared to hits); in the current experiment, however, performance was relatively low for *change* trials (i.e., hits). Because hits figure prominently into capacity estimates, this suggests that 5-year-olds' capacity estimates will be notably lower in this experiment. Moreover, this difference across trial types should be reflected in more *no change* response biases for this age group.

For each set size, three measures were computed from participants' responses: accuracy (A'), response criterion bias (β), and capacity. Note that, as in Experiment 2, many 3- and 4-year-old did not complete SS4 and SS5 due to difficulty keeping children motivated to complete blocks with harder trials; therefore, I analyzed accuracy and response criterion bias across SS1-SS3 only. For each participant, however, I included data from all completed set sizes to compute capacity estimates (described below).

Accuracy (A'). Accuracy was computed as in Experiment 2. Figure 21 shows mean accuracy scores (A) and percent correct (B) across set sizes separately for each age group. As in Experiment 2, both showed similar patterns, so only accuracy results are reported here. As can be seen in Figure 21, accuracy was again highest for adults and lowest for 3-year-olds, and generally decreased across set size. An ANOVA with Set Size (1, 2, 3) as a within-subjects factor and Age (3 years, 4 years, 5 years, adult) as a between-subjects factor revealed significant main effects of Set Size ($F_{2, 106} = 92.97$, $p < .001$) and Age ($F_{3, 53} = 46.80$, $p < .001$), which were subsumed by a significant Set Size x Age interaction ($F_{6, 106} = 4.31$, $p < .01$). Tests of simple effects for each age group showed a significant Set Size main effect for each age (3-year-olds, $F_{2, 26} = 31.62$, $p <$

.001; 4-year-olds, $F_{2, 26} = 24.16$, $p < .001$; 5-year-olds, $F_{2, 26} = 33.26$, $p < .001$; adults, $F_{2, 28} = 8.48$, $p < .01$). Follow-up Tukey HSD tests ($p < .05$) revealed different patterns across the age groups. For 3-year-olds, accuracy in SS1 was higher than in SS2 and SS3, while it 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.

Additional tests of simple effects separately for each Set Size showed significant Age main effects at all set sizes (SS1, $F_{3, 53} = 4.05$, $p< .05$; SS2, $F_{3, 53} = 21.33$, $p < .001$; SS3, $F_{3, 53} = 39.54$, $p < .001$). Follow-up Tukey HSD tests showed that adults' accuracy was higher than 3- and 4-year-olds' in SS1, but higher than all children's accuracy in SS2 and SS3; accuracy only differed across children in SS2, where 5-year-olds scored better than 3-year-olds. Thus, the Set Size by Age interaction was driven by different patterns of Set Size effects across the different age groups.

Response Criterion Bias (β). Response criterion bias was calculated as in Experiment 2 for each set size for each participant, and then analyzed as in Experiment 2. Figure 22A shows mean response criterion bias across set sizes separately for each age. As this figure shows, biases were larger than in Experiment 2, and tended to increase with set size. Moreover, adults tended to show smaller biases than children. I analyzed these scores in an ANOVA with Set Size (1, 2, 3) as a within-subjects factor and Age (3 years, 4 years, 5 years, adult) as a between-subjects factor. This analysis revealed significant main effects of Set Size ($F_{2, 106} = 8.99$, $p < .001$) and Age ($F_{3, 53} = 7.26$, $p < .001$), which were subsumed by a significant Set Size x Age interaction ($F_{6, 106} = 3.13$, $p < .01$). Tests of simple effects on each age group separately showed significant main effects of Set Size at both 4 years ($F_{2, 26} = 3.48$, $p < .05$) and 5 years ($F_{2, 26} = 10.21$, $p < .01$). In both cases, bias increased over set size. Thus, the interaction was driven by changes over set size for 4- and 5-year-olds, but not for 3-year-olds and adults. In addition, follow-up Tukey HSD tests ($p < .05$) on the overall Age main effect showed that mean bias differed

significantly between all children and adults, as well as between 3 and 4 years and between 4 and 5 years. Overall, 4-year-olds had the highest mean bias scores, followed by 5-year-olds, adults, and 3-year-olds. This suggests that 4- and 5-year-olds showed mostly *no change* bias (as suggested by Figure 20), which decreased over development; 3-year-olds, on the other hand, may have shown more *change* bias.

To investigate these biases further, I again categorized participants as showing "*change* bias", "*no change* bias", or "unbiased". The number of participants falling into these categories is shown in Figure 22B. As this figure shows, most participants were unbiased, especially for smaller set sizes (although more showed bias overall compared to Experiment 2). For the larger set-sizes, the number of *no change* biases increased, although 3-year-olds tended to show more *change* biases. This figure confirms the pattern suggested by the overall ANOVA presented above.

Capacity Estimates. Figure 23 shows capacity estimates across set sizes (bars) separately for each age group; recall that capacity can, at maximum, equal the set size in each block. As in Experiment 2, I also computed a "maximum" capacity for each participant; the mean maximum capacity estimates are also shown in Figure 23 (line). As this figure shows, capacity estimates tend to be similar across set sizes that are near or beyond capacity for that age. Specifically, children showed similar average capacity estimates for SS2 and up, and the maximum estimates for each of these age groups were approximately 2 items. Capacity estimates for adults, on the other hand, did not drop from ceiling until SS4 and SS5, and their maximum estimate was approximately 4 items.

I analyzed mean maximum capacity estimates in a one-way ANOVA with Age (3 years, 4 years, 5 years, adult) as a between-subjects factor, revealing a significant main effect ($F_{2, 30} = 4.87$, $p < .05$). Follow-up Tukey HSD tests ($p < .05$) showed capacity did not differ among children, but adults' capacity was significantly higher than each of the other age groups. Thus, the similar estimates of capacity between 3 and 5 years shown in Figure 23 were not statistically different, although the higher capacity for adults was.

Discussion

As in Experiment 2, accuracy scores showed that change detection performance increased over development, although capacity estimates did not reflect this improvement. As described above, this is likely due to the fact that performance on *change* trials (i.e., hits) was considerably lower than in Experiment 2. In addition, responses biases suggest that participants showed more bias relative to Experiment 2, especially for 4-year-olds, who showed *change* biases that increased over set size. In both experiments, however, children tended to show more bias than adults. A key finding from this experiment is again the maximum capacity estimates: children's capacity increased from 1.75 to 2.5 items between the ages of 3 and 5 years and adults showed capacity of approximately 4 items. In all of these cases, capacity estimates from change detection were again lower than the 6 items estimated from preferential looking, supporting the difference across tasks predicted by the DFT. Note that these capacity estimates were also slightly lower than those derived in Experiment 2, suggesting that there may be differences in capacity for different types of visual stimuli. In general, children seemed to find memory for shapes harder, presumably because the stimuli are more complex than colored squares. Still, memory capacity for shapes and colors could be related, as they are both part of the VWM system.

Comparisons across Tasks

Experiments 1-3 confirmed the first two predictions of the DFT account of VWM capacity: children and adults showed capacity estimates of 6 items in preferential looking, which is higher than infants show in this task, as well as the commonly accepted adult-like capacity of 3-4 items; these capacity estimates were also higher than the capacity estimates derived from the change detection task for the same individuals. The third prediction of the DFT is that performance in both tasks depends on the same

underlying VWM system. This prediction is tested in a multiple regression analysis across tasks for all participants who completed both tasks.

Method

Participants. The 118 children and adults who participated in both preferential looking and change detection (across experiments) were included in these analyses (see Table 5): 27 three-year-olds (*M* age = 3 years, 5.47 months, *SD* = 2.13 months), 28 four-year-olds (*M* age = 4 years, 3.25 months, *SD* = 1.47 months), 35 five-year-olds (*M* age = 5 years, 1.03 months, *SD* = 1.65 months; note that this includes the control condition for Experiment 2), and 27 adults (*M* age = 22 years, 3.88 months, *SD* = 3.96 years). To ensure that there were no differences between these participants and those who were included only in Experiment 1, I compared all performance measures from Experiment 1 across these two groups. In all cases, this comparison did not reach statistical significance (ps > 0.15). Thus, participants included in these analyses did not differ from those who are not included.

Variable Selection. The primary measure of interest from change detection is capacity; as such, the maximum capacity estimate for each individual was the dependent measure used in these analyses. In the preferential looking task, however, there is no direct calculation of capacity. Rather, change preference score have typically been used to infer capacity; because there is a separate score for each set size, all three scores were included as predictors in this analysis. Moreover, as described in Chapter 2, the DFT suggests that the number of switches in looking between the two displays in preferential looking can provide an index of the underlying working memory representations, and thus should also be related to capacity. As with change preference score, each individual contributed three switch scores (one for each set size), and all were included as predictors. Finally, because capacity in change detection clearly varied across age groups (see Figures 17 and 22), I included age as a predictor as well. Thus, in this analysis I

tested which of the following variables predicted change detection capacity: age; change preference scores for SS2, SS4, and SS6; and switches for SS2, SS4, and SS6.

Results and Discussion

Simple correlations among these factors are shown in Table 6. Change detection capacity was significantly correlated with most factors, with the strongest effect of age, followed by switches in all three set sizes. Note that the correlation between capacity and change preference in SS2 is one of the first indications from a looking task that the magnitude of a preference score may be an indication of the underlying memory representation. This is further supported by correlations among the preference scores for different set sizes. Given the relative weakness of these correlations, however, these findings are merely suggestive.

I conducted a stepwise multiple regression analysis to determine whether characteristics of performance in preferential looking were predictive of capacity in change detection. Given that capacity clearly increased with age, I selected age as the first predictor in this analysis. Note that, although age was a categorical variable in the analyses of Experiments 1-3, I converted it to a continuous variable for this analysis to capture the full range of ages in each group. Because I had no *a priori* expectations for the ordering of the remaining variables, I conducted two regression analyses to test both orders.

As described above and shown in Table 6, age accounted for a significant proportion of the variance in change detection capacity (standardized $\beta = 0.75$, $R^2 = 0.55$, $F_{1,116} = 144.62$, $p < .001$). For the remaining variables, the order of the variables did not impact the results: in both cases, only SS4 switches significantly accounted for additional variance. The combined model of age and SS4 switches provided a significantly better fit than age alone (standardized $\beta = 0.15$, $R^2 = .19$, F change$_{1, 115} = 4.24$, $p < .05$; for age, standardized $\beta = 0.64$, adjusted $R^2 = 0.56$). Recall from Experiment 1 that the analyses of

switches revealed significant main effects of both set size and age, driven by higher switches in SS2 and for adults. As Figure 13C showed, however, children and adults showed different patterns over set size: children's switches dropped between SS2 and SS4, whereas adults' switches dropped between SS4 and SS6. Note that these drops in switches roughly correspond to the mean change detection capacity estimates for the different age groups (2-3 items for children, 4.5 items for adults). Thus, switches in preferential looking appear to provide an index of change detection capacity.

This correlation across tasks strongly supports the DFT proposal that both tasks tap the same working memory system, even though they produce different capacity estimates. Indeed, the results of Experiments 1-3 alone do not suggest that such correlations would exist; there were no developmental changes in preference scores (and therefore capacity estimates) in Experiment 1. This could have been interpreted as support for the proposal 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 single unified model, I was able to predict 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 preferential looking performance. Moreover, the fact that switches provide an index of the underlying memory representations further supports the DFT account of how looking behavior relates to the formation of working memory in a task. I explore this relationship further in the next chapter, where I quantitatively fit the data from both tasks over development.

Figure 13. Preferential looking data separated by age and set size: (A) total looking time in seconds; (B) change-preference scores (dashed line indicates chance = 0.50); (C) number of switches between displays. Error bars show standard error.

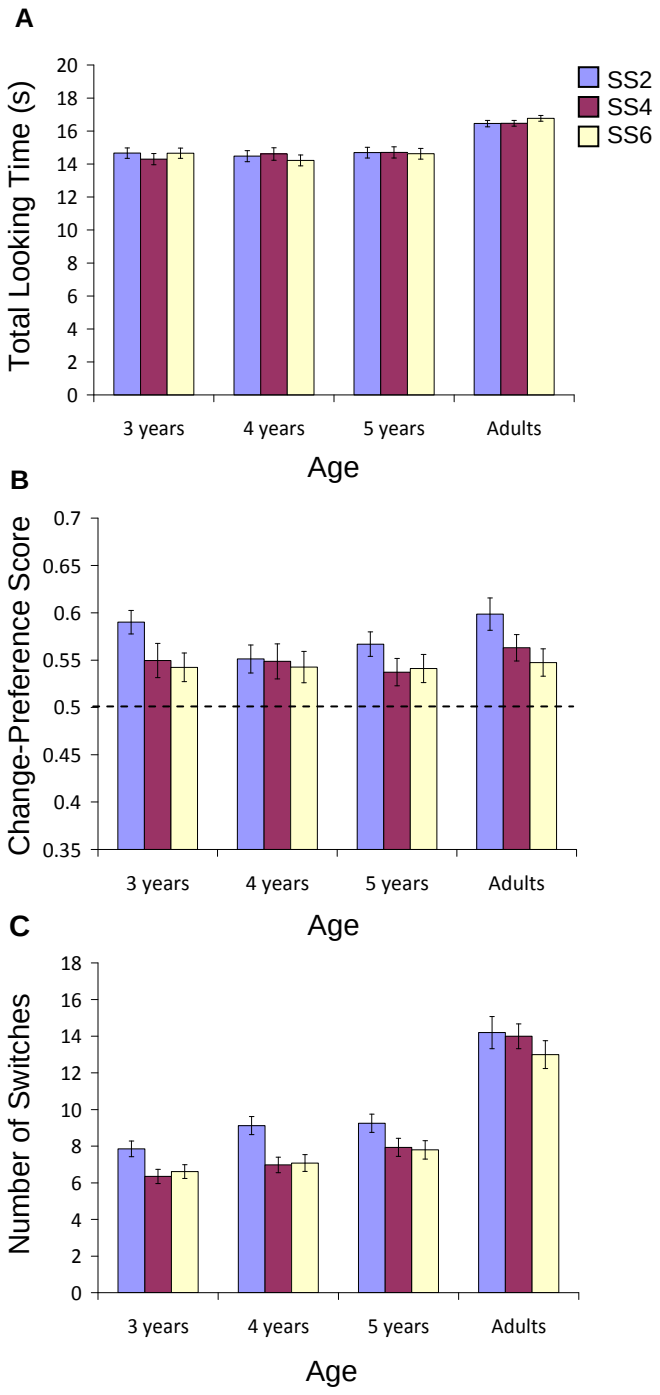


Figure 14. Sample trials in Experiment 2.

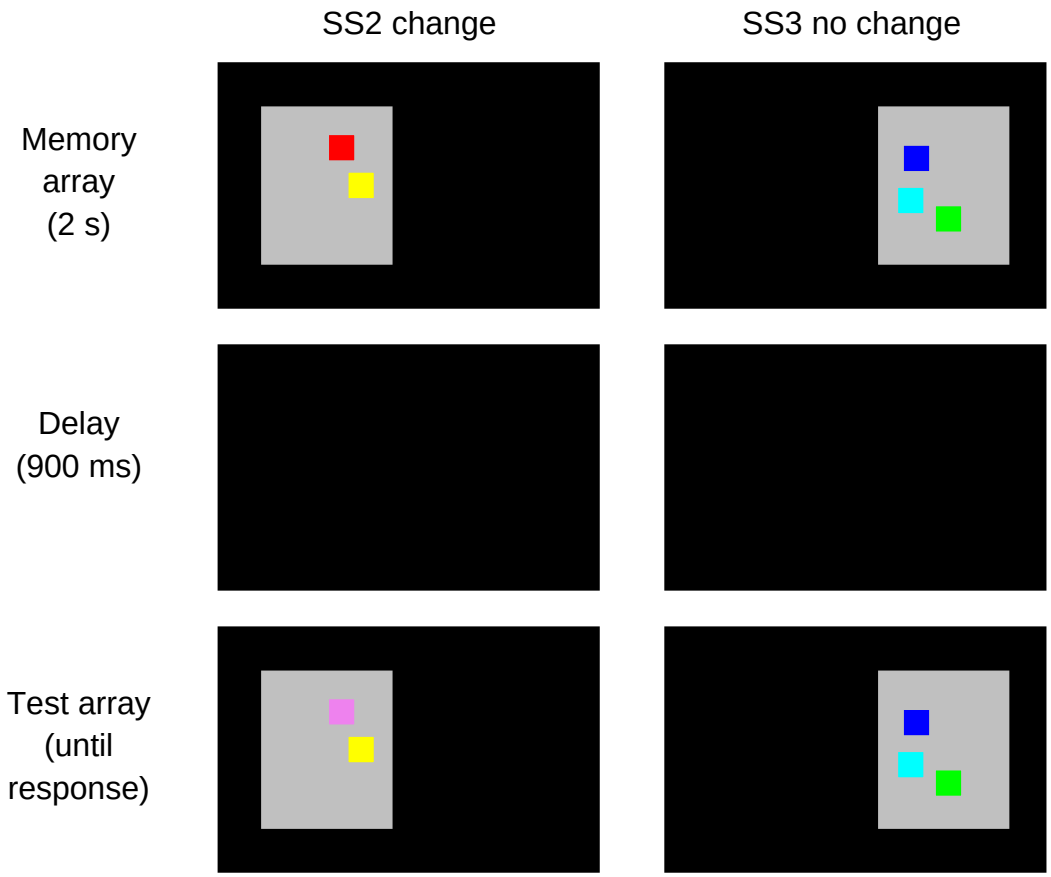


Figure 15. Response distributions across set sizes for each age group in Experiment 2.

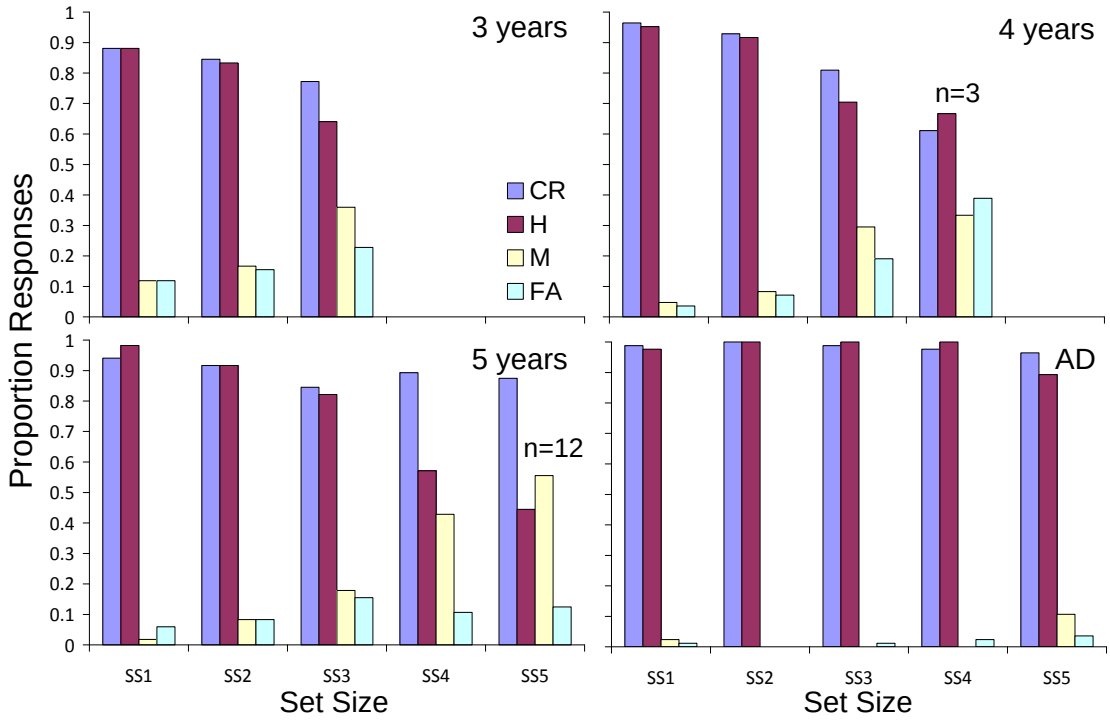


Figure 16. Performance in Experiment 2 across Set Sizes, separately for each Age group: (A) mean accuracy and (B) mean percent correct. Note that chance equals 0.5 for accuracy and 50% for percent correct. Error bars show standard error.

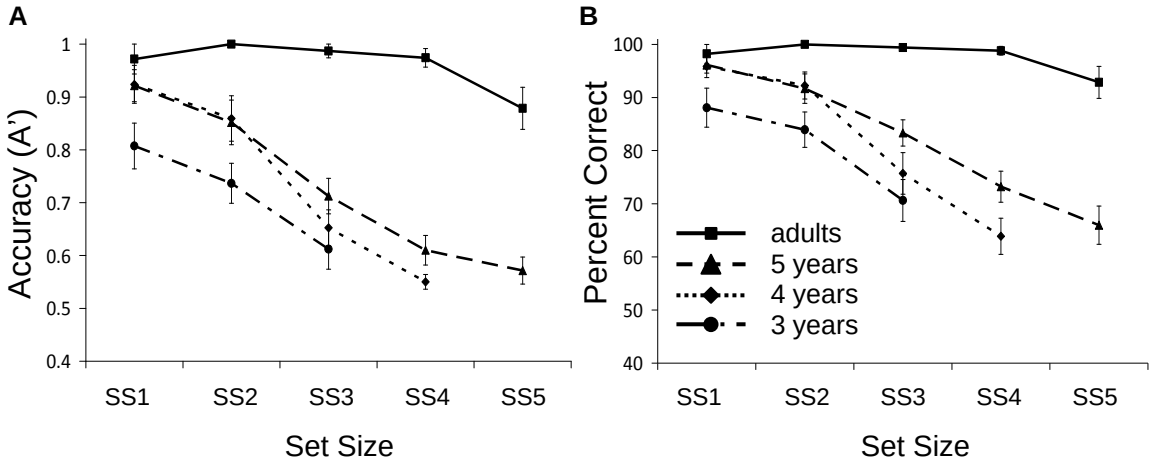


Figure 17. Response criterion bias in Experiment 2: (A) mean scores across Set Sizes, separated by Age group; (B) number of participants showing each category of bias. Error bars show standard error.

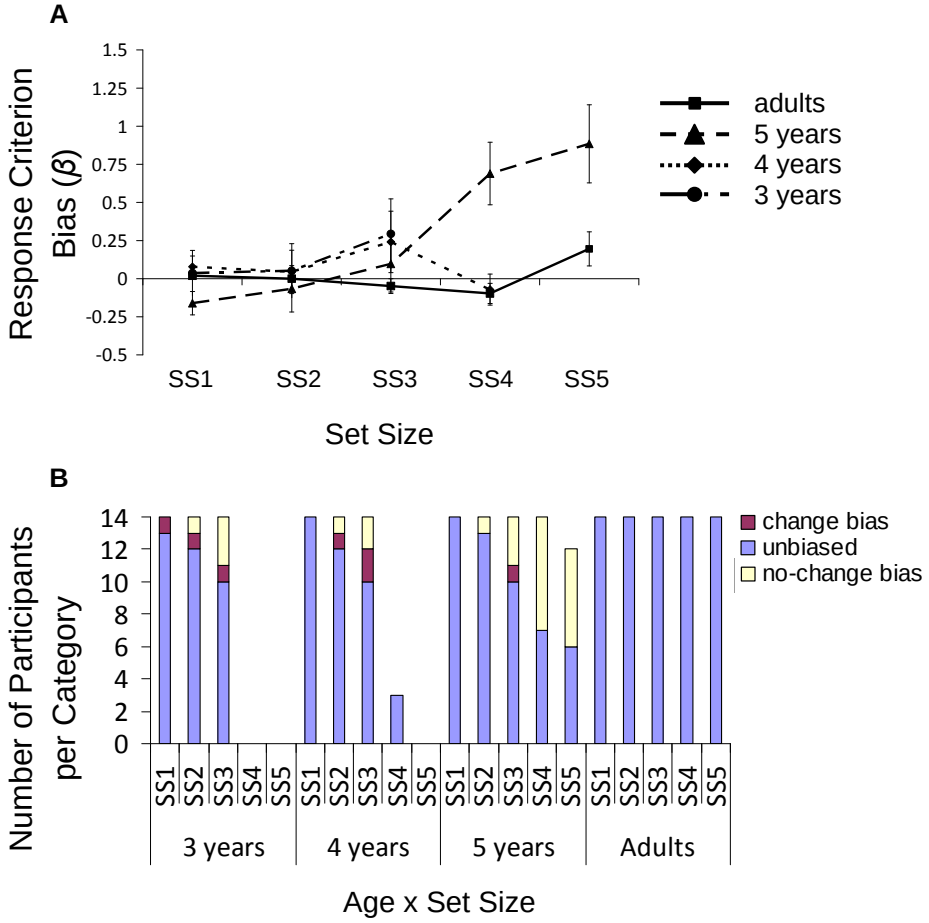


Figure 18. Mean capacity estimates from Experiment 2 across Set Sizes (bars) and mean maximum capacity for each participant (line), separately for each Age group. Error bars show standard error.

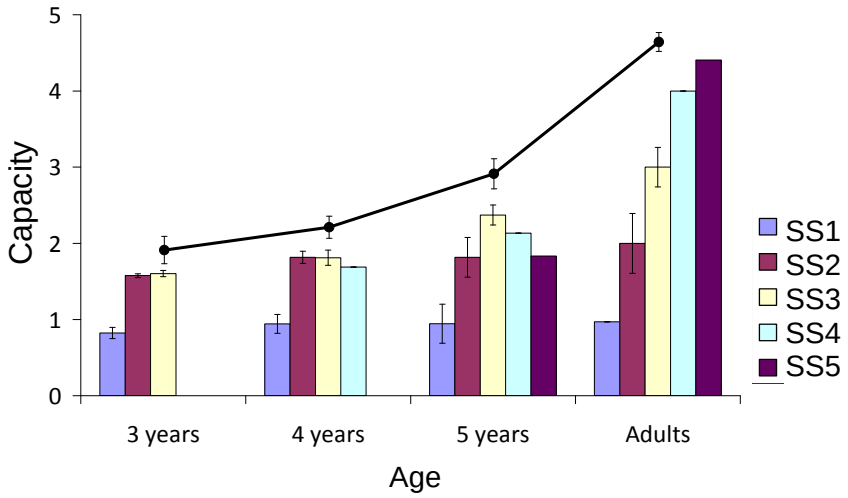


Figure 19. Shape stimuli from Experiment 3.

Figure 20. Response distributions across set sizes for each age group in Experiment 3.

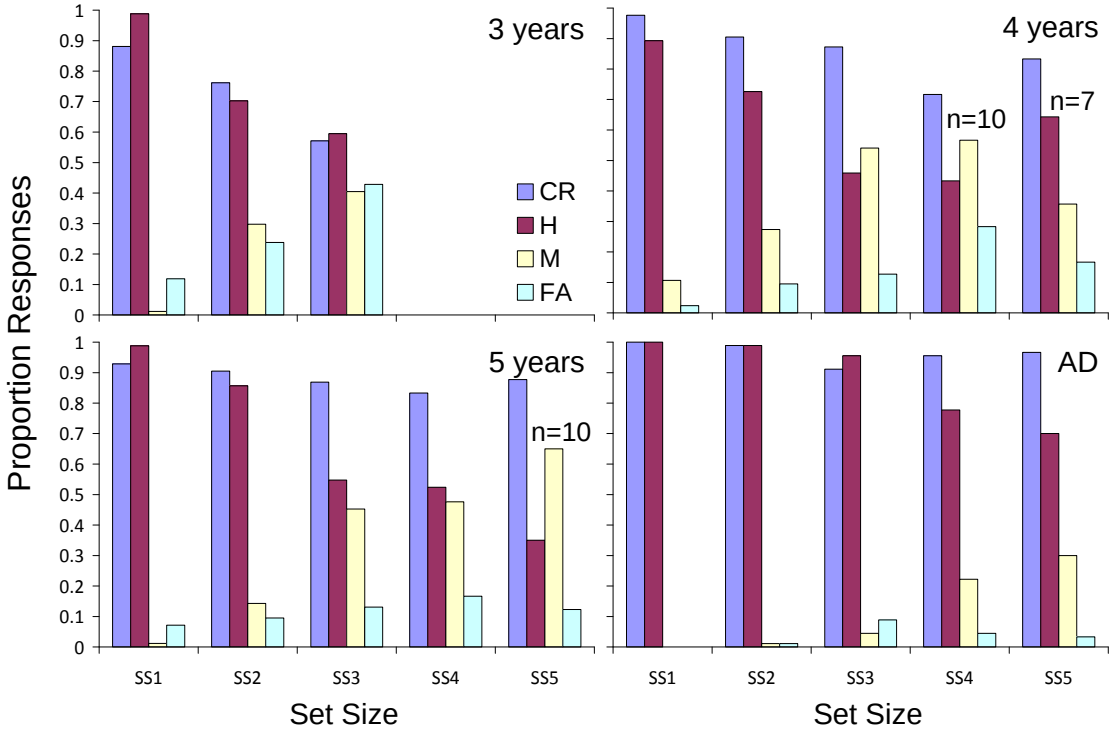


Figure 21. Performance in Experiment 3 across Set Sizes, separately for each Age group: (A) mean accuracy and (B) mean percent correct. Note that chance equals 0.5 for accuracy and 50% for percent correct. Error bars show standard error.

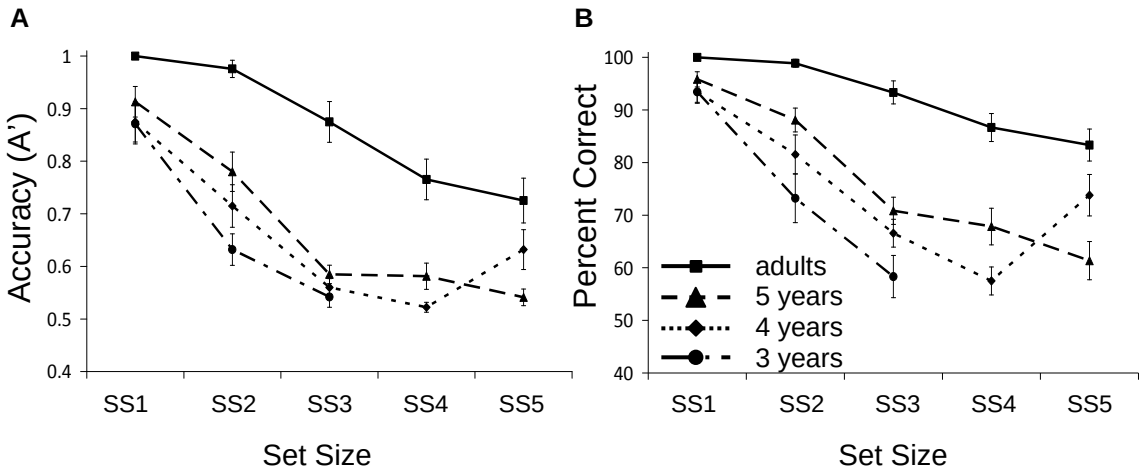


Figure 22. Response criterion bias in Experiment 3: (A) mean scores across Set Sizes, separated by Age group; (B) number of participants showing each category of bias. Error bars show standard error.

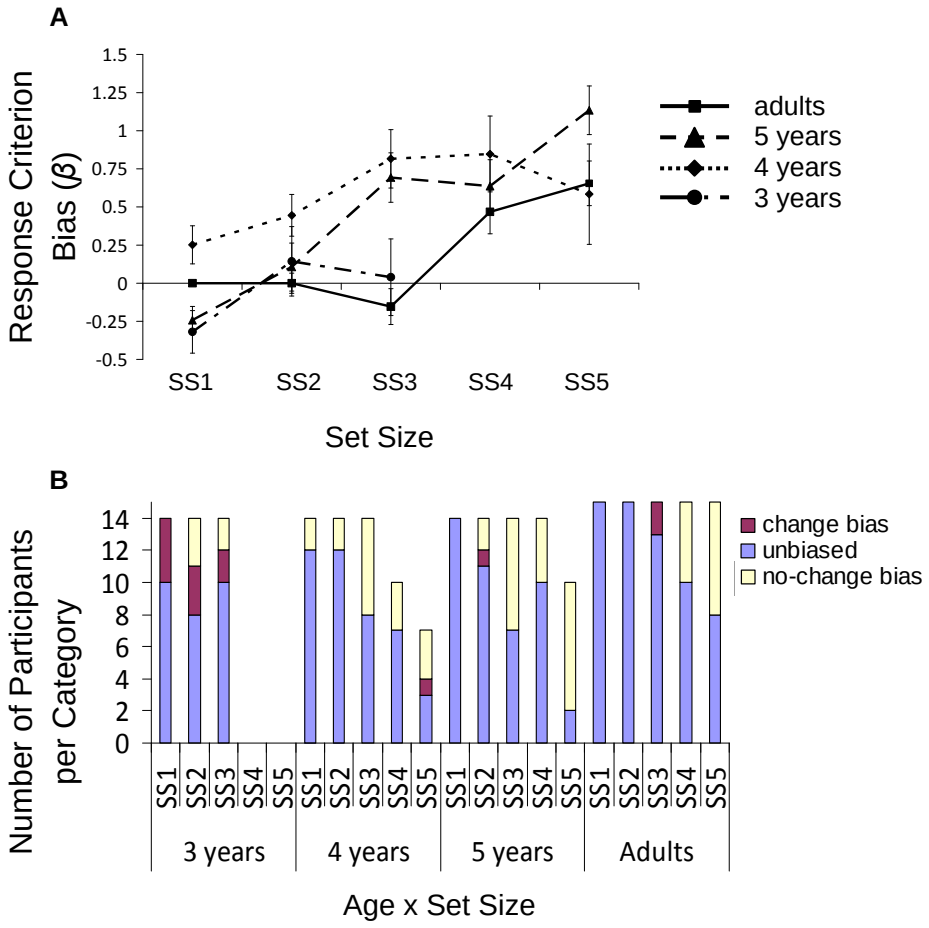


Figure 23. Mean capacity estimates from Experiment 3 across Set Sizes (bars) and mean maximum capacity for each participant (line), separately for each Age group. Error bars show standard error.

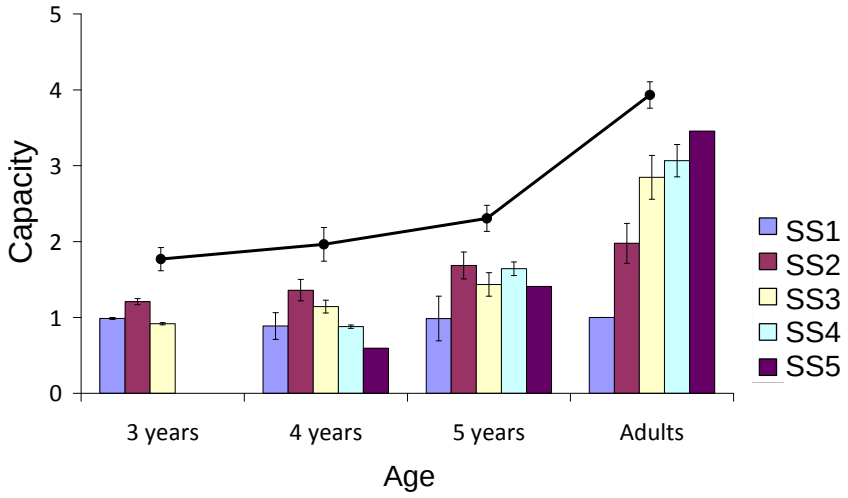


Table 3

RGB Values for Stimuli in Experiments 1 and 2

Color	R	G	B
black	0	0	0
blue	0	0	255
brown	153	51	0
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

Note. See Ross-Sheehy et al. (2003) and Vogel et al. (2001) for approximate CIE values.

Table 4

T-Test Comparisons of Change-Preference Scores with Chance (0.50) for Experiment 1

		Set Size		
Age	*df*	2	4	6
3 years	55	7.28***	2.75**	2.80**
4 years	43	3.46**	2.64*	2.57*
5 years	54	5.13***	2.57*	2.77**
Adults	39	5.76***	4.53***	3.29**

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

Table 5

Participants Included in Analyses for Experiments 1, 2, and 3

		Experiment				
	Total	1 only[a]	1 and 2	2 only	1 and 3	3 only
3 years	57	29	14	0	13	1
4 years	44	16	14	0	14	0
5 years	57	19	24[b]	0	12	2
Adults	42	13	12	2	15	0
Total	201	77	64[c]	2	54[c]	3

[a] These participants were only included in Experiment 1 for the following reasons: 47 were enrolled in experiments not reported here (9 3-year-olds, 13 4-year-olds, 12 5-year-olds, 13 adults); 19 participated in Experiment 2 and 10 in Experiment 3, but were excluded from analyses (see Experiment 2 and 3 method sections for details).

[b] Of these 5-year-olds, 14 participated in the primary condition Experiment 2 and 10 in the control condition (see Appendix).

[c] These 118 participants were included in the cross-task analyses.

Table 6

Simple Correlations (R^2) for Comparisons across Tasks

	Variable							
Variable	1	2	3	4	5	6	7	8
1. Capacity	--	0.74***	0.36***	0.53***	0.35***	0.17*	-0.01	0.03
2. Age		--	0.43***	0.57***	0.46***	0.15	0.13	0.06
3. Switches – SS2			--	0.65***	0.60***	-0.13	0.05	-0.04
4. Switches – SS4				--	0.66***	0.05	0.08	-0.05
5. Switches – SS6					--	0.11	0.08	-0.10
6. Preference Score – SS2						--	0.18*	0.06
7. Preference Score – SS4							--	0.25**
8. Preference Score – SS6								--

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

CHAPTER 4

QUANTITATIVE SIMULATIONS OVER DEVELOPMENT

Experiments 1-3 showed that preferential looking and change detection lead to different capacity estimates even when testing the same individuals. In particular, capacity estimates were higher in preferential looking (at least 6 items) and did not change over development from 3 to 5 years or to adulthood. Capacity estimates in change detection, on the other hand, were low for young children (about 2 items) and increased over development (about 3 items for older children, just over 4 items for adults). In addition, cross-tasks comparisons revealed that SS4 switch rate in preferential looking was predictive of capacity in change detection, even after removing variance due to age. These results support the first three predictions of the DFT: 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 across tasks is related over development. This support of the DFT predictions suggests that these two tasks do indeed rely on the same underlying VWM system. But can the DFT produce this pattern results in both tasks over development? This chapter includes quantitative simulations of both tasks testing the fourth prediction generated in Chapter 2: development in these tasks can be captured by the unified model using the Spatial Precision Hypothesis.

To test this prediction, I first simulated the data from the change detection task, beginning with the parameters used for the simulations in Chapter 2. Next, I simulated the data from the preferential looking task using the same developmental parameters. I chose to simulate the change detection task first because the parameters follow directly from the adult simulations presented in Chapter 2. In addition, this task showed more developmental effects with a complex pattern of results across set sizes and trial types

(i.e., *change* versus *no change*); therefore, these data provide more constraints for the model.

To model the developmental data, I began with the adult parameters and then varied model parameters according to the Spatial Precision Hypothesis. Recall that the SPH posits that neural interactions become stronger and more precise over development (see Chapter 5 for further discussion). Previous work within the DFT framework has used the SPH to capture developmental changes in spatial working memory tasks between infancy and 6 years. Simulations have shown that small, quantitative changes in neural interaction are sufficient for the DFT to reproduce a complex pattern of behavioral performance observed across multiple spatial cognitive tasks over development (Schutte & Spencer, 2008; Simmering, Schutte, & Spencer, 2008; Simmering & Spencer, 2008). The specific changes I implemented are described in detail below. Critically, these changes were derived from quantitative simulations from Schutte and Spencer (2008). These researchers used the 3-layer architecture described in Chapter 2 and used to model change detection performance (see Johnson, Spencer, & Schöner, in press-b; Simmering & Spencer, 2008) to capture a complex pattern of changes in spatial memory biases between 3 and 5 years. In particular, they increased the strength of locally excitatory interactions in perceptual and working memory fields, as well as the strength of lateral inhibition from the shared inhibitory layer over development. They also increased the strength and precision of inputs to the model over development to capture changes in the precision with which stimuli are encoded by early visual processes. As I describe below, these same parameter changes can effectively capture the developmental increase in capacity reported in Chapter 3 as well as differences in performance across change detection and preferential looking tasks.

Model Architecture

The architecture consisted of the five-layer model described in Chapter 2 with the response nodes for the change detection task and the fixation system for the preferential looking task. Each of the layers in the neural field model contained 541 units. Parameter values for the fields are shown in Table 7, for the response nodes in Table 8, and for the fixation system in Table 9. The majority of parameters (31 of 48) did not vary across all simulations; these values are shown above the dashed lines in Tables 6-8. In particular, there were no changes for the following parameters: the timescale for each equation (τ_u, τ_v, τ_w, τ_r, τ_f, τ_h, $\tau_{ltmbuild}$, $\tau_{ltmdecay}$); the resting levels in each equation (h_u, h_v, h_w, h_r, h_{rest}) the width of projections within or between layers (σ_{uu}, σ_{uv}, σ_{vu}, σ_{vw}, σ_{wu}, σ_{wv}, σ_{ww}); the strength of excitatory projections between layers (c_{uw}, c_{vu}, c_{wu}); the projections from the nodes to the fields (c_{uc} /c_{wn} for change detection, c_{ul} /c_{ur} for preferential looking); the projection from PF to the *change* node (c_{cu}); the self-excitation of the fixation nodes (c_{ll} /c_{rr} /c_{aa}); the slope for the dynamic resting level of the fixation nodes (a_h); the constant input to the *away* node (c_a); the transient input to all fixation nodes (c_{trans}); and the long-term memory projections (c_{ultm} /c_{wltm}).

The remaining (17) parameters were scaled over development, following our previous work with the SPH (Schutte & Spencer, 2008; Schutte, Spencer, & Schöner, 2003; Simmering et al., 2008; Spencer, Simmering, Schutte, & Schöner, 2007). These values are shown below the dashed lines in Tables 7-9; for comparison, the scaling values are also shown in Figure 24 (note that only parameters that also varied between 3 and 5 years are shown in this figure). As this figure shows, most parameters were scaled to smaller values for early development, and the increase between 3 and 5 years was generally linear.

To move from the adult parameters to early development (that is, backwards in developmental time), I decreased the strength of local excitation (c_{uu}, c_{ww}) and the strength of lateral inhibition (c_{uv}, k_{uv}, c_{wv}, k_{wv}). The effect of this scaling on WM peaks is

shown in Figure 25, with weaker peaks early in development. Next, I decreased the strength and precision (i.e., by increasing the width) of input to the fields (c_t, σ_t,) to approximate less precise projections from early visual areas. I also increased the direct projection into WM relative to the projection into PF (i.e., decreasing the ratio), although when coupled with the decreased input strength, this still reduced the overall strength of the projections into WM relative to the adult parameters. Lastly, I increased the strength of the spatially-correlated noise within the field (c_n) to capture the noisier performance of young children. All of these changes have been used in previous implementations of the SPH (see Schutte & Spencer, 2008), and correspond to changes in the strength of connections within and between brain areas postulated by the SPH (discussed in detail in Chapter 5).

For the decision nodes, I decreased the strength of the self-excitation (c_{cc} /c_{nn}) and competition (c_{cn} /c_{nc}) between the nodes, as well as decreasing the strength of the projection from WM to the *no change* node (c_{nw}). The changes in self-excitation and competition are conceptually comparable to the changes within the layers. The decreased strength of the projection to the same node can be thought of as tuning the system to be more sensitive to novelty or to be more input-driven (relative to the adult system). Although I implemented this as a decrease in the WM-*no change* projection, a similar affect could be achieved by increasing the PF-*change* projection.

For the fixation nodes, I again decreased the inhibition between nodes (c_{lr} /c_{la} /c_{rl} /c_{la} /c_{ar} /c_{al}). I also scaled the input to the fixation nodes by increasing the strength of the projection from PF to the nodes (c_{lu} /c_{ru}) early in development and decreasing the constant input (c_{lk} /c_{rk}), that is, the input which indicated that something was present on the left or right display. As with the changes to the response nodes, this scaling made the child system more input-driven relative to the adult system because the activation from PF, which largely reflects the details of the featural input at the fixated site, had a relatively stronger influence on fixation early in development. Next, I increased the low

attractor value (h_{down}) of the h-dynamics for the nodes, which cause the nodes to return to this attractor state more slowly. Lastly, I increased the magnitude of the noise on the constant input to the nodes (c_{n_in}).

Simulation Details

Simulations were conducted in Matlab 7.4 (Mathworks, Inc.) on a PC with an AMD Athlon 2.61 GHz dual-core processor. The dynamic field equations (see Chapter 2) were integrated using the Euler method with one time step = 2 ms. For these simulations, the model was presented with one to five inputs on each trial, depending on the set size; inputs corresponded to different colors selected from a continuous 360° color space (i.e., 1° in color space = 1.5 units in the model), in which colors were separated by 40° (60 units).

Change Detection Trials

Each simulation trial began with a 200 ms relaxation period that allowed the model to reach a stable resting state. This was followed by the 2 s presentation of the sample array and a 900 ms delay interval. Next, the test array was presented in which either all items matched the sample 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. Shortly after the appearance of the test array (100 time steps or 200 ms later), the resting level of the nodes in the response layer was increased and activation was projected from PF and WM to the respective decision nodes. This allowed the activation of the nodes to build beyond threshold, provided suitably strong input from the corresponding field. Because the nodes in the response layer were coupled via strong

inhibitory connections, only a single node surpassed threshold on each trial.[14] The node whose activation reached threshold first, therefore, determined the model's response. To parallel the design of Experiment 2, each "session" for the model included 12 trials (6 *change*, 6 *no change*) in each of the five set sizes, presented in the order 2 – 1 – 3 – 4 – 5. Note that, because young children did not complete the larger set sizes, I excluded these trials for the "young" simulations. The model was then tested 50 times for each age group (equivalent to running 50 participants per age group).

Preferential Looking Trials

Again, each simulation trial began with a 200 ms relaxation time that allowed the model to reach a stable resting state. Then, transient input was sent to all fixation nodes for 150 ms, indicating that something had appeared on the displays. In addition, during the presentations of the inputs (on for 500 ms, with 250 ms "blank" delays in between), non-specific constant input (with added white noise; see Table 9) was sent to the *left* and *right* fixation nodes; this corresponded to the appearance and disappearance of stimuli on the displays. When one of these nodes pierced threshold, the specific input from that display—that is, which colors were presented on that display—was projected to the model (strongly into PF and weakly into WM as described in Chapter 2). 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

---

[14] On a small proportion of trials (< 0.01%), the decision system showed a particular type of instability: the *change* node would rise above threshold for approximately 500 time steps (1 s), but then activation of this node would drop below threshold. At this point, the *no change* node would rise above threshold for the remainder of the trial (about 500 time steps). In these cases, the *change* response was recorded. This type of instability occurred almost exclusively with the child parameters, presumably due to the weaker self-excitation of the decision nodes. Behaviorally, this could be interpreted as the model "changing its mind"; interestingly, this is a behavior that some young children exhibited in Experiments 2 and 3.

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. For each "session", the model was presented with six 20-s trials, two at each set size, in random order. The model was again tested 50 times for each age group.

Results and Discussion

In this section, I describe the results of the simulations in change detection first, followed by the result in preferential looking. Within each section, I discuss how the model captures children's and adults' performance in each task, as well as how developmental changes in the model give rise to developmental changes in performance.

Change Detection

Figure 26A-D shows the distributions of response types[15] for the change detection simulations for each "age" (data from Experiment 2 are shown in Figure 26E-H for comparison); recall 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). As this figure shows, the results were similar to the data from Experiment 2, approximating not only the overall percent correct, but also capturing the relative rates of each type of error. Moreover, the simulations yielded mean maximum capacity estimates similar to those in Experiment 2: 1.96 items with the 3 year parameters (1.90 in Experiment 2), 2.59 with the 4 year parameters (2.20), 2.91 with the 5 year parameters (2.90), and 4.11 with the adult parameters (4.61). Table 10 shows the model responses and the data from Experiment 2 for comparison; note that I only included correct rejections and hits in this table because

[15] For simplicity, I only compare the model's response distributions directly to the distributions from Experiment 2; because the other measures of performance (accuracy, response bias, and capacity) are computed directly from these distributions, the fits will be similar across measures.

misses and false alarms are simply the differences of these measures from 1.0, with identical standard deviations.

In evaluating the fit of the model simulations to the data, it is first important to note the percentages shown in Table 10 may be misleading due of the small number of trials of each type. Because there were only 6 *change* trials and 6 *no change* trials, a difference of a single trial results in a 16.67% change (i.e., performance drops to 83.33% with only one incorrect trial) in hits or correct rejections, respectively. These large jumps in percentage based on few trials make it difficult to compare performance rates using percent correct. As an alternative, I converted these numbers back to numbers of trials (e.g., 100% equals 6 trials, 83.33% equals 5 trials, etc.); now differences between the model's performance and the data from Experiment 2 can be evaluated in terms of the numbers of trials that differed. Figure 27 shows these values, with the corresponding data from Experiment 2. As can be seen in this figure and in Table 10, the model not only produced the correct levels of mean performance, but also captured the variability across participants. To quantify this fit, I calculating the root mean squared error (RMSE) for correct rejections and hits separately for each set size and age group. Thus, I compared a total of 34 means and 34 standard deviations between the model and the data. The average RMSE across the 34 means was 0.26 (range: 0.00 to 0.97), and the average RMSE across the 34 standard deviations was 0.73 (range: 0.00 to 0.73). As these values indicate, the fits of the model were quite good, within a single trial of the data for every data point.

To understand how performance was influenced by the developmental changes I implemented, I have included some sample simulations with the 3-year-old parameters (see Table 8 for details of the parameter changes). Figure 28 shows a sample simulation of a SS3 false alarm trial (i.e., a *no change* trial on which the model responded *change*) using the 3-year-old parameters. When compared to the adult-parameter simulations shown in Chapter 2, the WM peaks in this simulation are relatively broad and weak.

Because of the weaker interaction profile, the model is not able to maintain all three peaks over the delay (see “peak dies” arrow in Figure 28). Thus, when the same colors are presented in the test array, the input for the forgotten color produces a self-stabilized peak in PF (see circle in Figure 28), leading to a *change* response.

In another sample simulation, Figure 29 shows a SS3 miss trial (i.e., a *change* trial on which the model responded *no change*). Again, the weak interactions prevent the model from maintaining all three WM peaks over the delay (see “peak dies” arrow in Figure 29). On this trial, however, the forgotten color (-160°, corresponding to black) happened by chance to be the one that changed at test. The new color (violet, at -40°) that comes in at test is relatively near a color that is being maintained in WM (red, at -80°); the slight inhibitory trough caused by this WM peak—along with weak neural interactions in PF—is enough to slow down the formation of a self-stabilized peak corresponding to the new color in PF (see circle in Figure 29). By the time this peak builds in PF (see dashed arrow in Figure 29), however, the *no change* node has already passed threshold due to the activation from the peaks still held in WM. Thus, the model responds *no change* and makes an error. These sample simulations illustrate how changes in the interactions in the fields can produce errors for different reasons relative to the adult model described in Chapter 2.

These simulations of the change detection task demonstrate that the DFT can generalize from the standard task simulated in Chapter 2 to the slightly modified version we tested with adults in Experiment 2, as well as capturing the details of performance over development from 3 to 5 years. Furthermore, these results provide strong support for the Spatial Precision Hypothesis as a mechanism to drive changes over development (discussed in more detail below). As the interaction among neurons was weakened to capture early development, the stability of the attractor states in the fields decreased. As a result, WM peaks were weaker and less likely to sustain throughout the delay. In addition, self-stabilized peaks were less likely to form in PF at test. The combination of

these two effects, both resulting from weaker neural interactions, produces the appropriate decline in performance on both *change* and *no change* trials. Thus, the SPH not only provides a specific mechanism for developmental change, but also specifies the process by which this change influences performance.

To understand how memory capacity arises in the DFT across all age groups, I calculated the number of peaks present in WM at the end of the delay period on each trial; this provides an index of how many peaks the model can stably maintain in memory through the delay. Figure 30A shows the mean number of peaks across set sizes for each point in development; Figure 30B-E also shows the mean number of peaks separately for each response type. As this figure shows, the number of peaks that could be held in WM increased over development, with an average of about 3.25 in SS4 for the 3-year-old parameters.[16] This increased to around 3.5 in SS4 for 4-year-old parameters and 3.85 in SS4 for the 5-year-old parameters. For SS5, the 5-year-old parameters average over 4.5 peaks per trial, and the adult parameters are near ceiling, averaging 4.97 peaks per trial. It is interesting to note how the mean number of peaks changes over response type; for all ages, the false alarm category has the lowest number of peaks, suggesting that these responses are the most likely to result from a failure in capacity (as shown in the simulations figures in Chapter 2 as well). After false alarms, hits show the next lowest number of mean peaks, further suggesting that correct responses can result from imperfect memory of the displays.

In light of these simulations, what should one conclude about capacity estimates in these tasks? As a primary contribution to capacity, the strength of neural interactions

---

[16] Even though there were no comparable data from Experiment 2, I included simulations in SS4 with the 3 year parameters to ensure that performance remained at reasonable levels. In this set size, correct rejections were at 70.67% (SD = 19.79) and hits were at 56.67% (SD = 21.83), which fits with the pattern seen across smaller set sizes in this age group, and across set sizes in the other age groups.

limits the number of items (i.e., peaks) that can be actively maintained at one time. Note, however, that the number of peaks does not *equal* the estimated capacity (see Figure 9 and related discussion in Chapter 2). This is the sense in which the DFT simulations map most closely onto traditional ideas about capacity. However, the DFT account moves beyond IP approaches to capacity, as Figure 30 illustrates: failing to hold all of the items in memory leads to particular response types (i.e., false alarms) more than others. This contrasts with the traditional interpretation of the task, in which misses are considered to be indicative of capacity failures. Conceptual or verbal theories can only hypothesize the relationship between capacity and responses in this task; by implementing this task in a process-based model, we can directly observe the relationship between the items being held in memory (i.e., WM peaks) and the model's response.

In addition to the number of items in memory, decision factors played a considerable role in the change detection simulations, especially over development. For example, note that the model held the SS1 item in memory on every trial with the 3-year-old parameters; however, performance was only around 90% correct. Errors on these trials, then, must result from the comparison and response process, not just from failures in capacity. Again, this could not be determined conclusively from conceptual theories, but is evident when the task is implemented in a process-based model.

These limitations of the IP account of capacity in change detection are also evident in the formula typically used to compute capacity. As Pashler (1988) acknowledged, the formula he developed is not equipped to address different sources of errors in the task. Moreover, the assumptions underlying this formula (see Chapter 2 for discussion) are not supported by the DFT simulations. This formula is derived from signal detection theory—the idea that the observer is merely monitoring for a *change* signal, and responds *no change* in the absence of that signal—which may be the wrong way to describe performance in this task. Rather, the decision process is actually about

the competition between two signals (one for *change* and one for *no change*), and this competition contributes significantly to performance.

To summarize, these simulations showed that the DFT can capture developmental changes in change detection performance by implementing the SPH. In addition, analysis of how the model performed the task provided insight into the nature of capacity limits, as well as how capacity may impact performance on *change* and *no change* trials. These results suggest the traditional IP approach to understanding capacity may not appropriately account for performance in this task. The next critical question for the model is to test the consequences of these developmental changes in the second task, preferential looking.

Preferential Looking

Figure 31A-C shows the mean total looking times, mean change preference scores, and mean switches across simulation "age" groups; the data from Experiment 1 are shown in Figure 31D-F for comparison. As can be seen in this figure, the model was able to capture the pattern of data across all three measures, all three set sizes, and all four age groups. Because these three behavioral measures in the model could vary independently, however, I compare each to data from Experiment 1 in turn.

Table 11 shows the model total looking times and the corresponding data from Experiment 1 for comparison; I again evaluated the model fits by calculating RMSE for each set size and age group. Thus, for this measure, I compared 12 means and 12 standard deviations between the model and the data. The average RMSE across the looking time means was 0.42 (range: 0.10 to 0.82), and across the standard deviations was 1.52 (range: 0.58 to 1.98). As these values indicated, the fits of the model to the mean total looking times were again quite good, within 1 s of the data for each mean. The standard deviations were not fit as well in this task as in change detection, with the model

showing less variability than the participants. There are a number of reasons this may be the case, which I discuss in detail below.

Change preference scores from the model simulations and Experiment 1 are shown in Table 12. I calculated RMSE for each set size and age group, again comparing 12 means and 12 standard deviations between the model and the data. The average RMSE across the preference score means was 0.02 (range: 0.01 to 0.05) and across the standard deviations was 0.03 (range: 0.00 to 0.05). Thus, the fits of the model to the means and standard deviations for change preference scores were quite good.

Finally, switches from the model simulations and Experiment 1 are shown in Table 13. Again, I calculated RMSE for each set size and age group, to compare 12 means and 12 standard deviations between the model and the data. The average RMSE across the switch means was 0.74 (range: 0.49 to 1.21) and across the standard deviations was 1.37 (range: 0.87 to 2.09). As with total looking, the fits of the means were quite good, within a single switch for all but 1 of the 12 means. However, the fits for the standard deviations were not as good, with the model again showing less variability than the participants.

Importantly, the model was able to capture the general pattern in switches across set sizes, with a drop in switching between SS2 and SS4 for children, and between SS4 and SS6 for adults (see Figure 31C, F). As described in Chapters 2 and 3, this provides an indication of capacity in change detection (between 2 and 3 items with the child parameters, and just over 4 items with the adult parameters). As with the change detection simulations, I will describe a few sample simulations to explore how the model is performing this task.

Figure 32 shows a simulation of a SS2 trial with the 3-year-old parameters. The activation of the nodes over the course of the trial is shown in Figure 32A. As this panel shows, the model switched between displays 11 times (i.e., there are 12 fixations that alternate = 11 switches), a relatively high number for this age group. To illustrate how the

underlying memory representation contributed to this pattern of looking, PF and WM are shown in Figure 32B-C (recall that in this "overhead" view of the fields, the fuchsia "hot spots" indicate peaks). The central point to take from this simulation is that the two peaks corresponding to the two colors on the No Change display sustain throughout much of the trial (see arrows in Figure 32C). By contrast, peaks associated with the colors from the Change display build more transiently; they are not able to sustain for very long.

Why do some peaks sustain while others do not? At first, this difference is due to chance—whichever items are fixated early have an advantage. Once there are self-sustaining peaks in WM, however, it becomes more difficult for new peaks to build, because existing peaks suppress the formation of new peaks via lateral inhibition. To illustrate this point, Figure 32D-E shows time slices through PF and WM at a point when a new peak is building in WM (see first vertical arrow across the fields). At this point in the trial, the model is maintaining fixation to the Change display, and one of the colors has just changed to a new value (at -40°). Because this peak is still building, it is not contributing as much inhibition back to PF as the peaks being sustained at -120° and 80°. This allows a self-stabilized peak in PF at that value (-40°; see circle in Figure 32D-E), which maintains fixation to this display. This illustrates one contribution to the mechanism behind the change preference: the model continues to fixate the change display when values are "new" in WM.

A second example of the same process is highlighted at the next time slice (see second vertical arrow across the fields; Figure 32F-G). In this example, the model is stably maintaining the two items from the No Change display (at -160° and 80°) while fixating two new colors in the Change display (at -80° and 40°). Over the course of this fixation, the WM peaks associated with the new colors begin to build, but much more slowly due to competition from the two No Change colors already being held in WM. During the time slice shown in Figure 32F-G, the peaks at the new colors have just emerged in WM (at -80° and 40°; see Figure 32G), which de-stabilizes one of the old

peaks (at 80°; see “dies out” in Figure 32G). This competition between peaks in WM arises due to constraints on the capacity of WM in the DFT.

Figure 32 also illustrates why switching is high for set sizes within the model’s capacity: because the model is able to form peaks from both displays relatively quickly (albeit, not necessarily simultaneously), activation in PF is quickly inhibited and fixation is released. Although switching is high, it is important to emphasize that a change preference can still emerge because peaks from the No Change side receive more support from traces in LTM_{WM}. Thus, release from fixation is rapid in simulations that switch frequently, but there can still be a difference across displays in how quickly the peaks are inhibited.

For contrast, Figure 33 shows a simulation of a SS4 trial with the 3-year-old parameters. Note, first, that the model switches infrequently, only four times over the course of the trial. As discussed previously, the low number of switches results from the limited capacity of WM. To illustrate this, consider the model’s performance during the first fixation of the No Change display (red line in Figure 33A). Because the model has a limit to how many peaks can be built at one time, the four colors being presented do not all form WM peaks at once (see dashed circle in Figure 33C); rather, the peaks build sequentially. Once built, these peaks are relatively unstable leading to fluctuations in their activation over time, which in turn affects the inhibitory projection sent back to PF. Because this projection varies through time, activation in PF is able to repeatedly rise above threshold which maintains fixation to this display even though the items are close to the self-sustaining state in WM.

Roughly 8 s into the trial, activation in PF is low enough to release fixation, and the model switches to fixate the Change display. Now there is competition between the No Change peaks and the new colors coming from this display. A time slice through the fields at this point (Figure 33D-E) shows that the two new peaks trying to build at 0° and 120° (see red circles) are slowed by the other peaks being held in WM (at -160°, 40°, and

80°). As a result, activation at these values remains above threshold in PF, maintaining fixation to this display. Eventually, however, the WM peaks associated with the new colors rise above threshold, and competition causes three of the previous No Change peaks to die out (see solid circle in Figure 33C).

The model's next fixation is back to the No Change display. Because only one of these colors (-160°) is still maintained in WM from the last fixation, the remaining three colors must build again (see red circle in Figure 33F-G). Critically, these peaks build relatively quickly as a result of input from LTM_{WM}. This, in turn, creates a relatively short look to the No Change side which results in a change preference score across the 20s simulation. Note once again, however, that change preference and switches are not isomorphic: the lower number of switches in this simulation are driven by the relatively slow build up of peaks in WM due to constraints on capacity, while the change preference reflects the fact that this build up—while slow relative to the previous simulation—is faster to the No Change display.

To summarize, as in the change detection simulations, weaker interactions in WM limit the number of peaks that can be maintained simultaneously. However, due to the more supportive nature of the preferential looking task, the combination of frequent input and LTM support allows more peaks to sustain (4 or 5, albeit briefly) than were stably maintained during the change detection delay (i.e., 3.5 on average for SS4). As a result, performance in preferential looking is not as strongly constrained by the number of peaks that can be maintained over delays. A more important factor is how quickly activation in PF is suppressed, either by quickly-building peaks, or by peaks sustaining in WM. Over development, as interactions strengthen, stability increases and more peaks may be built and maintained simultaneously. This in turn leads to faster inhibition, which causes to faster switching behaviorally. This type of stability is likely the underlying cause of the correlations across tasks revealed in Chapter 3, between switching in preferential looking and change detection capacity.

Overall Evaluation

This chapter presented quantitative simulations to test my fourth prediction, that implementing the SPH in the unified model will capture developmental changes in both preferential looking and change detection. 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 simulated 70 means and 70 standard deviations (34 each in change detection, 36 each in preferential looking), for a grand total of 140 data points. In general, these fits of the simulations to the data were very good, typically capturing the data within one unit of measure (i.e., one trial, one second, one switch).

These simulations provide strong support for the Dynamic Field Theory account of capacity in preferential looking and change detection, as well as the proposed developmental mechanism, the Spatial Precision Hypothesis. The simulations presented here are a particularly stringent test of the model—few models are tested across multiple tasks, multiple measures within tasks, or multiple points in development. Moreover, many models only simulate mean data points, but not standard deviations. Evaluating the DFT simulations compared to both means and standard deviations gives a better evaluation of the process by which the model generates behavior on each trial, not just averaged across trials and individuals. As such, the simulations presented here move beyond most other cognitive and developmental models in providing both specificity—on a trial-by-trial basis within each task—and generality—across tasks and development (see Simmering & Spencer, 2008, for discussion).

One remaining challenge for this model, however, is to also capture individual differences within each age group; this type of variance is necessary to examine whether the model shows the type of correlations across tasks described in Chapter 3. One approach to this issue would be to introduce slight variations in the parameters that were scaled over development—it is possible that such variation would be sufficient to capture individual differences (i.e., cross-task correlations). Conceptually, such variations would

correspond to differential rates of development across individuals (and across parameters). For example, there may be differences in the development of excitatory versus inhibitory connections across individuals that would lead to differences in performance across tasks. If these differences were stable within individuals, they may lead to the correlations across tasks described in Chapter 3. Testing this possibility in the model was beyond the scope of the current thesis, but will be an important task for future modeling work.

Figure 24. Developmental scaling for parameters in the fields (A) and the nodes (B). Note that lines are color coded to correspond to the columns in Tables 5-7.

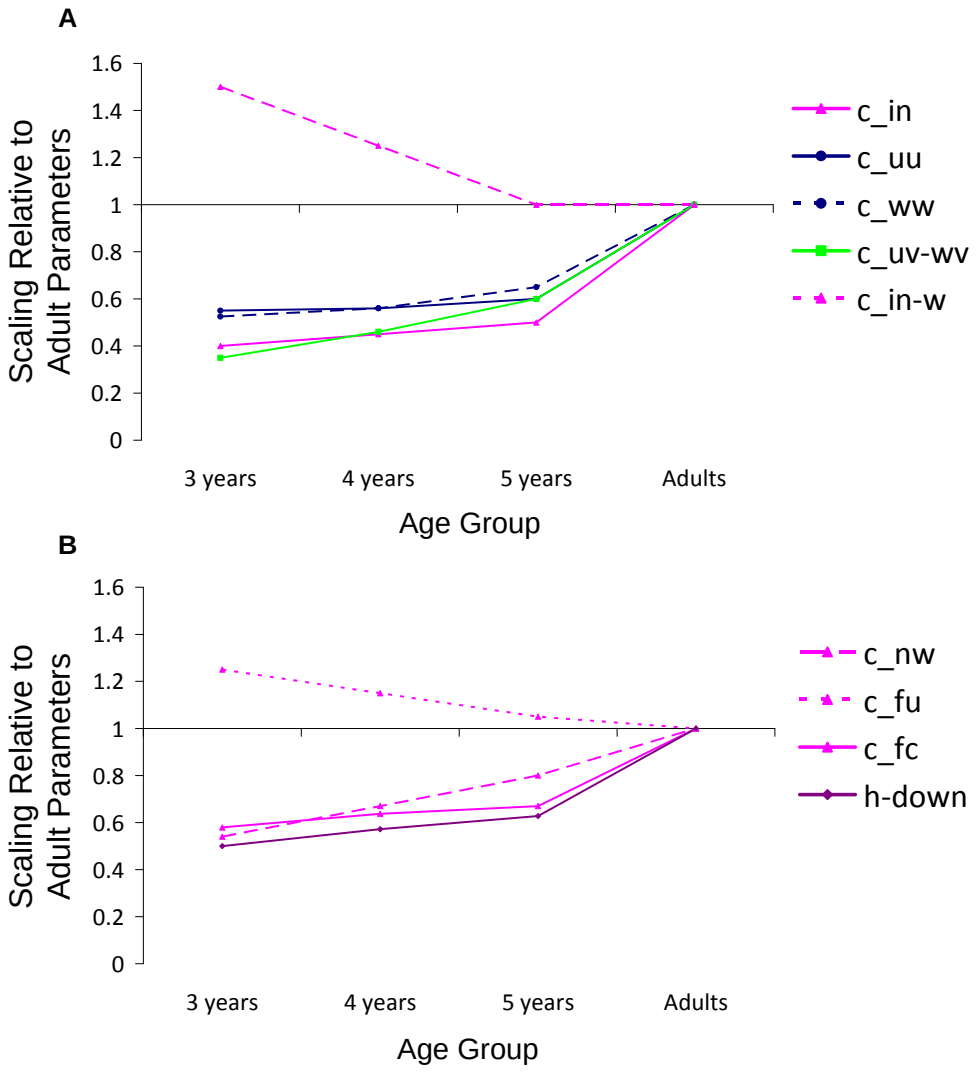


Figure 25. Changes in peaks in WM as a function of developmental changes using the Spatial Precision Hypothesis.

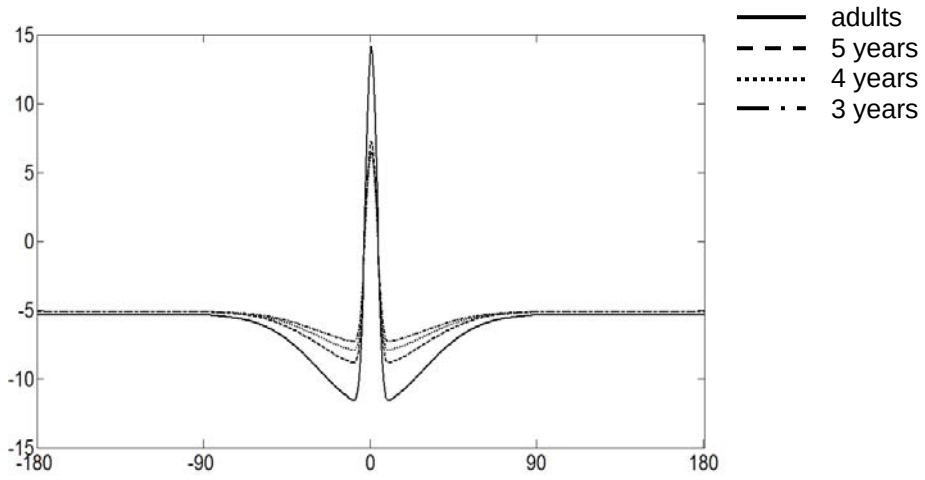


Figure 26. Response distributions across set sizes for each age group in the change detection simulations (A-D), and the comparable data from Experiment 2 (E-H), reproduced from Figure 14.

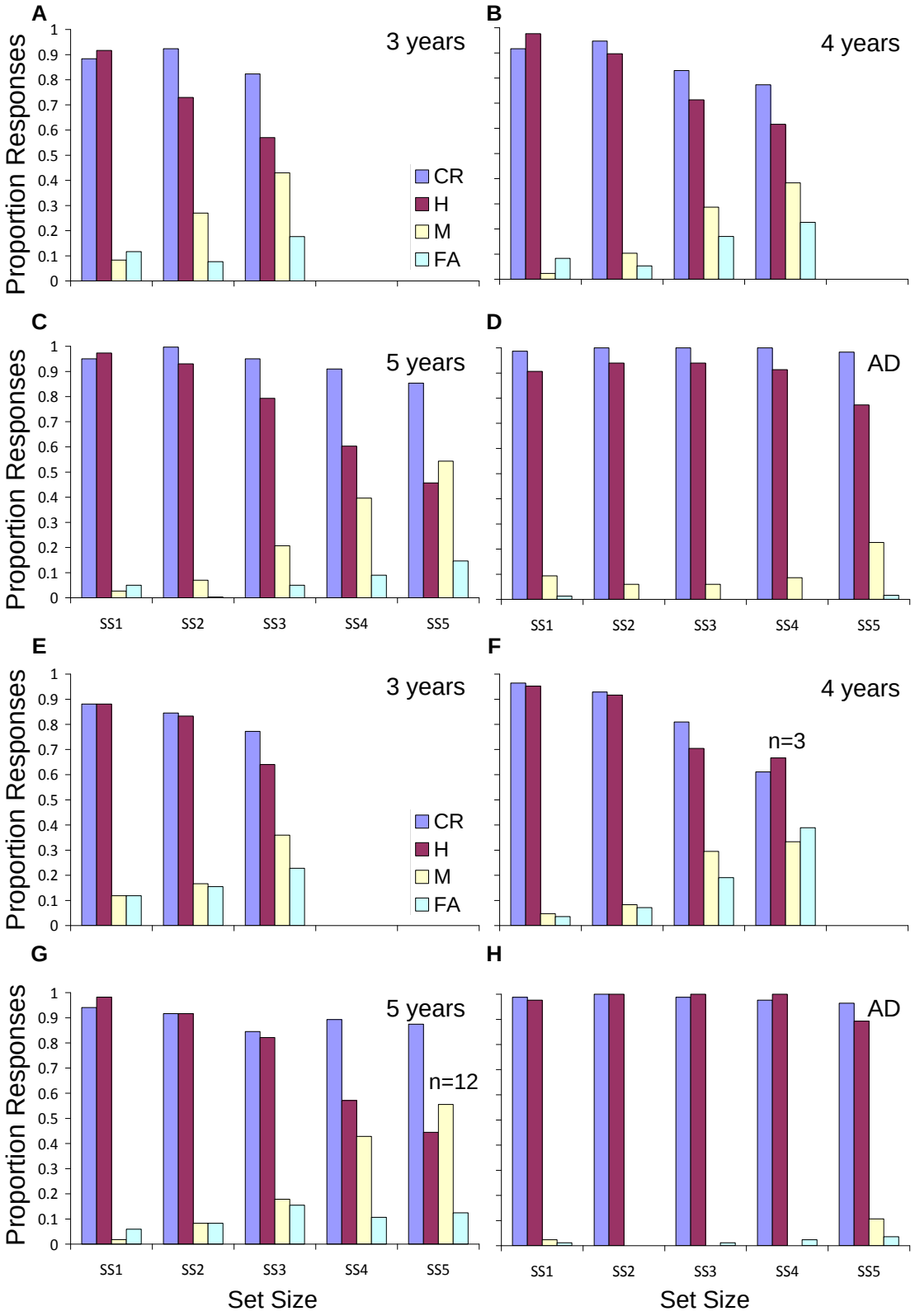


Figure 27. Responses adjusted as number of trials; correct rejections are shown on the right, hits are shown on the left. Error bars show standard error

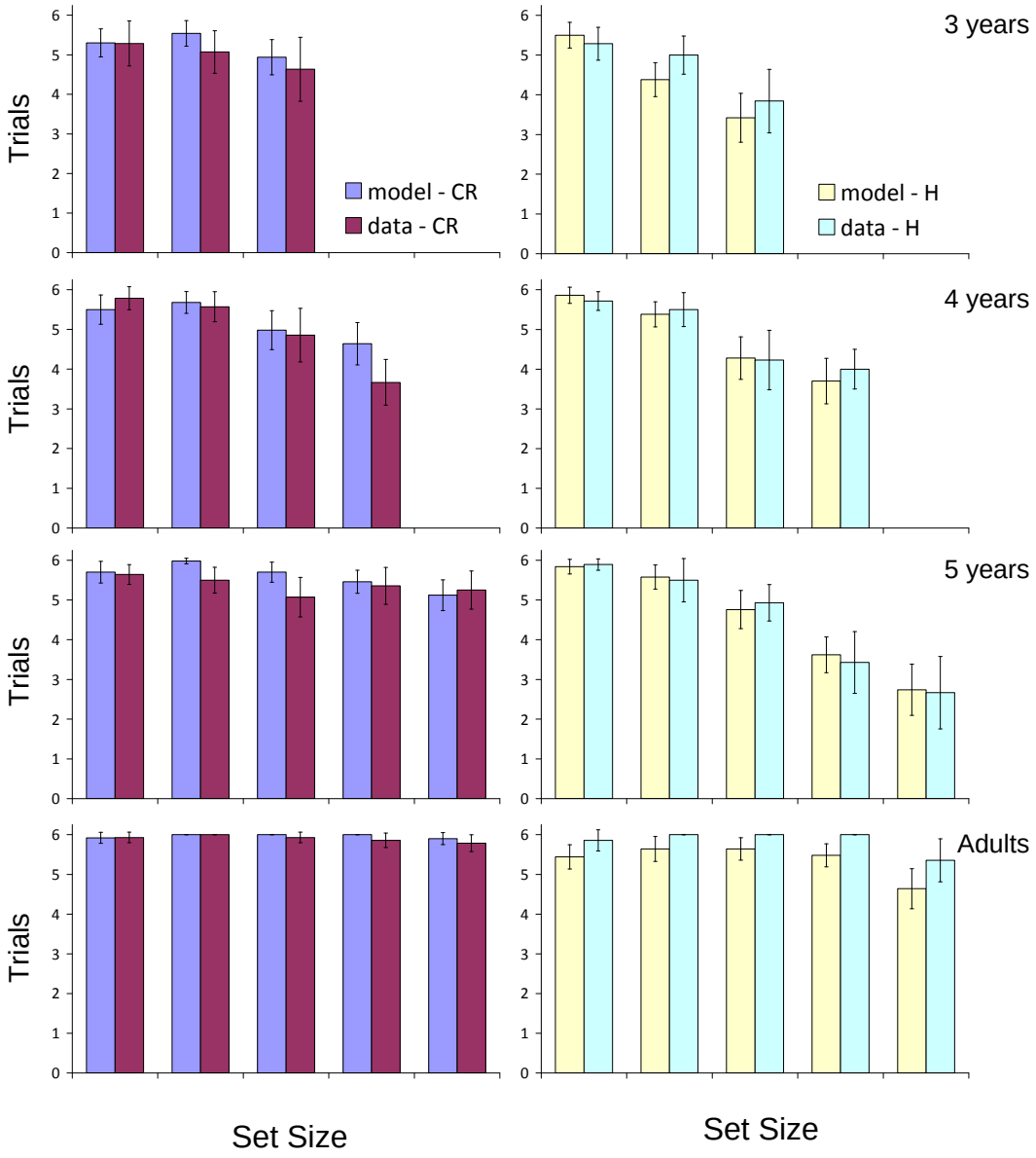


Figure 28. A simulation of the DFT with the 3-year-old parameters performing a SS3 false alarm trial in change detection. Panels and axes are as in Figure 5.

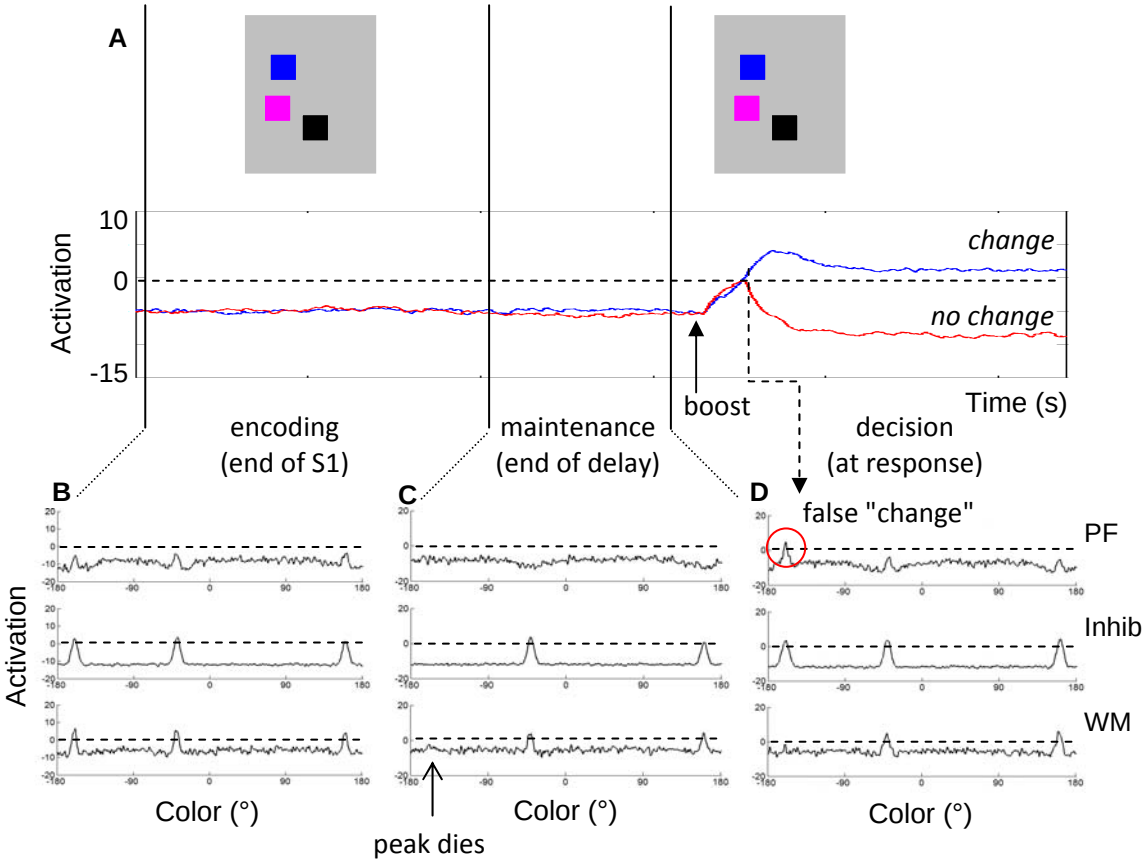


Figure 29. A simulation of the DFT with the 3-year-old parameters performing a SS3 miss trial in change detection. Panels and axes are as in Figure 5.

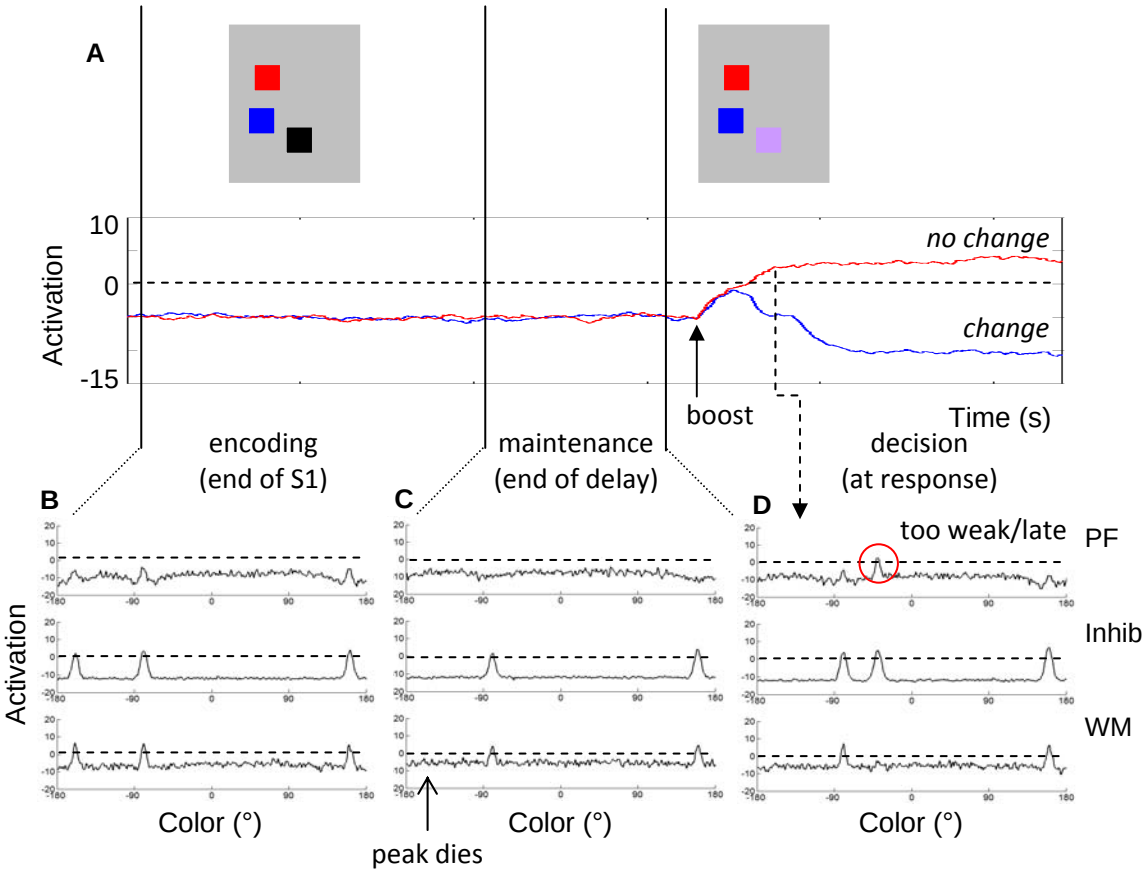


Figure 30. Mean number of peaks in WM at the end of the delay in change detection, both averaged across trials (A) and separated by response types (B), for each parameter set. Error bars show standard error.

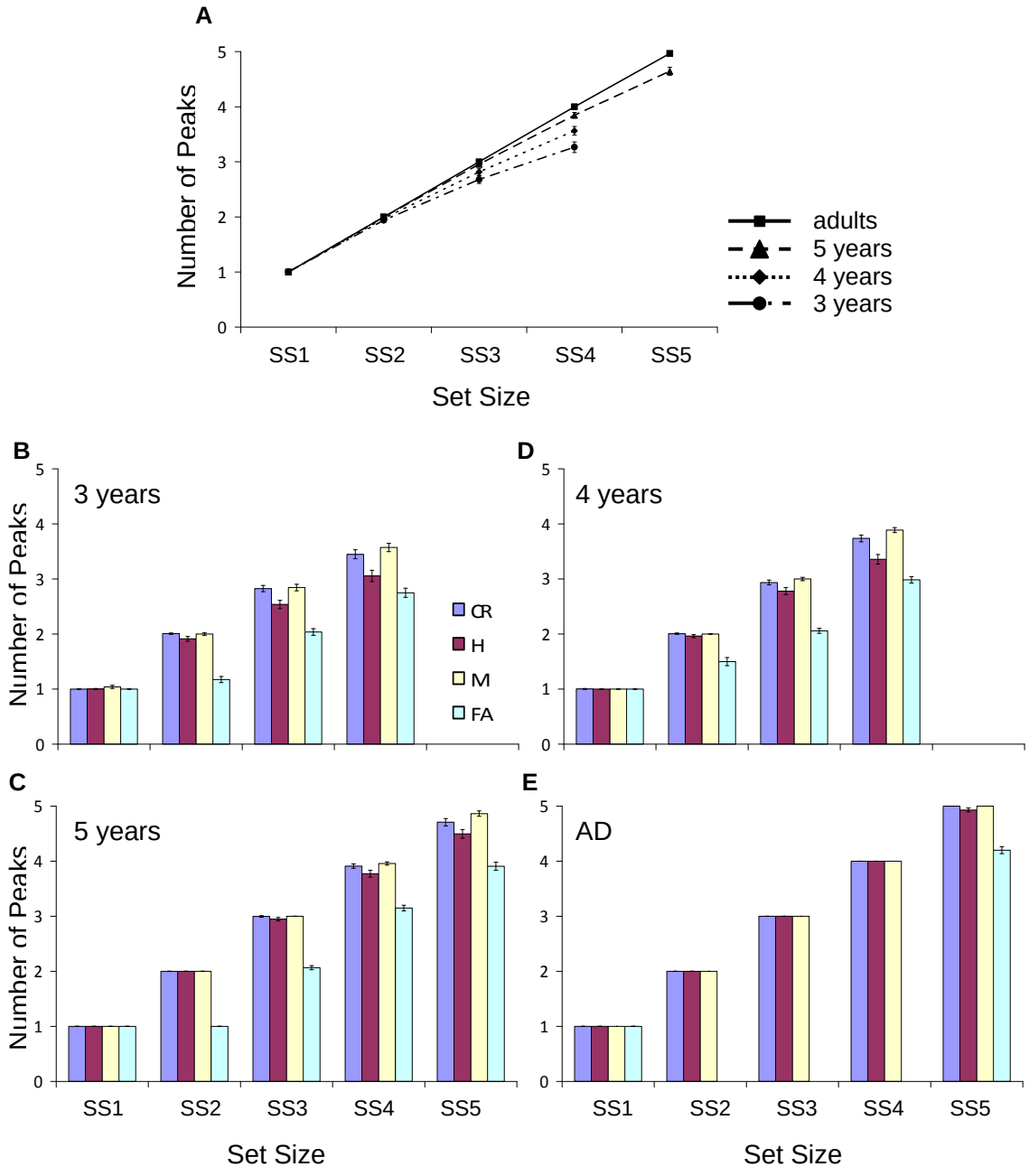


Figure 31. Simulation results for total looking time (A), change preference scores (B), and switches (C); data from Experiment 1 are reproduced from Figure 13 for comparison (D-F). Error bars show standard error.

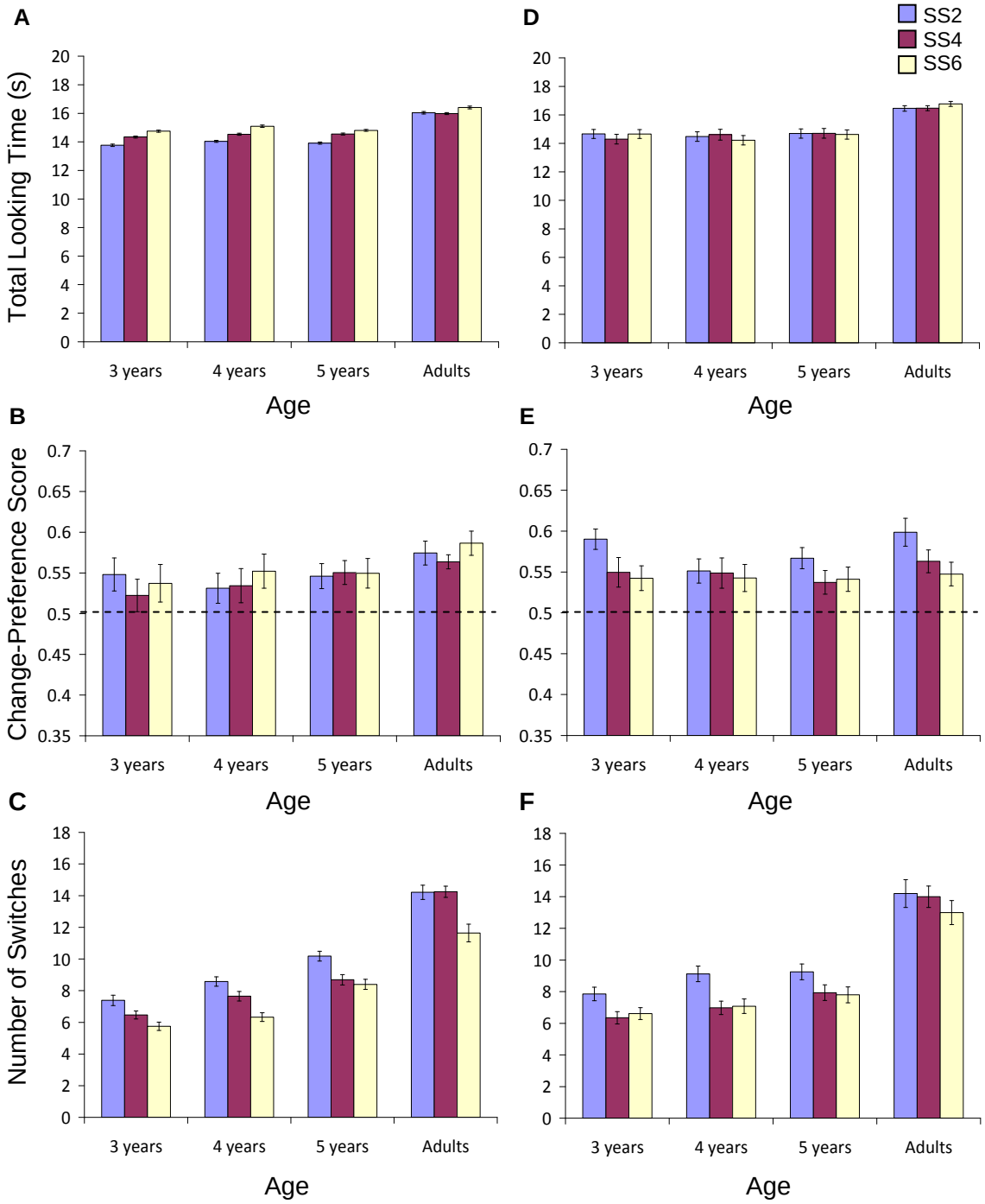


Figure 32. Sample simulation of a SS2 preferential looking trials using the 3-year-old parameters. Activation of the fixation nodes (A) and in PF (B) and WM (C) are shown with time along the *x*-axis; along the *y*-axis is activation (A) and color (B-C). Vertical arrows indicate time slices through PF and WM (D-G); in these panels, color is along the *x*-axis and activation along the *y*-axis.

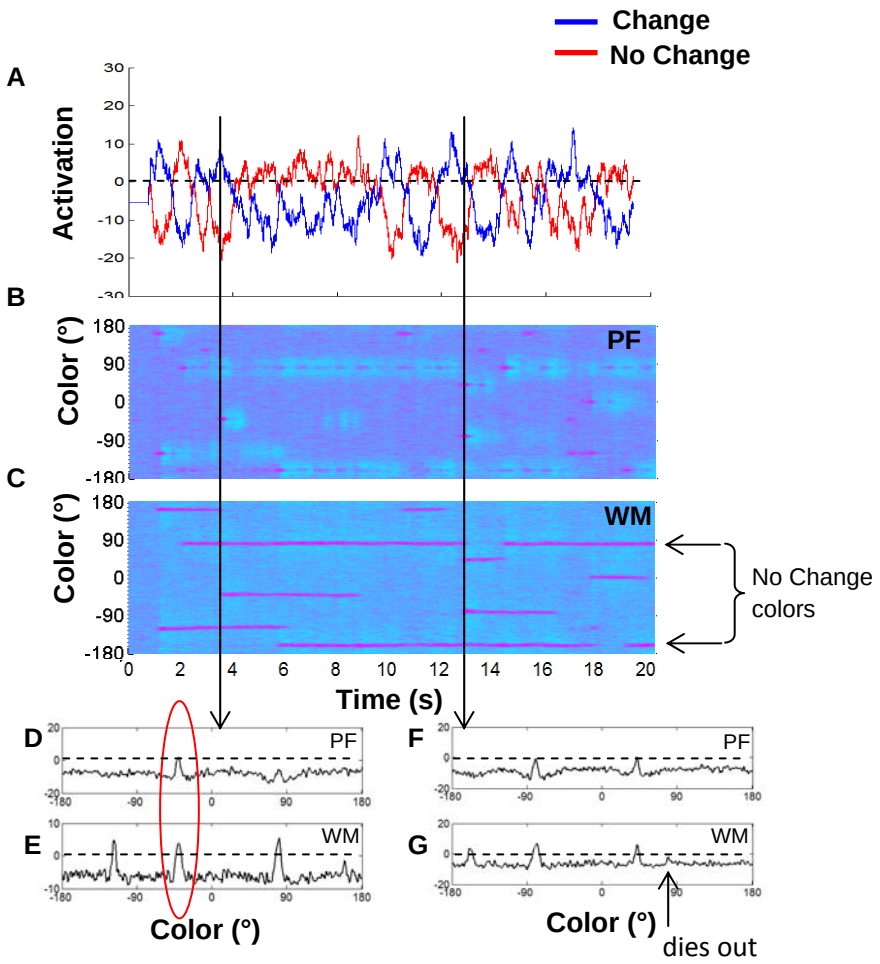


Figure 33. Sample simulation of a SS4 preferential looking trials using the 3-year-old parameters. Activation of the fixation nodes (A) and in PF (B) and WM (C) are shown with time along the *x*-axis; along the *y*-axis is activation (A) and color (B-C). Vertical arrows indicate time slices through PF and WM (D-G); in these panels, color is along the *x*-axis and activation along the *y*-axis.

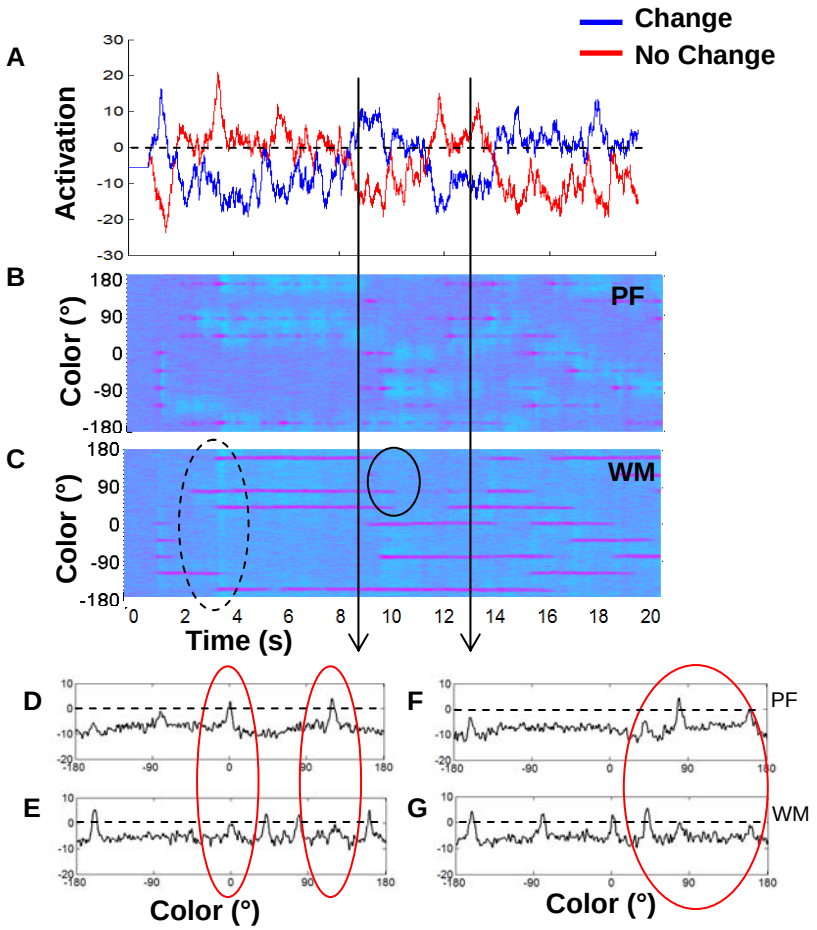


Table 7

Field Parameter Values

Layer	τ	h	self-excitation	excitatory projection(s)	inhibitory projection(s)	input	noise
PF(u)	80	-7	$c_{uu} = 3.10$		$c_{uv} = 1.85$	$c_{tar} = 20$	$q_u = 0.1$
			$\sigma_{uu} = 3$		$\sigma_{uv} = 26$	$\sigma_{tar} = 3$	$\sigma_q = 10$
					$k_{uv} = 0.05$		
Inhib(v)	10	-12		$c_{vu} = 2$			$q_v = 0.1$
				$\sigma_{vu} = 7$			$\sigma_q = 10$
				$c_{vw} = 1.95$			
				$\sigma_{vw} = 5$			
WM(w)	80	-4.5	$c_{ww} = 3.15$	$c_{wu} = 1.5$	$c_{wv} = 0.6$	$c_{tar} = 4$	$q_w = 0.1$
			$\sigma_{ww} = 3$	$\sigma_{wu} = 5$	$\sigma_{wv} = 42$	$\sigma_{tar} = 3$	$\sigma_q = 10$
					$k_{wv} = 0.08$		
PF(u)			$c_{uu3} = 1.705$		$c_{uv3} = 0.6475$	$c_{tar3} = 8$	$q_u = 0.125$
			$c_{uu4} = 1.736$		$c_{uv4} = 0.851$	$c_{tar4} = 9$	
			$c_{uu5} = 1.85$		$c_{uv5} = 1.11$	$c_{tar5} = 10$	
					$k_{uv} = 0.0375$	$\sigma_{tar} = 4.5$	
Inhib(v)							$q_v = 0.125$
WM(w)			$c_{ww3} = 1.65375$		$c_{wv3} = 0.21$	$c_{tar3} = 2.4$	$q_w = 0.125$
			$c_{ww4} = 1.764$		$c_{wv4} = 0.276$	$c_{tar4} = 2.25$	
			$c_{ww5} = 2.0475$		$c_{wv5} = 0.36$	$c_{tar5} = 2$	
					$k_{wv} = 0.06$	$\sigma_{tar} = 4.5$	

Note. Scaled developmental parameters are shown below the dotted line; numbers at the end of each subscript indicate the corresponding age group when the value changed between 3 and 5 years. Only parameters that were changed relative to the adult values are shown in this section; all other values were identical.

Table 8

Response Node Parameter Values

Node	τ	h	self-excitation	excitatory projection(s)	inhibitory projection(s)	input	noise
Change (r_c)	80	-5	$c_{cc} = 1.5$	$c_{uc} = 1$	$c_{nc} = 14$	$c_{cu} = 1.005$	$q_c = 0.075$
No Change (r_n)	80	-5	$c_{nn} = 1.5$	$c_{wn} = 1$	$c_{cn} = 14$	$c_{nw} = 0.14$	$q_n = 0.075$
			$c_{cc} = 1.125$		$c_{nc} = 10.5$		
			$c_{nn} = 1.125$		$c_{cn} = 10.5$	$c_{nw3} = 0.0756$	
						$c_{nw4} = 0.0938$	
						$c_{nw5} = 0.112$	

Note. Scaled developmental parameters are shown below the dotted line; numbers at the end of each subscript indicate the corresponding age group. Only parameters that were changed over development are shown in this section; all other values were identical to the adult parameters.

Table 9

Fixation Node Parameter Values

Node	τ	h[a]	self-excitation	excitatory projection(s)	inhibitory projection(s)	input	input noise[b]
Left(f_l)	80	-5.5	$c_{ll} = 2$	$c_{ul} = 1$	$c_{rl} = 20$ $c_{al} = 20$	$c_{lu} = 1$ $c_{const} = 14.5$ $c_{trans} = 2$	$q_{in} = 20$
Right(f_r)	80	-5.5	$c_{rr} = 2$	$c_{ur} = 1$	$c_{lr} = 20$ $c_{ar} = 20$	$c_{ru} = 1$ $c_{const} = 14.5$ $c_{trans} = 2$	$q_{in} = 20$
Away(f_a)	80	-5.5	$c_{aa} = 2$		$c_{ra} = 20$ $c_{la} = 20$	$c_{const} = 2.5$	$q_{in} = 20$
Left(f_l)/ Right(f_r)					$c_{rl3} = 19$ $c_{rl4} = 19$ $c_{rl5} = 19$	$c_{lu3} = 1.25$ $c_{lu4} = 1.15$ $c_{lu5} = 1.05$ $c_{const3} = 8.40$ $c_{const4} = 9.24$ $c_{const5} = 9.715$	$q_{in3} = 25$ $q_{in4} = 25$ $q_{in5} = 25$

Note. All inhibitory projections (c_{rl}, c_{al}, c_{lr}, c_{ar}, c_{ra}, c_{la}) were scaled the same over development; for simplicity, these values are listed only once. Scaled developmental parameters are shown below the dotted line; numbers at the end of each subscript indicate the corresponding age group. Only parameters that were changed over development are shown in this section; all other values were identical to the adult parameters.

[a] The fixation node equations used dynamic resting levels; this value indicates the resting level in the absence of input. The resting level changed according to the following function: $\tau_h \cdot \dot{h}_f = -a_{hf} h_f + h_{rest} + \Lambda(h_f) \cdot h_{down}$ (see Chapter 2 for further details). The slope (a_h) was 0.9 and did not change over development. The low attractor value (h_{down}) was 15 for adults and changed as follows over development: $h_{down3} = 7.5$, $h_{down4} = 8.575$, $h_{down5} = 9.42$.

[b] Recall from Chapter 2 that the noise for the fixation system was white noise applied to c_{const}, rather than the activation of the nodes. This value was multiplied by a random value drawn from a normal distribution with mean zero and standard deviation one (i.e., using the "randn" function in Matlab, set to a random state to begin each simulation), and the result was multiplied by c_{const}.

Table 10

Comparison of Results from Simulations and Experiment 2

		Simulations		Data	
		CR Rate	H Rate	CR Rate	H Rate
3 years	SS1	88.33 (11.79)	91.67 (10.78)	88.10 (18.98)	88.10 (13.76)
	SS2	92.33 (10.76)	73.00 (14.24)	84.52 (17.86)	83.33 (16.01)
	SS3	82.33 (14.83)	57.00 (20.51)	77.22 (26.93)	64.05 (26.67)
4 years	SS1	91.67 (12.26)	97.67 (6.74)	96.43 (9.65)	95.24 (7.81)
	SS2	94.67 (9.18)	89.67 (10.59)	92.86 (12.60)	91.67 (14.25)
	SS3	83.00 (16.32)	71.33 (17.83)	80.95 (22.51)	70.48 (24.94)
	SS4	77.33 (17.74)	61.67 (19.12)	61.11 (19.25)	66.67 (16.67)
5 years	SS1	95.00 (9.07)	97.33 (6.17)	94.05 (8.29)	98.21 (4.82)
	SS2	99.67 (2.36)	93.00 (10.15)	91.67 (10.84)	91.67 (18.20)
	SS3	95.00 (8.42)	79.33 (15.99)	84.52 (16.62)	82.14 (15.28)
	SS4	91.00 (9.65)	60.33 (15.02)	89.29 (15.48)	57.14 (25.91)
	SS5	85.33 (12.88)	45.67 (21.51)	87.50 (16.09)	44.44 (30.43)
Adult	SS1	98.67 (4.57)	90.67 (10.91)	98.81 (4.45)	97.62 (8.91)
	SS2	100.00 (0.00)	94.00 (10.52)	100.00 (0.00)	100.00 (0.00)
	SS3	100.00 (0.00)	94.00 (9.38)	98.81 (6.05)	100.00 (0.00)
	SS4	100.00 (0.00)	91.33 (9.67)	97.62 (6.05)	100.00 (0.00)
	SS5	98.33 (5.05)	77.33 (16.75)	96.43 (7.10)	89.29 (18.03)

Note. CR = correct rejections; H = hits; SS = set size. Means are shown with standard deviations in parentheses.

Table 11

Comparison of Total Looking Time from Simulations and Experiment 1

		Simulations	Data
3 years	SS2	13.77 (0.64)	14.67 (2.35)
	SS4	13.34 (0.54)	14.23 (2.54)
	SS6	14.75 (0.54)	14.67 (2.32)
4 years	SS2	14.04 (0.45)	14.48 (2.22)
	SS4	14.54 (0.50)	14.61 (2.50)
	SS6	15.10 (0.53)	14.23 (2.21)
5 years	SS2	13.92 (0.53)	14.69 (2.40)
	SS4	14.55 (0.49)	14.71 (2.48)
	SS6	14.80 (0.46)	14.63 (2.42)
Adult	SS2	16.06 (0.51)	16.46 (1.24)
	SS4	15.99 (0.41)	16.47 (1.11)
	SS6	16.43 (0.53)	16.77 (1.11)

Note. SS = set size. Means are shown with standard deviations in parentheses.

Table 12

Comparison of Preference Scores from Simulations and Experiment 1

		Simulations	Data
3 years	SS2	0.55 (0.15)	0.59 (0.09)
	SS4	0.52 (0.15)	0.55 (0.13)
	SS6	0.54 (0.17)	0.54 (0.11)
4 years	SS2	0.53 (0.12)	0.55 (0.10)
	SS4	0.53 (0.14)	0.55 (0.13)
	SS6	0.55 (0.14)	0.54 (0.11)
5 years	SS2	0.55 (0.11)	0.57 (0.10)
	SS4	0.55 (0.11)	0.54 (0.11)
	SS6	0.55 (0.13)	0.54 (0.11)
Adult	SS2	0.59 (0.09)	0.60 (0.11)
	SS4	0.55 (0.06)	0.56 (0.09)
	SS6	0.60 (0.09)	0.55 (0.09)

Note. SS = set size. Means are shown with standard deviations in parentheses.

Table 13

Comparison of Switches from Simulations and Experiment 1

		Simulations	Data
3 years	SS2	7.39 (2.44)	7.86 (3.18)
	SS4	6.47 (5.75)	6.35 (2.94)
	SS6	5.75 (1.98)	6.62 (2.78)
4 years	SS2	8.58 (1.94)	9.13 (3.28)
	SS4	7.65 (2.02)	6.98 (2.80)
	SS6	6.33 (1.81)	7.08 (3.03)
5 years	SS2	10.81 (2.32)	9.25 (3.68)
	SS4	8.69 (2.43)	7.94 (3.64)
	SS6	8.40 (2.33)	7.80 (3.73)
Adult	SS2	13.84 (2.93)	14.20 (5.54)
	SS4	14.25 (2.31)	14.00 (4.28)
	SS6	11.79 (3.25)	13.00 (4.80)

Note. SS = set size. Means are shown with standard deviations in parentheses.

CHAPTER 5

GENERAL DISCUSSION

The goal of this project was to use the Dynamic Field Theory to investigate the development of visual working memory capacity. Previously, preferential looking and change detection tasks had revealed inconsistent capacity estimates over development—10-month-old infants showed an adult-like capacity of 3-4 items in preferential looking (Ross-Sheehy et al., 2003), whereas 5-year-old children showed a capacity of only 1.5 items in change detection (Riggs et al., 2006). This discrepancy was previously thought to reflect differences across tasks, and both tasks were explained in the context of Information Processing theory. Nevertheless, this explanation left a number of questions unanswered. How does capacity relate to preferential looking behavior? What causes changes in capacity between 6 and 10 months in preferential looking, and between 5 and 10 years in change detection? How is performance related across these two tasks? To address these questions, I considered both tasks within a different theoretical framework, the Dynamic Field Theory.

In Chapter 2, I discussed how both tasks—change detection and preferential looking—have previously been captured using dynamic field models (Johnson, Spencer, & Schöner, in press; Perone & Spencer, 2008). I then used these models as a starting point to develop a unified model that could capture performance in both tasks. First I showed sample simulations demonstrating the different types of responses in change detection, and fit the general pattern of performance shown in adult studies of change detection (e.g., Vogel et al., 2001). Next, I tested this model in the preferential looking task by adding a fixation system used by Spencer and Perone (2008). Using this theoretical framework, I made three specific predictions for performance in these tasks over development: children and adults would show higher capacity than infants in preferential looking, that is, capacity estimates higher than the commonly accepted

capacity limit in adults of 3-4 items; capacity estimates would be higher in preferential looking than in change detection for the same individuals; and, even though these tasks were expected to produce different capacity estimates, there would be correlations in performance across tasks because both rely on the same underlying working memory system.

These three predictions were verified in a series of experiments with 3- to 5-year-old children and adults. Specifically, all age groups produced capacity estimates of at least 6 items in preferential looking, but in change detection, children produced estimates of only 2-3 items, and adults, about 4 items. Although I found the expected difference in capacity estimates when the same participants were tested in both tasks, I also found the predicted correlations in performance across tasks: capacity in change detection was predicted by the number of switches between displays on the SS4 trials in preferential looking. Recall that the number of switches in the DFT provides an index of how quickly and robustly the model can form a working memory of items in each display. More switches, therefore, should be related to an improved ability to maintain peaks in change detection, which simulations of the unified model showed was related to capacity estimates. These findings lend strong support to the DFT account of performance in these tasks and, more specifically, how performance in these tasks relates to the underlying capacity of VWM.

In Chapter 4, quantitative simulations of the DFT captured a complex pattern of performance across all ages in both tasks. In change detection, the model produced the appropriate rates of each type of response—correct rejections, hits, misses, and false alarms—across set sizes and development. When tested in preferential looking, this same system was able to account for all three measures in the task—total looking, preference scores, and switches—across set sizes at each point in development. These simulations provide insight into the underlying WM representations that would lead to the correlations found in Chapter 3. In particular, the number of peaks that can be stably

maintained over the memory delay in change detection is the primary source of capacity limits in this task. It is this number (between 2 and 4 for children, and between 4 and 6 for adults) that produces the change in switches observed over set sizes in preferential looking.

The simulations in Chapter 4 also confirmed my fourth prediction, that the Spatial Precision Hypothesis could capture the developmental changes in these tasks. This provides added support for this developmental mechanism in the DFT (Schutte & Spencer, 2008; Schutte, Spencer, & Schöner, 2003; Simmering, Schutte, & Spencer, 2008; Simmering & Spencer, 2008; Spencer, Simmering, Schutte, & Schöner, 2007). This hypothesis has successfully captured developmental changes in spatial working memory tasks (see Simmering et al., 2008; Spencer et al., 2007, for reviews), and has also been applied to infant development in preferential looking (Perone & Spencer, 2008; Perone, Spencer, & Schöner, 2007). The simulations presented here show how the SPH can be generalized to account for developmental changes in change detection and preferential looking between 3 and 5 years of age and into adulthood.

In the sections that follow, I outline the implications of the unified model that I have developed and tested here. First, I will discuss how the DFT compares to other theoretical approaches to understanding VWM capacity. Building off of this discussion, I will focus specifically on developmental accounts within the working memory literature. Next, I will discuss various aspects of working memory development (as discussed in Chapter 1) that have not been incorporated in the DFT. Finally, I will consider the broader implications of this project for behavioral development.

Contrasting the Dynamic Field Theory with Other Theoretical Approaches to Capacity

Recall from Chapter 4 that capacity, as measured by these tasks, emerges in the DFT from an interaction between the limited number of WM peaks that can be stably

maintained at one time and the comparison and decision processes in the model. This account validates the construct of capacity in one sense—that there is a limited amount of information (i.e., activation) that can be held in memory at once. In another sense, however, it contradicts one traditional view of capacity—in the DFT, there is no "pure" form of capacity, that is, capacity does not exist separately from the information being maintained. Interestingly, much data has been generated illustrating this latter point (e.g., recall from Chapter 1 that word span tasks depend heavily on the details of the stimuli). However, general acceptance of an Information Processing account of capacity has made it difficult for researchers to abandon the notion that there is a pure form of memory that can be measured in isolation.

In recent years, a debate has emerged within the adult VWM literature on the nature of capacity limits. In particular, Luck and colleagues (e.g., Luck & Vogel, 1997; Zhang & Luck, 2008) argue that capacity limits show properties of fixed-resolutions "slots"; that is, there is a specific number of items that may be "loaded" into working memory, regardless of what these items are. On the other hand, Alvarez and colleagues (e.g., Alvarez & Cavanagh, 2004) propose that capacity depends on a pool of resources that are allocated differentially depending on the details of the stimuli; complex items require more resources, and therefore fewer may be remembered at once, relative to simple items that require fewer resources. Both views have received empirical support, but because neither has been formalized in a process-based model, these results fail to falsify either perspective. For example, a conceptual theory can be modified such that slots may appear resource-like if one allows objects to be divided into smaller components that then become the units of representation. Conversely, resources may serve slot-like functions for certain types of stimuli (e.g., the simple colors and shapes that are often used in change detection tasks). Until these accounts have been formalized and embedded within processes that generate behavior, it is difficult to evaluate these explanations against one another.

The DFT can address this issue directly because it is a formal, process-based model: the nature of representations can be directly observed in simulations by analyzing the characteristics of the peaks in WM. In some ways, WM peaks are like slots—there is a discrete number of peaks that may be formed at one time, and they encode information in an all-or-none fashion (i.e., peaks either form or not, rather than encoding partial information; note that this does not imply that information in WM will be error-free, see Johnson, 2008). However, unlike most characterizations of slots, peaks do not have a fixed resolution (i.e., width). This is due to the continuous metric nature of the WM field—each peak being held in WM may be affected by other peaks being held at the same time. As an example, consider how long it takes a WM peak to form: peaks build more slowly and are weaker if two other peaks are building to either side. Moreover, the strength of a peak is generally higher when it is the only peak being maintained in WM. In this way, peaks resemble a pool of resources—memory for any given item will be affected by the number of other items, along with their complexity and similarity (Johnson, 2008; Johnson, Spencer, Luck, & Schöner, in press). Thus, the representations in the DFT—working memory peaks—provide a third alternative to slots and resources that can capture central aspects of both perspectives.

The DFT also moves beyond previous approaches to capacity in that it grounds working memory in real-time processes that are rooted in neural principles. Some previous theories of capacity have proposed neurally-inspired accounts for the maintenance of items in working memory (e.g., Raffone & Wolters, 2001), but none has yet proposed neural processes for each component of the change detection task—encoding, maintenance, comparison, and decision—that unfold in real time. How strongly tied to neural principles is DFT? Patterns of activation in neural fields have been directly linked to a population dynamics approach to cortical activation. For example, Bastian and colleagues (1998; 2003) compared time-dependent changes in neural activation in a dynamic field model of motor planning to single-unit neural activity in

motor cortex. To make this comparison, these researchers first mapped the responses of neurons in motor cortex to basic stimuli and created a continuous field by ordering the neurons based on their "preferred" stimulus values. This was followed by a behavioral pre-cueing task that probed predictions of a dynamic field theory of movement preparation (Erlhagen & Schöner, 2002). Note that this same theory has also been tested using ERP techniques (McDowell, Jeka, Schöner, & Hatfield, 2002). In both cases, the reported results suggested a robust relationship between predictions of a dynamic field model and neurophysiology.

There is also strong evidence for the basic local excitation / lateral inhibition form of neural interaction used in dynamic field models (Durstewitz, Seamans, & Sejnowski, 2000). Moreover, because cortical neurons never send both excitatory and inhibitory projections, the lateral inhibitory interaction must be mediated through an ensemble of interneurons. To capture this property, the DFT builds from a generic, two-layer formulation proposed by Amari and Arbib (1977) to realize this interaction, where a field of inhibitory interneurons receive input from an excitatory activation field and in turn inhibits that excitatory field.

The DFT incorporates additional insights gained from studies of the layered structure of cortex. In particular, the three-layered architecture of the model was inspired by a cortical circuit model of the neocortex, derived from decades of research investigating the cytoarchitecture of the neocortex (Douglas & Martin, 1998). This basic circuit model consists of two interacting populations of excitatory pyramidal cells distributed across different layers of cortex, coupled to a single population of inhibitory neurons. Although I did not specify a precise relationship between layers in the DFT and particular cortical layers, the basic structure and patterns of connectivity within the model are consistent with established principles of cortical organization.

Because the DFT is a real-time process model that is grounded in neural principles, it also provides a unique opportunity to explore development in this context.

In particular, the Spatial Precision Hypothesis provides a specific developmental mechanism—inspired by neural development—that can be incorporated into the real-time processes operating in the model. In the section that follows, I evaluate the simulations from Chapter 4 and the more general implications of the SPH for our understanding of the link between behavioral and brain development.

The Spatial Precision Hypothesis: A Specific Mechanism for Developmental Change

Previous research on developmental changes in working memory capacity has focused on when changes occur developmentally, mapping out the general patterns over development to see when children reach adult-like levels of performance (see Table 1). What has been missing from this literature thus far is an explanation of how these changes occur (see Siegler, 1996; and Spencer & Plumert, 2007, for similar critiques in other domains). In an attempt to synthesize the developmental findings, some researchers have proposed changes that fall into two general categories: strategic, as in the use of rehearsal or chunking, and maturational, as in improvements in processing or resistance to interference (Dempster, 1981; Pickering, 2001). In the case of strategic factors, no specific proposals have been put forth regarding how or why the use of strategies changes over development. For maturational accounts, myelination has been suggested as a likely candidate for improving the efficiency of memory processing. However, the process by which myelination might produce changes in capacity has not been specified. Thus, a primary goal of formalizing two VWM tasks in a single process-based model was to move beyond these previous proposals to explore whether a specific developmental mechanism might account for changes in performance in working memory tasks over development.

The simulations in Chapter 4 demonstrated that implementing the Spatial Precision Hypothesis in the Dynamic Field Theory effectively captures developmental

changes in both preferential looking and change detection performance. In particular, strengthening neural interactions over development produced developmental changes in capacity and generated the same behavioral patterns children showed in these tasks. Because the unified model built upon a modeling framework and developmental hypothesis that have been applied in other domains, this connects performance in these tasks to a host of other phenomena.

Beginning with infancy, the SPH can account for developmental changes in the Piagetian A-not-B task. In particular, early in development, weak interactions do not support self-sustaining peaks in WM; this leads to perseverative responding in the context of the A-not-B task (see Simmering et al., 2008, for details). Later in development, as interactions strengthen, WM peaks sustain and can survive the short delay in the A-not-B task, consistent with the correct performance of 10- to 12-month-old infants. Beyond infancy, as interactions continue to strengthen, WM peaks become more robust and may sustain over longer delays and in the absence of continued visual input (i.e., without lids marking the hiding locations, see Schutte et al., 2003).

Further strengthening of interactions moves working memory from a single-peak regime (in which only one peak may build and sustain at a time) to a multi-peak regime (in which multiple peaks may co-exist simultaneously). The simulations capturing 3-year-olds' performance in Chapter 4 exhibit an interaction function that has just transitioned beyond the single-peak regime: multiple peaks may be formed at one time, but only under supportive circumstances, and even in these cases the peaks are not particularly robust. For example, compared to simulations with "older" parameter sets, peaks in the 3-year-old simulations were prone to dying out during the delay. Only with further strengthening of neural interactions is the WM field able to reliably maintain multiple peaks throughout the change detection delay.

Through this same period of development, strengthening interactions play a vital role in the interface between perceptual and memory processes. In spatial working

memory tasks, for example, strengthening neural interactions enable children to simultaneously remember a target location in working memory and remain anchored to perceived reference frames in the task space. Between the ages of 3 and 6 years, children show a complex pattern of change in reference frame biases in spatial recall. This pattern has been quantitatively modeled in the DFT using the SPH (Schutte & Spencer, 2008). In addition, these findings have been generalized to predict performance in a second task, position discrimination (Simmering & Spencer, 2008).

The tasks presented here also reflect a richer and more robust interaction between perceptual and working memory processes over development. In particular, the interface between perceptual and working memory processes is at the center of the comparison mechanism in the DFT, that is, the comparison between items held in WM and items present in the display (represented in PF). This comparison mechanism produces the pattern of switching seen in preferential looking, as well as the capacity estimates in change detection. Thus, the comparison mechanism is an important contribution to the relationship across tasks revealed in the regression analysis.

In summary, the SPH has proven to be a powerful mechanism for developmental change within the DFT across multiple tasks—from the Piagetian A-not-B task to change detection—and across a broad range of ages—from infancy into adulthood. Critically, the full range of these developmental changes across tasks can be realized by relatively simple changes in the strength of neural interactions over development. Given this, to what extent might the changes captured by the SPH reflect simple biological changes in the brain during early development?

Various studies have shown that there are changes in multiple brain regions over development that might produce effects similar to the changes captured by the SPH. As an example, consider development of the prefrontal cortex (PFC). Many researchers have suggested that the PFC plays a central role in a variety of working memory tasks (e.g., Diamond & Goldman-Rakic, 1989; Luciana & Nelson, 2002; Nelson, 1995). For

example, one imaging studied demonstrated involvement of the PFC in both maintenance and manipulation of object-related information (i.e., when in a sequence a particular object was seen) in a working memory task (Crone, Wendelken, Donohue, van Leijenhorst, & Bunge, 2006). It is well-documented that the PFC continues to develop throughout childhood and adolescence (e.g., Huttenlocher, 1979, 1990, 1994; Jernigan, Trauner, Hesselink, & Tallal, 1991), suggesting that development of the PFC plays a likely role in developmental changes in working memory capacity.

In addition, reductions in grey matter volume that reflect pruning of neural connections is still occurring in early childhood, and even into the post-adolescent years (Gogtay et al., 2004). Myelination is also still occurring in the frontal lobe into early childhood (Sampaio & Truwit, 2001). It is possible that these neurophysiological changes underlie changes in the spatial precision of neural interactions over development. For example, Edin and colleagues (Edin, Macoveanu, Olesen, Tegnér , & Klingberg, 2007) examined neurophysiological changes related to the development of working memory by implementing changes related to synaptic pruning, synaptic strengthening, and myelination in a neural network model of visuo-spatial working memory. These researchers then used the model to generate five developmental predictions about BOLD signals. They compared predictions the network made to BOLD signals measured with fMRI in 13-year-olds and adults, and found that neural interactions with “higher contrast” over development effectively captured developmental changes in BOLD signals. Higher contrast in their model consisted of strengthening connections both within and between regions which yielded more precise patterns of neural activation. Thus, the most effective developmental hypothesis in their simulation and fMRI study mirrored the changes captured by the implementation of the spatial precision hypothesis in Chapter 4.

An example of a similar proposal is the representation acuity hypothesis proposed by Westermann and Mareschal (2004) to explain the development of visual object processing. According to the representation acuity hypothesis, the transition from

processing object parts to processing objects as wholes is the result of the narrowing of receptive fields in visual cortex. This narrowing of receptive fields is conceptually similar to the increase in the precision of neural interactions in the DFT. Schutte and Spencer (2008) showed that such narrowing can be an emergent result of strengthening excitatory and inhibitory interactions among layers of neurons in cortical fields.

Although these explanations of what is changing at the level of the brain are exciting, this simply shifts the developmental question to another level of description, leaving the question open as to what drives the change in cortex. It is likely that these cortical changes are regulated by 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. It is also likely that these changes are massively experience-dependent (see Johnson, 1999, for review). Given that the changes made to the model in Chapter 4 involved only changes in the strength of excitatory and inhibitory interactions among layers, it is easy to imagine that such changes could arise from a simple Hebbian process that strengthens cortical connections as a function of experience. For example, by incorporating Hebbian learning into the projections within and between layers in the unified model, these connections could be built up through long-term experience. Each time an item is perceived or remembered, the connections supporting those peaks (in PF and WM, respectively, with the corresponding connections to and from Inhib) would strengthen slightly. Given repeated experience with a full range of stimuli over days, weeks, months, and years, this suggests a mechanism by which the tuning of neural fields could emerge through experience. Future research will need to probe whether such a process can indeed give rise to the types of parameter changes implemented in Chapter 4.

In summary, by incorporating the SPH in the DFT, I have proposed a specific developmental mechanism—the strengthening of neural interactions—that is embedded within a real-time process account of performance in both change detection and preferential looking. This makes substantial advances beyond previous accounts of

capacity and development in the VWM literature, as it is the first complete account of how capacity could develop in a neural system. By contrast, no IP account of these tasks has proposed an explanation for how or why capacity is increasing between 6 and 10 months in preferential looking, or between 5 and 10 years in change detection.

Implications for the Development of Verbal and Spatial Working Memory Capacity

Although the experiments and simulations presented here have addressed the processes underlying the capacity of VWM over development, they do not yet speak to other tasks that test verbal or spatial working memory spans. There are a number of ways in which the tasks I have captured here—preferential looking and change detection—differ from more traditional working memory tasks. Recall from Chapter 1 that the tasks used here differed from other working memory tasks in that they do not depend on speed of processing or rehearsal and there is no sequential component. Therefore, to generalize the DFT beyond the tasks used here, each of these differences must be considered.

Within maturational explanations of capacity changes, increases in speed of processing are often proposed as a possible consequence of myelination. Speed of processing may be considered part of rehearsal (i.e., faster processing allows for faster rehearsal), or could be considered a separable component that influences how quickly items are encoded into or retrieved from memory. In Dempster's (1981) review, the one potential developmental mechanism he concluded had the most empirical support was increased speed of item identification. As evidence for this increase in speed over development, he reviewed studies showing that recognition times decrease over development. Within the DFT, this notion of processing speed is already addressed by the SPH. In particular, stronger interactions allow peaks to build more quickly in both PF and WM. For example, using the parameters from the simulations in Chapter 4, I compared the time for a WM peak to pierce threshold at the developmental end points: it took about

225 time steps (~450 ms) to build peaks in WM with the 3-year-old parameters, and about 150 time steps (~300 ms) to build peaks in WM with the adult parameters. Thus, as neural interactions strengthen over development in the DFT, there is an emergent increase in speed of processing.

A second change that underlies children's improved performance in verbal working memory tasks over development is the emergence and subsequent improvement of rehearsal. Although there is no specific rehearsal mechanism in the DFT, it is important to consider how a strategy like rehearsal might be realized in a neural system. One proposal, in the domain of spatial working memory, is that attention serves as a mechanism of rehearsal (Awh, Jonides, & Reuter-Lorenz, 1998; Awh, Smith, & Jonides, 1995). These authors propose that active maintenance of location information in a working memory task is mediated by focal shifts of spatial selective attention to the memorized locations. Behavioral testing of this hypothesis shows that memory performance is impaired by a secondary task only when it requires moving attention away from the location being remembered. In addition, identification of stimuli (i.e., colors or letters) was faster when the item was presented at a location being held in memory. These results led the authors to conclude that attention is "rehearsing" the location by maintaining focus at the remembered location.

Johnson and Spencer (2008) tested an alternative to this hypothesis, specifically, that the recurrence that sustains peaks through the memory delay in the DFT could produce these behavioral effects. In a series of experiments and simulations, they showed that a dynamic neural field model can produce the same pattern reported by Awh and colleagues without a specific attentional or rehearsal process. As such, no separate mechanism for rehearsal is needed to capture these effects. The "rehearsal" processes is essentially captured in the neural dynamics that sustain peaks in cortical fields.

Within the domain of verbal working memory, rehearsal is generally viewed as an active, iterative process by which information is continually articulated or "refreshed"

(Dempster, 1981). An important component of this strategy in verbal working memory is to maintain the sequence in which the stimuli were presented. Although the model presented here contains no mechanism for remembering items in sequence *per se*, it does remember items as they are presented in time. As such, if items were presented in sequence to the model described here, order effects can occur. For example, items presented early in a sequence are more likely to be kicked out of the self-sustaining state by the time the last item is presented. There is evidence for this type of sequential effect in change detection. Johnson and colleagues (2008; see also, Lin, 2005) used a sequential change detection task where three colored squares (at different locations) were presented sequentially to participants during encoding. After a short delay, a single item was presented at test and participants compared this item to the original item presented in that location to generate a *same/different* response. Participants showed a slight recency effect, that is, performance was higher for the item presented last in the encoding sequence. Thus, memory for the items was affected by the order of presentation, even though the sequence of the items was not explicitly tested in this task.

Most verbal working memory tasks, however, require that items be recalled in the order they were presented. Without a specific memory for the sequence in which the items were presented, the unified model reported here would be unable to generate an ordered list in response. The DFT framework has been extended in this direction. In particular, Sandamirskaya and Schöner (2008) developed a system for learning and sequence generation with autonomous robots. For this study, the robot was taught a sequence of colors by presenting colored blocks, one at a time, while fields corresponding to the order in the sequence were boosted. Then, the robot was set in an arena where these colored blocks were randomly distributed, and its “task” was to visit the blocks in sequence. By coupling the “order” fields such that satisfaction of one step in the sequence (i.e., visiting the color associated with that step) activated the next field in the order, the robot was able to reproduce the sequence it had learned. Thus, sequential

learning is possible within the DFT framework. Interestingly, this model suggests that sequencing is a separable process from working memory, that is, the dynamics that govern sequence learning and production may differ from the dynamics of working memory studied here. As such, it is unclear how developmental changes in visual working memory capacity may interact with developmental changes in this sequencing process, but this is certainly an important topic for future study.

Implications for Behavioral Development

The experiments and simulations presented here also have implications for behavioral development more generally. One question that arises from these developmental findings is whether there might be adaptive consequence of a capacity-limited system early in development. This question has been raised in other contexts (e.g., the "less-is-more" hypothesis in word learning; Newport, 1990), and can be informed by other applications of the DFT framework. For example, dynamic field models of word learning in robots (Faubel & Schoner, in press) and young children (Smith & Samuelson, 2005; Smith, Samuelson, & Spencer, 2008) can provide insight into how weaker interactions—and thus lower capacity—might be advantageous in a developing system.

In the robotic implementation, the robot sees an object and hears a label, and must learn this association within 1-3 trials in order to use that information later to retrieve objects when cued with a label. This is accomplished through a two-dimensional association field, with labels along one dimension and objects along the other. To form an association, a single peak builds at the intersections of the two values (i.e., the object and the label); this peak leaves a trace in an associated long-term memory field. Then, when the label is cued later in the task, this value is activated as a "ridge" along the entire object dimension (i.e., with no specific object information). The dynamics of the field allow a peak to build where there is support from long-term memory—where the trace of

the previously-learned association was laid down—and the robot is able to select the correct objects. A key mechanism in this process is *forming a single peak at the intersection of the two values in the association field,* to lay down the corresponding long-term memory trace. If a peak builds in the incorrect location (say, at another object value for the same label), the robot will learn an erroneous association.

What are the implications of this process for development? A balance is needed between forming associations quickly and avoiding spurious correlations. Learning the association between every two items that happen to be paired it time would clearly overwhelm the system. This is avoided in the robot partly by controlling the input (i.e., direct teaching by presenting single objects with single labels). But, of course, children do not learn about objects in such sparse situations. In the model, preventing associations between every label with every object means avoiding having too many peaks at once—that is, having a low capacity.

If the association field is tuned with weak interactions, like those presented here for young children, both the object and label inputs must be strong in order to form robust associations. This would be the case, for instance, in an ostensive naming context, where one object is highlighted visually and a label is presented repeatedly and/or in isolation. In these cases, the model may learn an object-label association quickly. In other cases, however, when multiple objects or labels are presented close together in time, it will take multiple repetitions to learn this association (see Horst & Samuelson, 2008, for evidence that this is true for children). This provides a concrete example of how weaker interactions, and thus lower capacity, may be adaptive early in development (see Spencer et al., 2007, for further discussion).

An additional insight gained from incorporating these tasks into a single model is to connect research in infancy with tasks used later in development. Research by Rose and colleagues has explored how visual attention in infancy may be related to other cognitive measures over development, with specific interest in outcome differences

between pre- and full-term infants. In particular, Rose, Feldman, Jankowski, and Van Rossem (2005) showed that measures of attention, processing speed, and memory (as assessed by looking behavior) in infancy mediated the effects of prematurity on later cognitive abilities at 2 and 3 years (as measured by the Bayley Mental Development Indexes). Based on structural equation modeling, these authors concluded that there was a cascade effect, in which speed of processing influenced memory in infancy, which in turn influenced later cognitive abilities. Rose and Feldman (1997) showed correlations between processing speed in infancy and IQ at 11 years of age. Thus, looking behavior in infancy seems to provide an index of individual characteristics that persist over long periods of development.

Although these findings by Rose and colleagues (Rose & Feldman, 1997; Rose et al., 2005) are compelling, they propose no specific mechanism or process linking looking behavior in infancy to these later cognitive abilities. The DFT, by contrast, provides a specific link between looking dynamics and memory formation—that is, specifying *how* and *why* looking and remembering are linked—in infancy (Perone & Spencer, 2008) as well as later development (as shown in my simulations of Experiment 1 data). Moreover, by combining this process into the unified model I presented here, I have shown how the same processes of memory formation contribute to other behaviors (i.e., change detection). This provides the first detailed, process-based account of how fixation dynamics relate to other behaviors later in development.

These findings have implications for the study of atypical development as well. As shown in Chapter 3, the dynamics of fixation in preferential looking tasks provides an index of working memory processes without requiring the participant to follow directions or interact verbally with an experimenter. Thus, it should be possible to use this task with children and adults with a variety of developmental disorders (including autism and Down syndrome). This could provide a fast, efficient way to assess the development of

working memory abilities early in development when interventions to enhance memory formation might be more effective.

Conclusions

The experiments and simulations presented here provide compelling evidence that change detection and preferential looking both depend on the same underlying VWM system, but that capacity estimates differ across these tasks because of the different ways in which the tasks tap this system. Using a single, processed-based model, the Dynamic Field Theory, I have provided the first account for *how* behaviors in these two tasks arise from the same memory system, and *why* they produce different capacity estimates when the same individuals are tested in both tasks. Moreover, developmental changes in both tasks can be captured using an established developmental mechanism, the Spatial Precision Hypothesis. By strengthening neural interactions over development in the model, I was able to capture multiple details of behavior across both change detection and preferential looking tasks, offering the first specific proposal for how capacity limits could develop in a real-time neural system. Incorporating these tasks into the Dynamic Field Theory with the Spatial Precision Hypothesis also connects the current findings to a larger literature on developmental changes in memory from infancy to 6 years and into adulthood. These findings illustrate the benefits of formalizing concepts such as "capacity" within a specific computational model, as well as the importance of probing detailed hypotheses regarding the mechanisms of developmental change.

APPENDIX

Two additional conditions were needed as controls for Experiments 2 and 3. In particular, Experiment 2 was the first use of an adapted procedure for testing change detection performance in young children. Because the task was adapted to be easier for young children to complete, it is possible that it also inflates performance relative to the standard version of the task. To test this possibility, I included a second group of 5-year-old children who completed the standard task for comparison with the 5-year-olds who participated in the new version in Experiment 2. Note that the children in the control condition in Experiment 2 also participated in Experiment 1 (see Table 5) and were included in the cross task analyses described in Chapter 3.

In addition, performance in Experiments 2 and 3 may have been affected by the fact that all children completed the preferential looking task (Experiment 1) before the change detection task. To ensure that change detection performance was not influenced by having participated in preferential looking first, I included a second group of 5-year-olds who participated in only the Experiment 3 task.

Experiment 2 Control: Color Change Detection—Replicating Riggs et al. (2006)

Because the primary condition in Experiment 2 used a version of the change detection task that was modified for younger children (see Chapter 3), I also collected data from a group of 5-year-old children using the original Riggs et al. (2006) procedure. Comparing these children's performance with the 5-year-olds who completed the modified version of the task provides validation that the two tasks are yielding similar performance. These children also completed the preferential looking task, and were included in the comparisons across tasks in Chapter 3.

Method

Participants. Ten 5-year-olds (*M* age = 5 years, 0.38 months; *SD* = 0.53 months) from Experiment 1 also participated in a replication of the Riggs et al. (2006) standard version of the color change detection task. One additional 5-year-old participated but was excluded from analyses because he chose to end early. All participants had normal or corrected-to-normal vision and no history of color-blindness.

Apparatus. The same computer was used as described in Experiment 2. The square stimuli were also identical, with three exceptions. First, instead of presenting the items within a gray "card" frame on a black background that alternated between left and right positions across trials, the colors were simply presented on a gray background. Second, the location of the squares was determined randomly on each trial. Third, rather than using flashcards to explain the task, the experimenter showed pictures of a SS3 trial on a single sheet of paper through a "window" cut into a piece of black cardstock.

Procedure. The procedure was identical to that describe for 5-year-olds in Experiment 2, except that the task was not described as card-matching. Rather, children were simply told they were to remember the colors and tell the experimenter if one changed between the first display and the second.

Results and Discussion

The same three measures of performance were computed for each set size: accuracy (A'), response criterion bias (β), and capacity; means for each measure are shown in Figure A1 with the data from Experiment 2 for comparison. As this figure shows, results were fairly similar across these conditions, with slightly better performance in the modified version of the task used in Experiment 2. Separate ANOVAs were carried out on each measure of performance, with Condition (card version, standard version) as a between-subjects factor. Because all 5-year-olds completed at least SS4, analyses include performance across these four set sizes. Only the comparison on bias

scores revealed a significant effect of Condition, ($F_{1, 22} = 6.13$, $p < .01$); bias was higher in the standard version, suggesting more of a *change* bias in this version of the task. In each other case, there was no significant effect of Condition (*p*s > 0.2). Note that, although capacity estimates in this task did not differ across conditions (mean maximum estimate of 2.82 in this condition, 2.90 in Experiment 2), both were higher than those reported for 5-year-olds by Riggs et al. (2006). The reasons for this difference are unclear; it is possible that the method for calculating capacity across set sizes differed. For example, averaging across set sizes would result in lower estimates than the maximum method used in Experiment 2.[17]

Experiment 3 Control: Shape Change Detection without Preferential Looking

An additional group of 5-year-olds participated in the shape change detection task without participating in the preferential looking task first. It is possible that capacity estimates were lower in change detection because children were fatigued from having completed the preferential looking task. One result that suggests this is not the case, however, is the higher capacity estimates in this task relative to those reported in the literature (i.e., 1.5 items for 5-year-olds; Riggs et al., 2006). Alternatively, it is possible that completing the preferential looking task first actually improved performance in change detection (although there are no clear indications as to how such an improvement might arise). In either case, comparing children's performance to a condition in which they did not participate in preferential looking should rule out these possibilities.

---

[17] The section on capacity in the Riggs et al. (2006) paper does not describe how they arrived at a single estimate for each age group (e.g., average versus maximum); when contacted for clarification, the first author was unsure what method was used (personal communication, June 2007).

Method

Participants. Fourteen 5-year-olds (*M* age = 5 years, 1.05 months; *SD* = 1.47 months) participated in this condition. All children had normal or corrected-to-normal vision.

Apparatus and Procedure. The apparatus and procedure were identical to that described in Experiment 3, with two exceptions. First, children did not participate in the preferential looking task. Second, blocks were presented in the same order for all children: SS2 – SS1 – SS3 – SS4 – SS5.

Results and Discussion

The same three measures of performance were computed for each set size: accuracy (A'), response criterion bias (β), and capacity; means for each measure are shown in Figure A2 with the data from Experiment 3 for comparison. As this figure shows, performance was nearly identical across conditions. Separate ANOVAs were carried out on each measure of performance, with Task Order (preferential looking first, change detection first) as a between-subjects factor. Note that, because all 5-year-olds completed at least SS4, analyses include performance across these four set sizes. In each case, there were no significant effects of Condition (all ps > 0.5). Additionally, mean maximum capacity estimates were similar across conditions: 2.34 in the control condition, 2.29 in Experiment 3. These results suggest that completing the preferential looking task before the change detection task did not influence performance.

Figure A1. Results from the control condition for Experiment 2: (A) accuracy, (B) response criterion bias, and (C) capacity. Error bars show standard error.

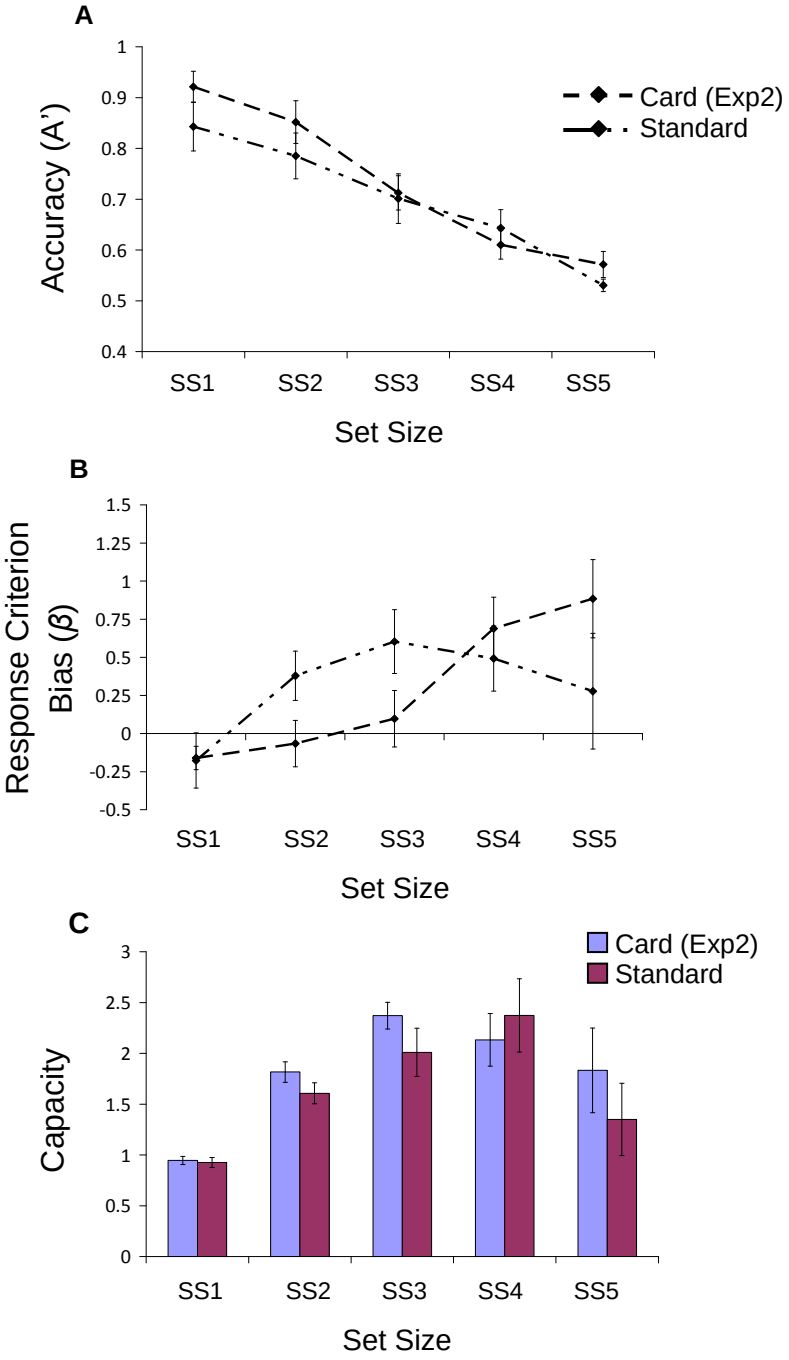
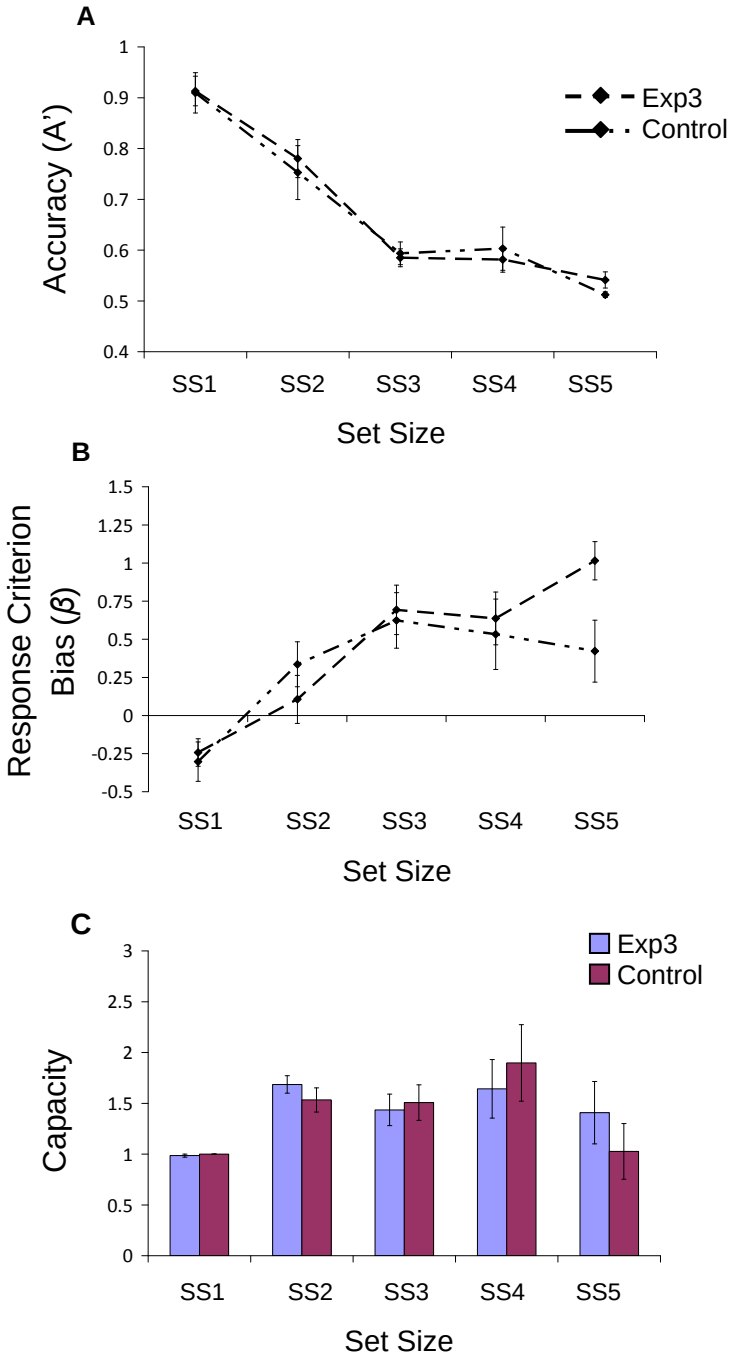


Figure A2. Results from the control condition for Experiment 3: (A) accuracy, (B) response criterion bias, and (C) capacity. Error bars show standard error.



REFERENCES

Aaronson, D., & Watts, B. (1987). Extensions of Grier's computational formulas for A' and B" to below-chance performance. *Psychological Bulletin, 102*, 439-442.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106-111.

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics, 27*, 77-87.

Amari, S., & Arbib, M. A. (1977). Competition and cooperation in neural nets. In J. Metzler (Ed.), *Systems Neuroscience* (pp. 119-165). New York: Academic Press.

Anderson, J. R., Redere, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*, 221-256.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory Volume 2* (pp. 89-195). New York: Academic Press.

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 780-790.

Awh, E., Smith, E. E., & Jonides, J. (1995). Human rehearsal processes and the frontal lobes: PET evidence. In J. H. K. J. Grafman (Ed.), *Structure and functions of the human prefrontal cortex. Annals of the New York Academy of Sciences* (Vol. 769, pp. 97-117). New York, NY, USA: New York Academy of Sciences.

Baddeley, A. D. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition, 10,* 17-23.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? . *Trends in Cognitive Sciences, 4*, 417-423.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *Recent Advances in Learning and Motivation* (Vol. 8). New York: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.

Baillargeon, R., & Graber, M. (1988). Evidence of location memory in 8-month-old infants in nonsearch AB task. *Development Psychology, 24*, 502-511.

Bastian, A., Riehle, A., Erlhagen, W., & Schöner, G. (1998). Prior information preshapes the population representation of movement direction in motor cortex. *NeuroReport, 9*, 315-319.

Bastian, A., Schöner, G., & Riehle, A. (2003). Preshaping and continuous evolution of motor cortical representations during movement preparation. *European Journal of Neuroscience, 18*, 2047-2058.

Bicho, E., Mallet, P., & Schöner, G. (2000). Target representation on an autonomous vehicle with low-level sensors. *The International Journal of Robotics Research, 19*, 424-447.

Broadbent, D. E. (1971). *Decision and stress*. London and New York: Academic Press.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in Long Term Memory*. Bristol: John Wiley & Sons.

Case, R. (1995). Capacity based explanations of working memory growth: A brief history and a reevaluation. In F. M. Weinert & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* Mahwah, NJ: Erlbaum.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). Habit 2000: A new program for testing infant perception and cognition (Version 10). Austin, TX: University of Texas.

Conrad, R. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429-432.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34,* 891B.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

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

Cowan, N., Morey, C. C., & Chen, Z. (in press). The legend of the magical number seven. In S. Della Sala (Ed.), *Tall tales about the brain: Things we think we know about the mind, but ain't so*: Oxford University Press.

Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Life-span development of visual working memory: When is feature binding difficult? *Developmental Psychology, 42*, 1089-1102.

Crone, E. A., Wendelken, C., Donohue, S., van Leijenhorst, L., & Bunge, S. A. (2006). Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Academy of Sciences, 103,* 9315-9320.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Deco, G., & Rolls, E. T. (2005). Sequential memory: A putative neural and synaptic dynamical mechanism. *Journal of Cognitive Neuroscience, 17*, 294-307.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo-spatial memory. *Neuropsychologia, 37*, 1189-1199.

Della Sala, S., Gray, C., Baddeley, A. D., & Wilson, L. (1997). *The visual patterns test: A new test of short-term visual recall.* Feltham, Suffolk: Thames Valley Test Company.

Dempster, F. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*, 63-100.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Reviews of Neuroscience, 18*, 193-222.

Diamond, A. (1990a). Development and neural bases of AB and DR. In A. Diamond (Ed.), *The Development and Neural Bases of Higher Cognitive Functions* (pp. 267-317). New York: National Academy of Sciences.

Diamond, A. (1990b). Developmental time course in human infants and infant monkeys, and the neural bases of inhibitory control in reaching. In A. Diamond (Ed.), *The Development and Neural Bases of Higher Cognitive Functions* (pp. 637-676). New York: National Academy of Sciences.

Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research, 74,* 24-40.

Douglas, R., & Martin, K. (1998). Neocortex. In G. M. Shepherd (Ed.), *The Synaptic Organization of the Brain* (4th ed., pp. 459-509). New York: Oxford University Press.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience Supplement, 3*, 1184-1191.

Edin, F., Macoveanu, J., Olesen, P., Tegnér , J., & Klingberg, T. (2007). Stronger synaptic connectivity as a mechanism behind development of working memory-related brain activity during childhood. *Journal of Cognitive Neuroscience, 19*, 750-760.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71*, 43-51.

Erlhagen, W., Bastian, A., Jancke, D., Riehle, A., & Schöner, G. (1999). The distribution of neuronal population activation (DPA) as a tool to study interaction and integration in cortical representations. *Journal of Neuroscience Methods, 94*, 53-66.

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review, 109*, 545-572.

Fagan, J. F. (1970). Memory in the infant. *Journal of Experimental Child Psychology, 9*, 217-226.

Fagan, J. F. (1973). Infants' delayed recognition memory and forgetting. *Journal of Experimental Child Psychology, 16*, 424-450.

Faubel, C., & Schoner, G. (2008). Learning to recognize objects on the fly: A neurally based dynamic field approach. *Neural Networks, 21,* 562-576.

Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B., & Massey, J. T. (1989). Mental rotation of the neuronal population vector. *Science, 243*, 234-236.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences, 101*, 8174-8179.

Goldberg, J. (2008). *When, not where: A dynamical field theory of infant gaze.* Unpublished doctoral dissertation, Indiana University.

Hitch, G. J., Halliday, S., Schaafstal, A. M., & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory & Cognition, 16*, 120-132.

Hitch, G. J., Woodin, M. E., & Baker, S. (1989). Visual and phonological components of working memory in children. *Memory & Cognition, 17*, 175-185.

Horst, J., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy, 13*, 128-157.

Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex: Developmental changes and effects of aging. *Brain Research, 163,* 195-205.

Huttenlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia, 28,* 517-527.

Huttenlocher, P. R. (1994). Synaptogenesis, synaptic elimination, and neural plasticity in human cerebral cortex. In C.A. Nelson (Ed.), *Threats to optimal development: Integrating biological, psychological, and social risk factors: Minnesota symposium on child psychology.* (Vol. 27, pp. 35-54). Hillsdale, NJ: Erlbaum.

Iossifidis, I., & Schöner, G. (2006). *Dynamical systems approach for the autonomous avoidance of obstacles and joint-limits for a redundant robot arm.* Paper presented at the IEEE International Conference On Intelligent Robots and Systems (IROS), Beijing, China.

Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology, 7*, 377-380.

Jancke, D., Erlhagen, W., Dinse, H. R., Akhavan, A. C., Giese, M., Steinhage, A., et al. (1999). Parametric population representation of retinal location: neuronal interaction dynamics in cat primary visual cortex. *Journal of Neuroscience, 19*, 9016-9028.

Jancke, D., Erlhagen, W., Schöner, G., & Dinse, H. R. (2004). Shorter latencies for motion trajectories than for flashes in population responses of cat primary visual cortex. *Journal of Physiology (Paris), 556*, 971-982.

Jernigan, T. L., Trauner, D. A., Hesselink, J. R., & Tallal, P. A. (1991). Maturation of human cerebrum observed in vivo during adolescence. *Brain, 114,* 2037-2049.

Johnson, M. H. (1999). Ontogenetic constraints on neural and behavioral plasticity: Evidence from imprinting and face processing. *Canadian Journal of Experimental Psychology, 53*, 77-90.

Johnson, J. S. (2008). *A dynamic neural field model of visual working memory and change detection.* Unpublished doctoral dissertation, University of Iowa.

Johnson, J. S., Hollingworth, A., & Luck, S. J. (2008). The role of attention in the maintenance of feature bindings in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 41-55.

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (in press). A dynamic neural field model of visual working memory and change detection. *Psychological Science*.

Johnson, J. S., Spencer, J. P., & Schöner, G. (2008). Moving to a higher ground: the dynamic field theory and the dynamics of visual cognition. *New Ideas in Psychology, 26*, 227-251.

Johnson, J. S., Spencer, J. P., & Schöner, G. (in press). A layered neural architecture for the consolidation, maintenance, and updating of representations in visual working memory. *Brain Research*.

Kopecz, K., & Schöner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural, dynamic fields. *Biological Cybernetics, 73*, 49-60.

Lin, P.-H., Hollingworth, A., & Luck, S.J. (2005). Similarity does not produce interference between visual working memory representations [Abstract]. *Journal of Vision, 5*(8), 616a.

Logie, R. H. (1995). *Visual-spatial working memory*. Hove, UK: Erlbaum.

Logie, R. H., & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology, 9*, 241-257.

Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function in children using the Cambridge Neuropsychological Testing Automated Battery (CANTAB): Performance in 4 to 12 year-olds. *Developmental Neuropsychology, 22,* 595-623.

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

Mareschal, D., & Thomas, M. S. C. (2007). Computational modeling in develpomental psychology. *IEEE Transactions on Evolutionary Computation, 11*, 137-150.

McDowell, K., Jeka, J. J., Schöner, G., & Hatfield, B. D. (2002). Behavioral and electrocortical evidence of an interaction between probability and task metrics in movement preparation. *Experimental Brain Research, 144*, 303-313.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Milner, B. (1971). Interhemispheric differences in the localisation of psychological processes in man. *British Medical Bulletin, 27*, 272-277.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, MA: Cambridge University Press.

Nelson, C. A. (1995). The ontogeny of human memory: A cognitive neuroscience perspective. *Developmental Psychology, 31*, 723-738.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 1911-1928.

Oakes, L. M., Messenger, I. M., Ross-Sheehy, S., & Luck, S. J. (in press). New evidence for rapid development of color-location binding in infants' visual short-term memory. *Visual Cognition*.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Science, 102*, 8776-8780.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologia, 32*, 301-345.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44*, 369-378.

Perone, S., & Spencer, J. P. (2008). Establishing a link between infant looking and memory formation in a dynamic layered neural architecture. *Manuscript in preparation*.

Perone, S., Spencer, J. P., & Schöner, G. (2007). A dynamic field theory of visual recognition in infant looking tasks. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the Twenty-Ninth Annual Cognitive Science Society* (pp. 1391-1396). Nashville, TN: Cognitive Science Society.

Pickering, S. J. (2001). The development of visuo-spatial working memory. *Memory, 9*, 423-432.

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience, 13*, 766-785.

Rensink, R. A. (2000). Visual search for change: a probe into the nature of attentional processing. *Visual Cognition, 7*, 345-376.

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

Rose, S. A., & Feldman, J. F. (1997). Memory and speed: Their role in the relation of infant information processing to later IQ. *Child Development, 68*, 630-641.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Visual short-term memory in the first year of life: Capacity and recency effects. *Developmental Psychology, 37*, 539-549.

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Van Rossem, R. (2005). Pathways from prematurity and infant abilities to later cognition. *Child Development, 76*, 1172-1184.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development, 74*, 1807-1822.

Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of grouping. *Quarterly Journal of Experimental Psychology, 21*, 137-147.

Sampaio, R. C., & Truwit, C. L. (2001). Myelination in the developing human brain. In C. A. Nelson & M. Luciana (Eds.), *Handbook of Developmental Cognitive Neuroscience* (pp. 35-44). Cambridge, MA: MIT Press.

Sandamirskaya, Y., & Schöner, G. (2008). Dynamic field theory of sequential action: a model and its implementation on an embodied agent. *Proceedings of the International Conference on Development and Learning*.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501-518.

Schutte, A. R., & Spencer, J. P. (2007). Planning "discrete" movements using a continuous system: Insights from a dynamic field theory of movement preparation. *Motor Control, 11*, 166-208.

Schutte, A. R., & Spencer, J. P. (2008). Tests of the dynamic field theory and the spatial precision hypothesis: capturing a qualitative developmental transition in spatial working memory. *Manuscript under review*.

Schutte, A. R., Spencer, J. P., & Schöner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. *Child Development, 74*, 1393-1417.

Shore, D. I., Burack, J. A., Miller, D., Joseph, S., & Enns, J. T. (2006). The development of change detection. *Developmental Science, 9*, 490-497.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Simmering, V. R., Johnson, J. S., & Spencer, J. P. (2008). Does change detection underestimate capacity? Insights from a dynamic field theory of visual working memory. *Manuscript in preparation.*

Simmering, V. R., Schutte, A. R., & Spencer, J. P. (2008). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. In S. Becker (Ed.), *Computational Cognitive Neuroscience [special issue] Brain Research, 1202,* 68-86.

Simmering, V. R., & Spencer, J. P. (2008). Generality with specificity: The dynamic field theory generalizes across tasks and time scales. *Developmental Science, 11*, 541-555.

Smith, L. B., & Samuelson, L. K. (2005). From the A-not-B error to early word learning: Building cognitive contents through space, Paper presented at the *Biennial Meeting of the Society for Research in Child Development.* Atlanta, GA.

Smith, L. B., Samuelson, L. K., & Spencer, J. P. (2008). The role of space in binding names to things. *Manuscript in preparation*.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34-50.

Spencer, J. P., & Perone, S. (2008). A dynamic field theory of development of visual working memory in infancy. *Manuscript in preparation*.

Spencer, J. P., Perone, S., & Johnson, J. S. (in press). The Dynamic Field Theory and Embodied Cognitive Dynamics. In *Toward a Unified Theory of Development: Connectionism and Dynamic Systems Theory Re-Considered.* New York, NY: Oxford.

Spencer, J. P., & Plumert, J. M. (2007). What makes thinking about development so hard? In J. M. Plumert & J. P. Spencer (Eds.), *The emerging spatial mind* (pp. 375-386): Oxford University Press.

Spencer, J. P., Simmering, V. R., Perone, S., & Johnson, J. S. (2007). Using dynamic field theory to understand the integration of where and what over development, *67th Biennial Meeting of the Society for Research in Child Development.* Boston, MA.

Spencer, J. P., Simmering, V. R., Schutte, A. R., & Schöner, G. (2007). What does theoretical neuroscience have to offer the study of behavioral development? Insights from a dynamic field theory of spatial cognition. In J. M. Plumert & J. P. Spencer (Eds.), *The emerging spatial mind* (pp. 320-361). New York, NY: Oxford University Press.

Squire, L. R. (1986). Mechanisms of memory. *Science, 232*, 1612-1619.

Steinhage, A., & Schöner, G. (1998). Dynamical systems for the behavioral organization of autonomous robot navigation. In P. S. Schenker & G. T. McKee (Eds.), *Sensor Fusion and Decentralized Control in Robot Systems: Proceedings of SPIE* (Vol. 3523, pp. 160-180): SPIE Publishing.

Thelen, E., & Bates, E. A. (2003). Connectionism and dynamic systems: are they really different? *Developmental Science, 6*, 378-391.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral & Brain Sciences, 24*, 1-86.

Thelen, E. & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *The organization of memory*. New York: Academic Press.

Vicari, S., Bellucci, S., & Carlesimo, G. (2003). Visual and spatial working memory dissociation: evidence from Williams syndrome. *Developmental Medicine & Child Neurology, 45*, 269-273.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 92-114.

Westermann, G., & Mareschal, D. (2004). From parts to wholes: Mechanisms of development in infant visual object processing. *Infancy, 5*, 131-151.

Wheeler, M., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*, 48-64.

Wilimzig, C., Schneider, S., & Schöner, G. (2006). The time course of saccadic decision making: dynamic field theory. Special issue: Neurobiology of decision making. *Neural Networks, 19*, 1059 - 1074.

Wilson, J. T. L., Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology, 5*, 249-255.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233-235.