2008

A dynamic neural field model of visual working memory and change detection

Jeffrey S. Johnson
University of Iowa

A DYNAMIC NEURAL FIELD MODEL OF VISUAL WORKING MEMORY AND CHANGE DETECTION

by

Jeffrey S. Johnson

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

May 2008

Thesis Supervisor: Associate Professor John P. Spencer

ABSTRACT

Many tasks rely on our ability to hold information about a stimulus in mind after it is no longer visible and to compare this information with incoming perceptual information. This ability relies on a short-term form of memory known as visual working memory. Research and theory at the behavioral and neural levels has begun to provide important insights into the basic properties of the neuro-cognitive systems underlying this form of memory. However, to date, no neurally-plausible theory has been proposed that addresses both the storage of information in working memory and the comparison process in a single framework. To address these limitations, I have developed a new model where working memory is realized via peaks of activation in dynamic neural fields, and comparison emerges as a result of interactions among the model's layers.

In a series of simulations, I show how the model can be used to capture each of the components underlying performance in simple visual comparison tasks—from the encoding, consolidation, and maintenance of information in working memory, to comparison and updating in response to changed inputs. Importantly, the proposed model demonstrates how these elementary perceptual and cognitive functions emerge from the coordinated activity of an integrated, dynamic neural system.

The model also makes novel predictions that were tested in a series of behavioral experiments. Specifically, when similar items are stored, shared lateral inhibition produces a sharpening of the peaks of activation associated with each item in memory. In the context of the model, this leads to the prediction that change detection will be enhanced for similar versus dissimilar features. This prediction was confirmed in a series of change detection experiments exploring memory for both color and orientation.

In addition to sharpening, shared lateral inhibition among similar items produces mutual repulsion between nearby peaks. This leads to the prediction that when similar features are held, they will be systematically biased away from each other over delays.

This prediction was confirmed in a cued color recall experiment comparing memory for a "far" color with memory for two "close" colors.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

A DYNAMIC NEURAL FIELD MODEL OF VISUAL WORKING MEMORY AND CHANGE DETECTION

by

Jeffrey S. Johnson

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

May 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

Jeffrey S. Johnson

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

Thesis Committee: _____
John P. Spencer, Thesis Supervisor

_____
Mary E. Campbell

_____
Andrew Hollingworth

_____
Inah Lee

_____
Bob McMurray

To my wife, Joanna, and my children, Drew, Amber, and Emma, for sticking with me through it all

ACKNOWLEDGMENTS

Despite the title page's claims to the contrary, the thesis you now hold in your hands was not the product of a sole author. Although it is true in the narrow sense that I, Jeffrey S. Johnson, did in fact put pen to paper, as it were, that this document should be preserved for posterity, I cannot in fairness lay claim to sole credit for its realization. First of all, I owe a deep debt of gratitude to my two mentors, John P. Spencer and Steven J. Luck, for their continual support, both financial and otherwise, which allowed me to pursue my doctoral studies on a full-time basis.

John Spencer is, simply put, a wonder to behold. His dedication and enthusiasm for our science is truly inspiring. He fearlessly goes where most would fear to tread, and is a source of light and inspiration wherever he travels. He has taught me much about the proper role of theory in experimental psychology, and has served as a constant reminder that the act of discovery in science often involves stepping outside of standard ways of thinking about problems. He has also exhibited a kindness and understanding towards me that has gone well beyond his duties as my faculty advisor. I am truly grateful for the opportunity that I have had to work in his lab and to get to know him and his wonderful family over the last several years.

I have also benefited tremendously through my interactions with Steve Luck. I owe him a great debt of gratitude for his willingness to take me on as a student, and to engage in an active collaboration with John Spencer and I in our efforts to bring the tools of dynamical systems theory to bear on issues in visual cognition. Steve is a brilliant empirical scientist, and I have learned a tremendous amount about how to be a scientist working with him. He has instilled in me a strong drive to pursue perfection in my scientific endeavors, even when doing so means holding off on writing up data that might be good enough to publish, while another, better version of the experiment is run.

In addition to John and Steve, I want to extend a heartfelt thank you to my third mentor, Gregor Schöner, without whom this project would not have been possible. I have learned quit a lot through our interactions over the last several years, and consider myself very lucky to have wandered into your camp. In addition to the part you played in bringing my thesis project to fruition, I want to thank you for your tireless efforts to expand our understanding of the complex processes underling behavior and cognition using the concepts and tools of dynamical systems theory. The breadth and depth of your scholarship and your contributions across multiple fields are a continuing source of inspiration to me.

I would also like to thank my dissertation committee members, Andrew Hollingworth, Ina Lee, John Lee, and Bob McMurray for their support and valuable input on this project. I would not have been able to finish this off in the time that I did without their understanding and flexibility. Thanks Guys! Also, I'd like to thank Mary Campbell for giving up her afternoon to serve as an alternate committee member at my thesis defense. I hope it was time well spent.

I also owe many thanks to my numerous colleagues here in the Department of Psychology. Of particular importance have been my fellow graduate students. Foremost among these have been, from John Spencer's lab, Vanessa Simmering, John Lipinski, and Sammy Perone. Each of you has been a great source of feedback, inspiration, and friendship over the last several years. Our time in the trenches, through grant revisions and model overhauls will not be forgotten. I also want to thank Anne Schutte and Evelina Dineva for their friendship and valuable input at different points in my graduate career, and Sebastian Schneegans, from Gregor's lab, for help with the interactive simulator.

In Steve Luck's lab, I have benefited greatly through my interactions with Joo-Seok Hyun, Wei Wei Zhang, Po Han Lin, and Adam Neise, who have all given me valuable assistance in everything from running experiments, to thinking about topics in

visual cognition. Wei Wei Zhang deserves special mention for walking me through his RAC model fits.

I also want to acknowledge my Mom and Dad and brothers and sisters, who have been extremely supportive throughout this whole process, and never seem to lose interest in my adventures. Your enthusiasm for my life and studies has been a great source of comfort and happiness to me over the years.

Finally, and most importantly, I want to thank my wife, Joanna, and my children, Drew, Amber, and Emma. They have suffered through the harsh Iowa winters for six long years while I worked on my PhD. They have endured, and continue to endure the uncertainty (and low pay!) that goes along with a career in academia. I don't know how I will ever repay the debt that I owe to each of you. Joanna, thank you so much for sticking this out with me. It means the world to me, and will not be forgotten. I love you all and appreciate your sacrifices so much. Hopefully Emma's dream will come true sooner rather than later, and we'll all make it to the promised land together!

ABSTRACT

Many tasks rely on our ability to hold information about a stimulus in mind after it is no longer visible and to compare this information with incoming perceptual information. This ability relies on a short-term form of memory known as visual working memory. Research and theory at the behavioral and neural levels has begun to provide important insights into the basic properties of the neuro-cognitive systems underlying this form of memory. However, to date, no neurally-plausible theory has been proposed that addresses both the storage of information in working memory and the comparison process in a single framework. To address these limitations, I have developed a new model where working memory is realized via peaks of activation in dynamic neural fields, and comparison emerges as a result of interactions among the model's layers.

In a series of simulations, I show how the model can be used to capture each of the components underlying performance in simple visual comparison tasks—from the encoding, consolidation, and maintenance of information in working memory, to comparison and updating in response to changed inputs. Importantly, the proposed model demonstrates how these elementary perceptual and cognitive functions emerge from the coordinated activity of an integrated, dynamic neural system.

The model also makes novel predictions that were tested in a series of behavioral experiments. Specifically, when similar items are stored, shared lateral inhibition produces a sharpening of the peaks of activation associated with each item in memory. In the context of the model, this leads to the prediction that change detection will be enhanced for similar versus dissimilar features. This prediction was confirmed in a series of change detection experiments exploring memory for both color and orientation.

In addition to sharpening, shared lateral inhibition among similar items produces mutual repulsion between nearby peaks. This leads to the prediction that when similar features are held, they will be systematically biased away from each other over delays.

This prediction was confirmed in a cued color recall experiment comparing memory for a "far" color with memory for two "close" colors.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
VISUAL WORKING MEMORY AND CHANGE DETECTION

Human thought and behavior arises within dynamic and often highly complex visual environments. Within such environments, objects are distributed in space and the events within which they are embedded unfold over time. Although the general properties of a complex scene can be obtained in a single fixation (Potter, 1976; Schyns & Oliva, 1994), acquiring detailed visual information from spatially separated regions requires the sequential inspection of different objects through movements of the eyes (see review in Henderson & Hollingworth, 1999). This allows objects of interest to be centered over the fovea, a region of the retina containing over 30,000 densely packed photoreceptors that provides high-acuity information to the visual system. However, the detailed iconic representations formed during each fixation quickly fade as the eyes move on to inspect new objects (Irwin, 1993; Simons & Levin, 1997). As a result, some form of visual memory is needed to maintain continuity in perceptual processing. In addition, the use of visual memory is necessary whenever we need to compare visual percepts created at different points in time.

Imagine, for example, that you are sitting at your desk drinking a cup of morning coffee. At some point, you set the coffee cup down and turn away to retrieve a paper from your briefcase. When you turn back towards the desk and reach for the coffee mug, you notice that the cup is different from the cup you were drinking from previously (e.g., its color has changed). What's the cause of this change? Looking around, you realize that a colleague sitting at a nearby desk has picked up your coffee mug, mistakenly identifying it as her own. To detect simple changes such as this, the properties of the first cup must have been held for a brief period of time in memory and subsequently compared to the second cup.

In this chapter, I will provide a selective review of the literature examining the cognitive and neural systems underlying the short-term form of memory utilized in such tasks. This discussion will focus primarily on behavioral and neural recording studies of change detection, a laboratory task that captures the basic challenges inherent in the office scenario described above. In addition, I will discuss two different levels of theorizing that have been formulated to capture findings in this area. Theories at each level have provided important insights into the basic properties of the neuro-cognitive systems underlying specific aspects of working memory. However, to date, no neurally-plausible theory has been proposed that addresses both the storage of information in working memory and the process of change detection in a single framework. The development of such theories is necessary if we are to effectively bridge the gap between neural data and behavior. My dissertation will focus on the development of a neurally-based process model that begins to address these challenges.

Fundamentals of visual working memory

Performance of simple visual comparison tasks like the one described above have been shown to rely on a short-term (~200 ms to 30 s) form of visual memory that can be differentiated from much shorter-term iconic memory (see, e.g., Averbach & Coriell, 1961; Irwin, 1992; Sperling, 1960), and from much longer-term forms of memory (see discussion in Luck, in press). This variety of memory was known as *primary memory* in the early days of psychology (see, e.g., James, 1890), and later became known as *short-term memory* (STM; see, e.g., Atkinson & Shiffrin, 1968). Beginning in the 1970s (see, e.g., Baddeley & Hitch, 1974; G. A. Miller, 1960), however, usage of the term short-term memory, which emphasizes the storage of information in memory, was largely supplanted by the term *working memory*. This concept subsumes the storage functions attributed to short-term memory, but places greater emphasis on the manipulation of information in service of ongoing cognition and behavior, as opposed to passive maintenance.

A number of models have been proposed that address the architecture and functioning of working memory. For example, Baddeley (Baddeley & Hitch, 1974; Baddeley & Logie, 1999) proposed a multi-component model of working memory which posits a set of specialized storage systems that are controlled by a central executive. Early versions of the model distinguished two different storage systems: a verbal storage system (the *articulatory loop)*, and a visual-spatial storage system (the *visual-spatial sketchpad*). This distinction has been supported by the finding that brain damage can lead to impaired verbal STM and spared visual STM, and vice versa (De Renzi & Nichelli, 1975). Similarly, behavioral studies have demonstrated that holding information in verbal STM does not impair one's ability to remember visual information, and vice versa (Scarborough, 1972; Vogel, Woodman, & Luck, 2001). More recently, it has been proposed that visual objects are stored in yet a third storage system, the *visual cache* (Baddeley & Logie, 1999), which is distinct from the spatial storage system. Although some work has supported this distinction (Logie & Marchetti, 1991; Woodman & Luck, 2004; Woodman, Vogel, & Luck, 2001), other work has suggested that spatial and object information are stored together in working memory (see discussion in, Luck, in press).

Another influential model, proposed by Cowan (Cowan, 1988, 1999), holds that information in working memory is not maintained within dedicated storage buffers, but instead reflects the subset of information that is currently in the focus of attention. According to this *embedded-process* model, working memory arises within a hierarchical system comprised of long-term memory, the subset of long-term memory that is currently activated, and the subset of activated memory that is currently within the focus of attention. Thus, in this view, information is not stored *in* working memory, but neural populations representing this information can be more or less activated depending on the current focus of attention, recent behavioral history, and so forth.

A related view from the domain of cognitive neuroscience (D'Esposito, 2007; Fuster, 2003; Postle, 2006) holds that working memory is not a separate system, with

dedicated storage buffers localized in particular brain regions, but rather is an emergent property arising as a result of functional interactions between the prefrontal cortex (PFC) and posterior brain systems involved in the perception of stimuli in the environment. In this view, the PFC, and perhaps other multi-modal cortical areas such as the posterior parietal cortex and hippocampus, serves as a source of top-down input to unimodal posterior cortical areas representing task-relevant information. Such inputs may serve to enhance task-relevant information, or, conversely, to suppress task-irrelevant information (see review in D'Esposito, 2007).

Whether viewed as a dedicated storage system, or as an emergent property reflecting interactions within and between cortical areas, it is clear that information about objects in the environment can be acquired and remembered over short-term intervals after the stimulus is no longer visible. In the present work, I adopt the term *visual working memory* (VWM) to refer to the cognitive and neural systems supporting this form of memory. One of the primary tasks used to study VWM is the change detection task depicted in Figure 1. A typical trial in this task involves the presentation of a *sample array* containing one or more simple objects (e.g., colored squares), which observers are asked to remember. This is followed by a brief (e.g., 1000-ms) delay interval, and the appearance of a test array that is either identical to the original sample array, or differs from it in the color of a single object. Participants then make a two-alternative forced choice (2-AFC) response, indicating whether the items in the test array are the same as or different than the items that were present in the sample array.

Information processing has been the dominant theoretical framework used to explain performance in this task. From an information processing perspective, this task can be broken down into a series of discrete stages. The first stage involves the perceptual *encoding* of the sample array and the *consolidation* of this information in VWM. The second stage involves the *maintenance* of this information throughout the delay interval.

In a final *comparison* stage, the items present in the test array are compared to the contents of VWM and a same/different decision is generated.

Adoption of this stage-like view has allowed rapid progress in understanding the properties of the VWM system. For example, behavioral studies of change detection have begun to elucidate the properties of the systems involved in the encoding/consolidation of information in VWM. This work has suggested that consolidation and maintenance in VWM may rely on independent processes (Woodman & Vogel, 2005); that VWM representations are established very rapidly, with objects being encoded at a rate of 25-50 ms per item (Gegenfurtner & Sperling, 1993; Shibuya & Bundeson, 1988; Vogel, Woodman, & Luck, 2006); and that consolidation reflects an all-or-none process, rather than a gradual accumulation of information where VWM representations become more and more precise over time (Zhang, 2007; Zhang & Luck, in press).

Similarly, studies of change detection focusing on maintenance have suggested that the capacity of VWM is highly limited to around 3-4 items worth of information (Luck & Vogel, 1997; Phillips, 1974; Vogel et al., 2001). Moreover, the studies of Luck and colleagues have revealed identical capacity limits when the to-be-remembered stimuli consisted of multi-feature objects (e.g., colored, oriented bars of different lengths), rather than single features. Because feature changes could occur along any one of the represented dimensions at test (i.e., the orientation, color, or line length of one of the items could change), participants needed to store all of the features of each object in VWM to successfully perform the task. Thus, it appears that participants can remember a relatively large number of individual features as long as the total number of objects to be remembered does not exceed the 3-4 item capacity of VWM. This finding has been taken as support for the notion that VWM stores integrated object representations, rather than free-floating features (see, e.g., Luck & Vogel; see also, Irwin & Zelinsky, 2002; Johnson, Hollingworth, & Luck, in press).

Finally, with respect to comparison, there is very little known at present. However, combined behavioral and eye-movement recording studies of change detection have demonstrated that the detection of changes at test elicits rapid movements of the eyes to the location of the change (Hyun, Woodman, Vogel, Hollingworth, & Luck, in press).

Clearly, the behavioral analysis of change detection as a series of discrete stages has begun to tell us quite a lot about the separate components involved in this task, and in VWM more generally. In addition, neural recording studies adopting a similar methodological strategy have begun to localize the specific brain areas supporting each stage. Critically, however, the picture emerging from this work suggests a more complex picture than a simple stage-based view can support. For example, functional magnetic resonance imaging (fMRI; see, e.g., Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002; Todd & Marois, 2004; Xu & Chun, 2006) and event-related potential (ERP; Vogel & Machizawa, 2005) studies of change detection have shown that maintenance in VWM engages a number of different brain areas, including specific regions of the inferior temporal, posterior parietal, and frontal cortex, which are known to interact in complex ways. Moreover, although behavioral work has been taken to support a clean division between consolidation and maintenance in VWM (Woodman & Vogel, 2005), the fMRI studies of Pessoa and colleagues suggest the involvement of a broadly overlapping network of brain regions in both phases of the task. Similarly, the detection of changes at test has been shown to engage a widespread network of brain areas implicated in the control of attention, in addition to the specific brain areas involved in perceptual and memory encoding (Pessoa & Ungerleider, 2004; see also, Hyun, et al., in press). This is not surprising, given that the comparison process is fundamentally about the integration of perception and WM in the moment, allowing the determination of sameness or difference.

Compounding the problem, single unit recording studies with non-human primates have begun to uncover the rich spatio-temporal dynamics underlying working memory at the cellular and network levels (see, e.g., Amit & Mongillo, 2003; Fuster & Alexander, 1971; Goldman-Rakic, 1987). This work has suggested that, rather than reflecting an orderly feed-forward progression of stages, even simple tasks probing WM recruit diverse populations of neurons distributed throughout the brain. Distributed neural systems such as these are noisy, densely interconnected, and time-dependent; they pass continuous, graded, and metric information to one another; and they are continuously coupled via both short-range and long-range connections (see, e.g., Constantinidis & Steinmetz, 1996; Rao, Rainer, & Miller, 1997). The distributed nature of such neuro-cognitive processes does not necessarily rule out a stage-based view of cognition, however. For instance, cognition could unfold as a series stages, as proposed by information processing theorists, but with multiple interconnected brain regions participating in different aspects of each stage. However, once this step is taken, one is confronted with the challenge of explaining how this seemingly stage-like behavior arises from the activity of a neural system with the properties described above. Addressing these issues raises deep theoretical challenges that have yet to be confronted.

Theoretical approaches to VWM and change detection

Given the complexity of the issues described above, it is perhaps not surprising that no single theory has been proposed that sheds light on both behavioral and neural findings within research on visual working memory. However, several theories have been developed that capture aspects of the literature reviewed above. For example, at the cognitive/behavioral level of theorizing, a number of models have focused on the maintenance stage, addressing capacity limits and the nature of the representations held in VWM (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997; Wheeler & Treisman, 2002; Zhang & Luck, in press). For example, Zhang and Luck have developed a quantitative approach where capacity limits are proposed to reflect the functioning of a VWM system

that stores a small number of discrete, fixed-resolution slot-like representations. When more than a few items are presented, observers form high-resolution representations for a subset of the objects and have no information about the remaining objects. As a result, when one of the items that failed to make it into VWM is probed at test, observers' responses are based on their best guess, which results in a high number of errors when the number of to-be-remembered items is high.

In contrast to this proposal, Alvarez and Cavanagh (2004) have argued that working memory consists of a pool of resources that can be allocated flexibly depending on the demands of the task. Thus, when presented with more than a few items, observers can store either a small number of high-resolution representations, or a larger number of low-resolution representations. Capacity limits arise because, presumably, relatively high-resolution representations are necessary to support performance in most change detection tasks, and only a small number of such representations can be formed and maintained simultaneously.

Despite lingering disagreements, models such as these have begun to clarify the nature of the representations stored in VWM. However, they have not addressed the question of how such representations are integrated with incoming percepts to arrive at the comparison decisions required in change detection. In addition, although possible links to neurophysiology have been suggested, it has yet to be shown how the properties suggested by these models emerge from the complex dynamical processes underlying neural function. That said, a particular strength of theories at this level is their firm commitment to rigorous behavioral research exploring VWM in human subjects.

At a neurodynamical/biophysical level of theorizing, models have focused on the cellular and network basis of maintenance in working memory using biophysically realistic models of single neurons and cortical circuits (see reviews in, Durstewitz, Seamans, & Sejnowski, 2000; Wang, 2001). Models at this level have emphasized the

complex interactive dynamics evident in neural systems and have maintained a tight interface with neural recording studies of working memory in non-human primates.

For example, Camperi and Wang (1998) have proposed a neurally realistic model of the prefrontal cortex that captures working memory for single spatial locations. Within this model, neurons coding for similar spatial locations are linked together through recurrent excitatory connections, with the strength of excitatory coupling decreasing as a function of the distance between their preferred cues (see discussion of space code models in Chapter 2). Spatial tuning is further shaped by broad lateral inhibition, in keeping with known physiological interactions (Rao, Williams, & Goldman-Rakic, 1999). This type of tuning allows localized peaks activation or "bump attractors", representing particular spatial locations, to be sustained in the absence of continuing external input.

Models in this class provide a neurally plausible basis for maintenance in working memory. However, much of this work has focused on the maintenance of single spatial locations and has not addressed multi-item VWM and the process by which VWM representations are compared to new perceptual inputs. As a result, these models have had little to say about the storage capacity of VWM (for an exception, see Trappenberg, 2003). Additionally, such approaches have made little contact with the human behavioral literature on VWM, and, in particular, with the large literature using the change detection paradigm.

In summary, although the models reviewed here show promise for elucidating the properties of the systems underlying performance in working memory tasks, there exists no single theory that addresses how the encoding and maintenance of multiple items in VWM *and* the process of change detection arises within an integrated, real-time neural system. Given this state of affairs, an important goal is to develop models that can bridge the gap between the behavioral and neural levels of theorizing. Specifically, we need

models that embrace the complex real-time dynamics of neural systems while maintaining a tight interface with behavioral research examining VWM.

Overview of the dissertation

My dissertation focuses on the development of a neural process model of multi-item VWM and change detection that begins to address these challenges. *The first goal of the dissertation is to formalize this new model*. To this end, Chapter 2 will provide a more extensive discussion of neural approaches to working memory, and will introduce the broader framework within which my model is situated. This will be followed by a description of the specific model architecture I've developed, together with a series of exemplary simulations demonstrating how the basic perceptual and cognitive functions supporting VWM and change detection can emerge from a fully integrated, complex dynamic neural system operating in real time.

The second goal of the dissertation is to empirically test novel predictions of the new model. Chapter 2 will introduce the notion that, rather than being stored entirely independently, items in VWM interact in specific ways depending on their metric relations (i.e., how similar they are along the represented dimension). Chapters 3 and 4 will explore the consequences of such interactions on performance in tasks used to probe VWM. Specifically, Chapter 3 will present a series of change detection experiments testing the prediction that interactions among items in VWM will lead to enhanced change detection when metrically-similar items are stored (e.g., similar colors and orientations). Chapter 4 will use a cued color-recall paradigm to test a second, more subtle prediction—that interactions in VWM will lead to mutual repulsion between items when metrically-similar features are stored.

Finally, it should be noted that the work I present here focuses on the active maintenance of memory for metric feature dimensions such as color and orientation over short-term delays, and does not address the spatial aspects of such representations. This should not be taken as a specific claim about the relative independence of spatial and

object subsystems in VWM, but rather as a convenient starting point for my investigations. Although spatial location does play a role in each of the experiments described below, a detailed consideration of the spatial aspects of the object representations I probe will not be addressed in this first pass of theory development. Ultimately, however, I believe it will be necessary to address how object and spatial information come together if we are to understand how vision can be used to support adaptive behavior, a topic I will return to in the General Discussion.

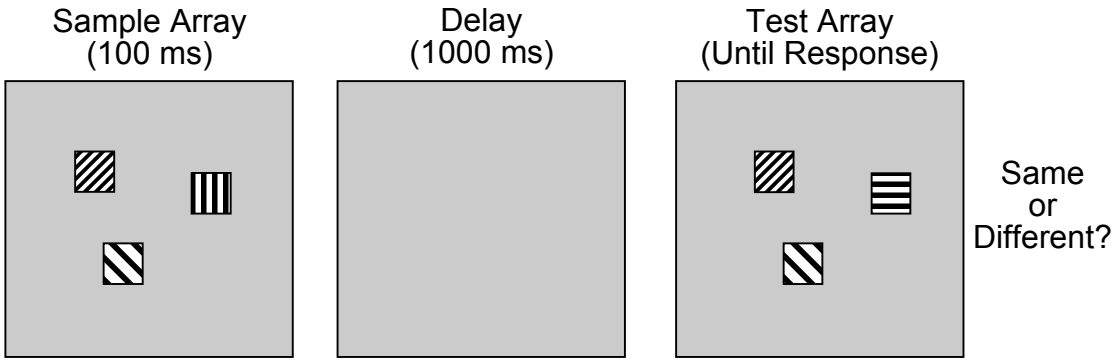

Figure 1. Change detection task used to explore properties of visual working memory for simple features (adapted from Luck & Vogel, 1997). Different fill patterns represent different solid colors.

CHAPTER 2
A DYNAMIC NEURAL FIELD MODEL OF VISUAL WORKING MEMORY AND CHANGE DETECTION

Many everyday tasks rely on our ability to actively maintain information in memory. At the neural level, sustained activation of appropriately tuned populations of cells has been proposed as a mechanism for maintenance in working memory (see, e.g., Amit, 1995; Grossberg, 1978; Wang, 2001). In the present chapter, I will extend the discussion of neural models of working memory that began in Chapter 1. This discussion will focus on three different forms of neural coding that could be used to capture working memory for visual features. Next, I will introduce a theoretical framework, the Dynamic Field Theory, that adopts one of these forms of coding, and which has been used as a framework for thinking about issues in visual-spatial cognition. This will be followed by a detailed consideration of a new dynamic neural field model I have developed that extends this framework to address VWM and change detection, together with a series of simulations that demonstrate the functionality of the model. I will end by considering several novel empirical predictions that can be derived from this model.

Information is represented in the brain by the electrical activity of neurons, and a central question in neuroscience concerns which aspects of a neuron's activity convey information about the environment and our mental states. That is, how is information encoded in the firing patterns of neurons? To date, three different forms of neural coding have been identified and used to capture aspects of memory for metric information. The first applies a *rate coding* principle for representing metric information (Romo, Brody, Hernández, & Lemus, 1999; Salinas, Hernández, Zainos, & Romo, 2000). With this type of coding, information about the metric dimension is carried by the average sustained

firing rate of selectively tuned populations of neurons, with different firing rates representing different values along the dimension.

Several models have been developed that use a rate coding principle to capture performance in the context of specific discrimination paradigms (see, e.g., Machens, Romo, & Brody, 2005; P. Miller & Wang, 2006). For instance, Miller and Wang described a neural model of two-interval vibrotactile frequency discrimination using a rate-coding principle. In this task, the specific vibrational frequency of an initial stimulus (S1) is remembered across a short delay interval and is compared to the frequency of a second stimulus (S2). The animal then makes a binary decision, indicating whether the frequency of S2 is greater than or less than S1. In their model, graded mnemonic activity reflecting the frequency of S1 in working memory provides an inhibitory signal to upstream neurons responding to S2. The inhibitory signal gates later inputs to the model on the basis of the difference in amplitude between S1 and S2. When S2 > S1, upstream neurons overcome the inhibition from working memory and their firing rate increases, whereas when S2 < S1, upstream neurons do not respond. Such differential frequency-dependent responding to S2 provides a plausible means of generating the binary decision required in the frequency discrimination task.

Rate coding models like this can provide a framework for thinking about the integration of perceptual and memory processes in specific discrimination paradigms. In addition, they can be used to capture properties of the sustained patterns of neural activity found in neural recording studies of somatosensory working memory in non-human primates (Romo et al., 1999). However, because the same population of neurons cannot simultaneously maintain different average firing rates, an approach to multi-item memory and change detection using the rate coding principle is not possible.

In addition to encoding information using the specific average firing rate of a population of neurons, a *temporal coding* principle has also been proposed, where the time structure of neural activity is used to represent information. Although the coding of

information in the time structure of neural firing can take many different forms (Shadlen, 2002), the form of temporal coding that has been most closely associated with working memory functions is the *neural synchrony* code. As the name suggests, this form of coding carries information about metric features via synchronized oscillatory reverberation among similarly tuned neurons in separate populations. Models using this form of coding have received some support from studies reporting gamma-band oscillations in the electroencephalogram (EEG) that are correlated with working memory load (see, e.g., Howard et al., 2003; Tallon-Baudry, Bertrand, Peronnet, & Pernier, 1998). Unlike rate coded models, temporal coding models have been proposed as a plausible neural basis for both multi-item memory for features, and as a means of binding together features belonging to the same object in memory.

For instance, Raffone and Wolters (2001) have proposed a neural synchrony model of multi-item VWM where assemblies of neurons modeling inferotemporal cortex (IT) are linked to matching assemblies that model prefrontal cortex (PFC). Individual neurons in IT code for the presence of individual feature values, whereas the assemblies represent perceived objects. When stimulated, the IT-PFC system enters into sustained oscillations, with cells representing different features firing out of phase with each other. This allows multiple feature values to be sustained simultaneously by the network. Additionally, synchronized activity within the IT-PFC assemblies establishes links among features, "binding" these features into objects. Inhibition among these assemblies effectively chunk input into separate objects, which are separately maintained in WM. The model accounts for the limited capacity of VWM in terms of spurious synchronization, that is, the increasing instability of temporal binding as the number of chunks increases.

Neural synchrony models, therefore, represent a promising candidate framework for modeling VWM and change detection. Although some research has supported a role for synchronous neural oscillations in working memory, and in general there has been

much enthusiasm for this coding principle, there is in fact little evidence supporting it, and direct ties to the behavioral literature have remained limited (see, e.g., Shadlen, 2002; Shadlen & Movshon, 1999). In addition, such models have not been applied to the comparison process. Finally, many temporal coding models, including the neural synchrony model described above, rely on reciprocal loops between different brain regions for maintenance (e.g., between the PFC and IT or posterior parietal cortex). Although functional interactions between areas are likely to play a role in working memory, existing evidence suggests that such loops are not necessary to generate sustained activation (E. K. Miller, Erickson, & Desimone, 1996).

A third approach represents metric information through sustained peaks, or "bumps", of activation in neural fields (see discussion of Camperi & Wang, 1998 above; see also, Compte, Brunel, Goldman-Rakic, & Wang, 2000; Trappenberg, 2003). Such approaches propose a *space code* principle for representing metric information. This principle holds that information about the metric dimension is represented by the position of one or more peaks of activation across a population of location- or feature-selective neurons. Support for this proposal has come from single-unit studies showing that memory for a given spatial location is stored by neural populations in the form of a spatially localized sustained firing pattern (see, e.g., Funahashi, Bruce, & Goldman-Rakic, 1989; Fuster & Alexander, 1971). Similar localized patterns of activation representing memory for non-spatial object properties and other task-relevant information have also been found (see, e.g., Andersen, Bracewell, Barash, Gnadt, & Fogassi, 1990; Fuster & Jervey, 1981; E. K. Miller, Li, & Desimone, 1993).

Space code models have been proposed at both the biophysical and the neural population level (see, e.g., Camperi & Wang, 1998; Spencer, Simmering, Schutte, & Schöner, 2007). Such approaches have been rigorously tied to both behavioral and neural data looking at spatial working memory. In addition, space code models have in some cases been adapted to capture VWM for metric feature dimensions (Trappenberg, 2003).

However, as suggested in Chapter 1, most models within this framework have been focused on memory for single spatial locations and, to date, have not been extended to address the comparison process (but see Simmering, Spencer, & Schöner, 2006 for steps in this direction).

To summarize, several different neural codes have been proposed that could be used to capture memory for metric feature information. Of these, only space code and temporal code models are in a position to account for multi-item memory for visual information. In addition to maintaining information about stimuli that are no longer present, however, working memory must be integrated with perceptual representations that are stimulus driven. Some rate coding models have addressed the integration of perceptual and memory processes (Machens et al., 2005; P. Miller & Wang, 2006). However, this has been achieved within the context of specific discrimination paradigms, and concrete limitations of such models make them unsuited to multi-item VWM and change detection. In contrast, Simmering et al. (Simmering et al., 2006) have proposed a space coding model of position discrimination that shows promise in this regard. Although the model has only been applied to single-item position discrimination, space-coding models can be adapted to multi-item memory. For these reasons, in the present work, I adopt a space coding approach for representing metric information in working memory, extending these approaches to capture the processes underlying change detection. It should be noted, however, that the three forms of coding discussed above are not mutually exclusive, and current evidence supports a role for each in different aspects of the neural coding of information.

Dynamic Field Theory as a framework for thinking about the integration of perceptual and memory processes

To begin addressing VWM and change detection within an integrated neural system, I have developed a new model that builds on the Dynamic Field Theory (DFT) of spatial cognition (Simmering, Schutte, & Spencer, 2008; Spencer & Schöner, 2003). This

approach shares the same goal as many biophysical neural network approaches—to think about cognition as the emergent product of complex, contextually-specific neural dynamics—but accomplishes this goal in a way that is more tightly linked to human behavioral work in cognitive psychology that has formed the backbone of the information processing view of VWM.

The DFT grew out of a program of research and theory that has come to be known as the 'motor approach' (Spencer et al., 2006; Spencer & Schöner, 2003). This approach originated in the work of Scott Kelso, Gregor Schöner, and colleagues looking at coordinated motor activity (Kelso, 1995; Kelso, Scholz, & Schöner, 1986; Schöner & Kelso, 1988a, 1988b), and the work of Esther Thelen and her colleagues on infant sensori-motor development (Thelen & Smith, 1994; Thelen & Ulrich, 1991).

According to this approach, human behavior reflects a dynamic balance among *stability*, *instability*, and *flexibility*. In dynamic systems terms, a system is said to be stable when it settles into one of its preferred 'attractor' states, wherein behavioral or neural states can persist in the face of systematic or random perturbations (Braun, 1994). The ability to maintain such stable states (e.g., stably remembering the properties of your coffee cup following a look away) is a key component of adaptive behavior. However, of equal importance is the ability to flexibly change behavior when the circumstances require it (e.g., to update WM in the face of changed information). To achieve this, the current stable state must be de-stabilized so that a new stable state, reflecting, for instance, the changed state of the world, can arise. In dynamic systems terms, a change of a system that leads a particular state to become unstable is referred to as an instability (see Braun, 1994, Chap. 4).

In contrast to an information processing approach, dynamical systems approaches emphasize how behavior and cognitive processes emerge from the coordinated activity of integrated sensorimotor, perceptual, and cognitive systems. Although this emphasis provides a leg up in terms of thinking about integration (e.g., the integration of spatial

and non-spatial object features), the challenge created by this view is the challenge of analysis: If cognition and behavior arise from an integrated system, how can we make sense of the processes that underlie behavior? The approach that has been pursued within this framework, and which is further explored here, is to build cognitive/neural systems that are reentrant, reciprocally coupled, fully integrated dynamical systems, but that have subsystems whose dynamical properties leave behavioral 'signatures' that can be observed in particular situations. In this way, it is possible to engage in rigorous hypothesis testing and to generate novel behavioral predictions—a core strength of the information processing approach—while addressing the integrated nature of cognitive processes.

Basics of dynamic neural activation fields: Elementary perceptual and memory processes in two-layer networks

The number of neurons and synapses within a given patch of cortical tissue is immense, and the patterns of connectivity among them is truly staggering. However, at a mathematical level, it has been shown that the problem can be usefully simplified if we regard cortical neural tissues as continuous neural fields that form and propagate interacting patterns of excitation and inhibition over time (see, e.g., Amari, 1977; Buerle, 1956; Griffith, 1963, 1965; Grossberg, 1980; Wilson & Cowan, 1972). A highly influential formulation of this idea was proposed by Amari (1977; Amari & Arbib, 1977) on the basis of studies examining neural dynamics within the visual cortex of cats. A simple two-layer network of the type analyzed by Amari can be seen in Figure 2. The basic model consists of a single population of visually-selective excitatory neurons reciprocally coupled to a second population of similarly tuned inhibitory neurons. These neurons are arranged by their topographic position in cortex, thereby forming a continuous dynamic neural field. Activation in the field represents, for instance, the retinal position of a stimulus. Thus, the discrete sampling of inputs by individual neurons

that is more typical in neural network approaches is replaced in this formulation by a continuous neural field that represents the metric structure of the represented dimension.

Although the field concept was originally developed to address neural dynamics in topographically organized visual areas (e.g., V1), the same methods have also been used where no clear topographic organization is apparent (e.g., in the motor cortex; see Bastian, Riehle, Erlhagen, & Schöner, 1998; Erlhagen, Bastian, Jancke, Riehle, & Schöner, 1999; Georgopoulos, 1995). In this case, neurons within the field are ordered according to their functional topography—that is, by the stimuli that they prefer, with nearby neurons coding for similar properties (e.g., similar colors), and distant neurons coding for distinct properties (e.g., different colors).

Patterns of activation within such fields can live in different attractor states. For instance, in the absence of input, the activation level across the population of neurons remains at a stable baseline rate, indicating that no relevant features are currently present. However, in the presence of input, reflecting, for instance, the appearance of a particular colored object in the task space, activation increases for those neurons selectively tuned to this feature. If input is sufficiently strong, the stable baseline firing state is destabilized and the field moves into an "on" state, characterized by the formation of a localized peak of activation within the field. The location of the resultant peak along the represented dimension reflects the field's estimation of the metric values present in the task space (e.g., the detection of a particular color), in keeping with the space coding principle. Two or more instances along the dimension are represented by a double or multi-peaked distribution, with the level of activation providing an estimate of the certainty of each informational source. For instance, a high level of certainty about the presence of a particular feature (e.g., blue) is represented by a higher level of activation than a less probable feature (e.g., red).

Although the picture sketched thus far parallels many of the concepts used by feed-forward networks, it is important to emphasize that activation in dynamic neural

fields does not simply mimic the structure of input (though that is one potential limit case of the model's dynamics). Rather, activation in dynamic neural fields can take on a life of its own due to the internal dynamics that govern the evolution of activation through time. The model developed here uses the generic locally excitatory and laterally inhibitory, or "Mexican hat", form of interaction among neurons described by Amari (1977), and commonly found in cortical feature maps (Durstewitz et al., 2000; Rao et al., 1999). With this form of interaction, neurons coding for similar properties (e.g., similar locations, features, objects, etc.) enter into mutually supportive interactions via excitatory synaptic connections (solid curved arrow in Figure 2), and neurons coding for very different properties enter into mutually suppressive interactions (dashed arrows in Figure 2) mediated by a field of inhibitory interneurons.

The simulations in Figure 3 illustrate the three basic attractor states that arise in the 2-layer dynamic neural field model. The first attractor state is illustrated in Figure 3A, which shows the state of the field prior to input. At this point, the field is characterized by a stable *sub-threshold* state of activation, reflecting the absence of information along the metric dimension. In Figure 3B, the field has transitioned to a *self-stabilized* state, wherein one or more above-threshold peaks of activation are formed in response to specific input. Such peaks remain above threshold as long as the input remains on, but quickly transition back to the sub-threshold state when input is removed. Finally, in the simulation shown in Figure 3C, the strength of excitatory recurrence among neurons in the excitatory layer has been increased slightly. In this case, the field enters a *self-sustaining* state where peaks of activation can remain above threshold after the input has been removed, a form of working memory that is central to the work presented here.

Which regime a field is working in depends on several factors, but of most importance is the balance between excitatory and inhibitory interaction. If inhibition is too strong relative to excitation, peaks will transition back to the sub-threshold state once the stimulus has been removed. However, changing the strength of excitation and

inhibition is not the only way to move the network into a given regime. With all other parameters held constant, a given network can be made to function in either the self-stabilized, input-driven regime *or* the self-sustained regime by global modulations of the resting level of neurons in the field. This changes the functional balance between excitation and inhibition without requiring specific changes to the parameters that determine the strength of these factors. Such modulations could play a role in allowing "executive" processes (e.g., processes believed to be implemented by the PFC), which are responsible for maintaining task demands to determine what types of information are maintained in working memory for use in ongoing tasks (see, e.g., D'Esposito, 2007; Fuster, 2003). For instance, in a task where color is the relevant behavioral dimension, the resting level of neurons in color-selective regions of cortex could be boosted slightly, allowing them to enter the self-sustained state following stimulus presentation (Buss & Spencer, 2008). Conversely, the resting level of neurons coding for other task-irrelevant dimensions (e.g., orientation) could be reduced through global inhibitory input, diminishing their role in the control of ongoing behavior and mental processes.

Integrating perceptual and memory processes in a dynamic neural field model with three layers

The simple two-layer networks introduced in the previous section can serve either a perceptual or a working memory function. Thus, taken separately they capture some of the requirements of a model of VWM and change detection. For instance, a two-layer network operating in the self-stabilized regime, where peaks of activation are formed in response to input but die out when input is removed, could be used to capture the perceptual encoding of the test array. Additionally, a two-layer network operating in the self-sustaining regime, where peaks of activation remain above threshold after input has been removed, can be used to implement maintenance in VWM. As discussed previously, however, modeling working memory and change detection requires that each of these functions be integrated in the same model operating in real time.

To address this, I have developed a three-layer architecture consisting of two layers of excitatory neurons coupled to a single layer of inhibitory neurons. This framework allows simple perceptual and working memory functions, including the detection of stimuli in the task space and their retention in working memory, to be implemented in the same model. In addition, excitatory and inhibitory interactions among the layers gives rise to emergent decisions about detected change. The next section describes the architecture and patterns of connectivity of the model, and presents a series of simulations illustrating the functionality alluded to above. These simulations demonstrate that the model can serve as an integrated neural framework that captures the encoding, consolidation, and maintenance of information in WM, in addition to the coordination of WM and perception leading to updating and change detection in the face of changing sensory inputs.

Model architecture

To combine basic perceptual and working memory functions in a single architecture, I have developed the three-layer dynamic neural field model depicted in Figure 4. The model consists of an excitatory perceptual field (PF(u); Figure 4A), an excitatory working memory field (WM(w); Figure 4C), and a shared inhibitory field (Inhib(v); Figure 4B). 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. In addition, nearby neurons within both PF and WM interact via local excitatory connections. Finally, to capture performance in change detection tasks, I have added a response layer containing two nodes (Figure 4D). One neuron receives summed excitatory activation from PF, and is responsible for making "different" responses, and the other receives summed excitatory activation from WM and is responsible for "same" responses (see discussion below). In addition, these nodes are equipped with self-excitatory connections and are mutually inhibitory, implementing a type of "winner-take-all" decision. Thus, the

response nodes compete for response output in the context of change detection. Note that, for simplicity, time is not shown in this figure, although patterns of activation evolve continuously over time.

Model equations

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

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

where $\dot{u}(x,t)$ is the rate of change of the activation level for each neuron across the feature 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 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 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 the Inhibitory layer (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 was specified by:

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

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

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

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.

Inputs to the model take the form of a Gaussian:

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

centered at x_{center}, with width, σ, and strength, c. These inputs could be turned on and off through time (e.g., the target appears and then disappears). This time interval was specified by the function χ(t). This is referred to as the "index function", because it is set to one in a given interval, when the stimulus is on, and zero elsewhere.

Finally, activation within the field is influenced by the addition of a stochastic component consisting of spatially correlated *noise* with a particular strength and width.

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

$$\tau\dot{v}(x,t) = -v(x,t) + h_v + \int c_{vu}(x - x')\Lambda_{vu}(u(x',t))dx'$$

$$+ \int c_{vw}(x - x')\Lambda_{vw}(w(x',t))dx' + noise\,. \qquad 2.5$$

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; $v(x,t)$ captures the current activation of the field; and h_v sets the resting level of neurons in the field. Note that Inhib receives activation from two projections—one from PF, $\int c_{vu}(x - x')\Lambda_{vu}(u(x',t))dx'$, and one from WM, $\int c_{vw}(x - x')\Lambda_{vw}(w(x',t))dx'$. 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).

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

$$\tau \dot{w}(x,t) = -w(x,t) + h_w + \int c_{ww}(x-x')\Lambda_{ww}(w(x',t))dx'$$

$$- \int c_{wv}(x-x')\Lambda_{wv}(v(x',t))dx' + \int c_{wu}(x-x')\Lambda_{wu}(u(x',t))dx'$$

$$+ c_s s(x,t) + noise \qquad 2.6$$

Again, $\dot{w}\ (x,t)$ is 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; *w(x,t)* captures the current activation of the field; and h_w sets the resting level. WM receives self excitation, $\int c_{ww}(x-x')\Lambda_{ww}(w(x',t))dx'$, lateral inhibition from Inhib, $\int c_{wv}(x-x')\Lambda_{wv}(v(x',t))dx'$, and input from PF, $\int c_{wu}(x-x')\Lambda_{wu}(u(x',t))dx'$. WM also receives direct target input, *s(x,t)*, scaled by c_s.

Note that the PF and WM layers are governed by identical equations. The primary difference between them is in the strength of direct target input, which is stronger to PF, and the strength of self-excitatory interactions, which is stronger in WM (i.e. $c_{ww} > c_{uu}$, see Table 1). Additionally, WM receives excitatory input from PF, whereas WM only contributes inhibition to PF via the inhibitory layer. These differences make it easier for peaks to be sustained in the absence of input in WM versus PF. Additional functional consequences of these differences are considered in the following sections.

Finally, each of the nodes in the response layer are governed by the following equation (example shows 'same' node equation):

$$\tau \dot{r}_s(t) = -r_s + h_r + c_{rr}\Lambda_s(r_s) - c_r\Lambda_d(r_d)$$

$$+ \int c_{su}\Lambda_{su}(u(x',t))dx' + noise \qquad 2.7$$

The rate of change of each node's activation, r_s (where the constant τ determines the time scale and subscript *s* denotes the 'same' node), is determined by the current activation

level, $-r_s$, and the resting level of the node activation, h_r. Each node has a self-excitatory connection, $c_{rr}\Lambda_s(r_s)$, and receives inhibition from the other node, $c_r\Lambda_d(r_d)$. Additionally, the 'same' node receives summed excitatory input from WM(w), $\int c_{su}\Lambda_{su}(u(x',t))dx'$, and the 'different node receives summed excitatory input from PF(u). Each node also has noise added.

Model simulations

In this section, I will present a series of simulations demonstrating the properties of the three-layer model. For simplicity, I will begin by showing how the model behaves when a single input is applied, moving from perceptual encoding, to working memory consolidation and maintenance, to comparison and updating. After this, I will move on to the multi-item case, showing how the model can capture key results from the literature thought to emerge across separate stages of information processing. Critically, I will show how these "stages" can reflect the time-dependent activity of a single, integrated dynamic neural system.

Single-item perception and working memory

Perceptual encoding. The resting level of neurons in each field, $h < 0$, ensures that interaction among neurons only plays a role in the presence of sufficiently strong input. For localized inputs, $S(x)$, with low amplitude, PF, and to a lesser extent WM, begins to assume the form of the input, but activation values across all field sites remain negative (see Figure 5A). That is, PF and WM remain in the stable sub-threshold activation state. When the amplitude of the input is sufficiently strong (see Figure 5B), however, interactions among neurons come into play. At this point, PF transitions to the self-stabilized state consisting of a localized peak of activation centered at the location of

the input. The size and shape of the peak reflects both the input state, as well as locally excitatory and laterally inhibitory interactions among neurons (via recurrence between PF and Inhib). The transition from the sub-threshold state to the self-stabilized state as a function of input strength involves a dynamical instability and is the mechanism by which detection decisions, or perceptual encoding, emerge in the model. Thus, for the model to detect and stably encode a stimulus, input to PF needs to be sufficiently strong and of sufficient duration to generate a self-stabilized peak of activation.

Consolidation and maintenance in working memory. In addition to activating neurons in Inhib, above-threshold activation in PF is propagated to WM. With sufficiently strong input, a single peak of activation begins to build in WM (see Figure 5B). As before, this peak is stabilized by locally excitatory interactions in WM together with broad lateral inhibition from Inhib. Figure 5C shows the consequences of removing the input from PF. This event leads to the destabilization of the peak in PF. In contrast, a self-sustaining peak of activation remains in the WM field. This occurs as a result of the large reduction of excitatory input to PF when the input is removed, and a consequent reduction of inhibitory input to WM (note lower levels of activation in middle layer of Figure 5C). The strength of recurrent excitatory connections among neurons in WM is somewhat stronger than in PF, giving WM a competitive advantage once the input is removed. Although this contributes to both the maintenance of the peak in WM as well as the destabilization of the peak in PF via shared inhibition, sustained activation arises even in a variant of the model with identical patterns of excitatory interaction. This shows that working memory emerges in the WM layer as a result of the flow of incoming activation directly to PF and indirectly to WM.

Emergent change detection and updating in WM. In the absence of continuing perceptual input, the only input to PF comes from Inhib, whose activation is driven by the presence of a self-sustaining peak in WM. As a result, a region of inhibition is formed in PF at the location of the sustained peak in WM (see Figure 5D, top layer). This leads to a failure to re-ignite a peak in PF when a second input that matches the value being maintained in WM is presented to the model (see Figure 6A). Therefore, the model remains in its previous state, with a single stable peak of activation sustained in WM, and negative activation in PF. The presence of a peak in WM, but no peak in PF provides strong input from WM to the response layer, generating a "same" response in the context of change detection (response layer not shown here). Note that although PF fails to re-ignite a peak, this layer shows sub-threshold deformations that reflect the input presented at test. Thus, one might say that the system perceived the stimulus but the on-going maintenance of an identical stimulus in WM prevented the formation of a self-stabilized peak at test.

Compare this with a trial where a metrically very different input is applied at test. In this case, the input enters PF at a relatively uninhibited site, and an above-threshold peak is built in PF (see Figure 6B). At this point, both WM and PF send strong activation to the response layer, where competition to control response output ensues. However, because the peak in PF is strongly supported by perceptual input, it has a competitive advantage over WM, which allows a "different" response to be generated.

Thus, visual comparison in the model amounts to a relatively low-level process where the amount of activation in PF versus WM drives a same or different decision at test. That is, there is no item-by-item comparison, where each target in the test array is

sequentially compared to the contents of working memory. At test, input comes into the model all at once, and comparison occurs globally and in parallel. If an above-threshold peak is established in PF at test, signaling that a novel input is present in the test array, the 'different' node receives strong activation, making it more likely for a different response to be generated. Otherwise, the same node tends to win out in the response layer. Thus, 'same' responses are generated by activation that is already present in WM, whereas 'different' responses must wait for activation to build in PF and overtake activation of the 'same' node. As a result, 'same' responses tend to be generated faster than 'different' responses, in keeping with previous findings (see review in Farell, 1985; Hyun, 2006). Thus, the model has clear potential for addressing RT effects in visual memory tasks, although this topic is not explored further here.

In addition to signaling that a change has occurred, the presence of an above threshold peak in PF at test leads to the updating of WM. When the new peak is metrically relatively close to the old peak, interference can arise. In some cases this can result in the deletion of the old peak and its replacement by a new peak at the value specified by input from PF (see WM field in Figure 6C). However, as shown in Figure 6D, when peaks are metrically far apart, a multi-peak solution can arise where the new peak is added to WM without destabilizing the old peak. This would reflect, for instance, the sequential loading into WM of items with different featural information obtained across successive eye movements or successive presentations of stimuli.

Multi-item perception and working memory

The goal of the present section is to show that the basic functionality of the model illustrated for single items in the previous section can be extended to the multi-item case.

Recall that rate coded models have been used to address discrimination, but are conceptually incapable of multi-item memory. Reversely, temporal coding models have addressed multi-item memory, but the mechanism behind comparison is not clear. Thus, it is important to demonstrate that the neural field model proposed here goes beyond such models by addressing multi-item VWM and change detection.

Multi-peak encoding, consolidation, and maintenance. The presence of multiple high-amplitude inputs results in the formation of peaks of activation in PF, representing the detection and stable encoding of multiple features in the task space. In general, because PF operates in the self-stabilized mode, where locally excitatory recurrence is somewhat weaker than in WM, several peaks can be simultaneously formed as long as the input remains on. This is demonstrated in Figure 7A. In Figure 7B these three inputs have been successfully consolidated and are self-sustained in the WM layer even though the input has been removed. In this case, the "Mexican hat" interaction profile, where inhibition declines as a function of the distance from the focus of excitation, allows the locally excitatory interactions associated with each peak to be isolated by lateral inhibition, while keeping the total amount of inhibition in the field relatively low. This makes it possible for multiple items to be maintained simultaneously in VWM (see also, Trappenberg, 2003).

Multi-peak change detection. The simulations shown in Figure 7A-D demonstrate robust change detection performance when three items are held in working memory. As with the single-item case, when a new input that matches one of the original inputs is presented (Figure 7C), a new peak fails to be re-ignited in PF, and the model generates a "same" response. Similarly, when a new input that does not match one of the original

items is presented (Figure 7D), a robust peak is built in PF, which serves as the basis for the generation of a "different" response. Note that the model behaves identically when three items, rather than one, are presented at test. When all three items match the original items, no new peak builds in PF. Conversely, when one of the items is different, a peak builds in PF at a location matching the changed input.

Updating in multi-peak WM. Figures 8A-B show two different forms of updating in multi-item VWM. In the first case, the presentation of a new input that is distinct but metrically relatively close to the original value produces interference in WM. This leads to the deletion of the original peak in WM and its replacement by a peak at the new value, that is, the value specified by the peak in PF (compare WM peaks in Figures 7D and 8A).

Figure 8B shows a case where new inputs are presented that are not identical to the original inputs, but are too similar to support explicit change detection. Note that the scale of the simulations has been changed so that this rather subtle effect is visible in the figure. Also, for these simulations, the inputs presented to the model are illustrated in the top panel, rather than in the PF layer, with dashed lines representing the first input (S1), and solid lines representing the second input (S2), which was presented following a 1000-ms delay interval. The bottom two panels show the PF and WM layers of the three-layer model—Inhib has been left out for simplicity. The dashed lines in PF and WM show the patterns of activation in these layers after the offset of the first input. As before, two self-sustained peaks of activation are present in WM following stimulus offset, and neurons representing similar features are inhibited in PF. When the second input is applied, peaks try to build in PF, but are unable to do so because of strong inhibition at those locations.

However, because WM receives weak direct input when S2 is on, the peaks in WM begin to shift in the direction of the new inputs. That is, WM is updated in a continuous fashion to reflect the values of the new stimulus, even though the change was not large enough to provoke the generation of a "different" response (see Hollingworth & Henderson, 2004).

Capacity limits in VWM. Thus far, I have demonstrated the functionality of the model when the number of items to be remembered (i.e., the number of peaks) is relatively small. What happens when the number of items is increased beyond three? Figures 9A-C show the consequences of adding more and more items to WM. Although the Mexican hat function underlying maintenance in the model allows a multi-peak solution of the field dynamics, the capacity of working memory is not unlimited. As more items are added to working memory, the net excitatory activation increases, which in turn increases the overall amount of inhibition. Figures 9A-B show that the model can stably maintain up to four items in WM. However, when two more inputs are added, the level of inhibition overtakes excitation, and several peaks drop out of WM, leaving three out of the six items in WM. Under these circumstances, if a new input were to be presented that matches one of the "forgotten" items, the model would incorrectly generate a "different" response by building a peak at that location in PF.

Selection of inputs for consolidation in working memory. Although the model shows robust change detection performance at smaller set sizes, self-sustained peaks can fail to be established in WM in some cases even though the number of items to be remembered is within the capacity of WM. Thus far, we have assumed that all inputs were more or less identical in strength. This is not entirely unreasonable. For example, normalization of input strength could be mediated through other layers of perceptual

processing not implemented here. It is likely, however, that the strength of perceptual inputs arising from different objects in multi-item arrays could vary considerably. Such variations might arise through spontaneous fluctuations, or due to factors such as the relative salience of the objects (e.g., their relative luminance), or the spatial distribution of attention within the display. In this case, less salient objects, or objects that are unattended would produce lower amplitude inputs to PF. This would decrease their likelihood of going through the detection instability and their probability of being consolidated in WM.

The simulations presented in Figure 10 show the consequences of presenting variable amplitude inputs to the model. For these simulations, the model was presented with one strong input, and two relatively weak inputs (Figure 10A). In this case, the strongest of the three inputs is successfully consolidated and sustained in WM after the input is turned off (Figure 10B), whereas the two weaker inputs remain below threshold. However, when the model is probed a short time later (Figure 10C), a second peak has formed at the location of one of the weaker inputs. Thus, with variable amplitude inputs, the model shows an emergent property of sequential consolidation in WM. In addition, these simulations provide a plausible basis for the finding that attended inputs are more likely to be consolidated in WM (Schmidt, Vogel, Woodman, & Luck, 2002).

Discussion

The model presented here represents the first neurally-plausible framework that addresses both the maintenance of multiple items in VWM *and* the process of change detection. This work demonstrates that the components involved in change detection can be realized within a fully integrated neural system operating in real time. Specifically, I

presented a series of simulations demonstrating each of the basic components of change detection, from initial encoding and consolidation, to maintenance, to comparison and updating. The proposed model, therefore, can serve as a useful framework for modeling performance in tasks used to probed VWM.

In addition to capturing existing data, however, a central challenge for any model is to make novel behavioral predictions. How does the model perform in this respect? In the context of the three forms of neural coding discussed above, space code models have the interesting property of metrics. Recall that neurons within each field are topographically ordered according to the stimuli that they prefer. As a result, when multiple items are held in WM, they interact in specific ways depending on their metric similarity. In the context of my model of VWM and change detection, this raises the possibility that the relative similarity of the items being held in memory could have specific measurable consequences on performance in tasks used to probe VWM. In the following chapters, I describe and test two novel predictions arising as a result of close metrics in VWM. Specifically, in Chapter 3 I use the change detection paradigm to test the prediction that metric-dependent sharpening of peaks in WM will lead to enhanced change detection. In Chapter 4, I use a cued color recall paradigm to test a second prediction: that metrically-similar items will be repelled away from each other across delays.

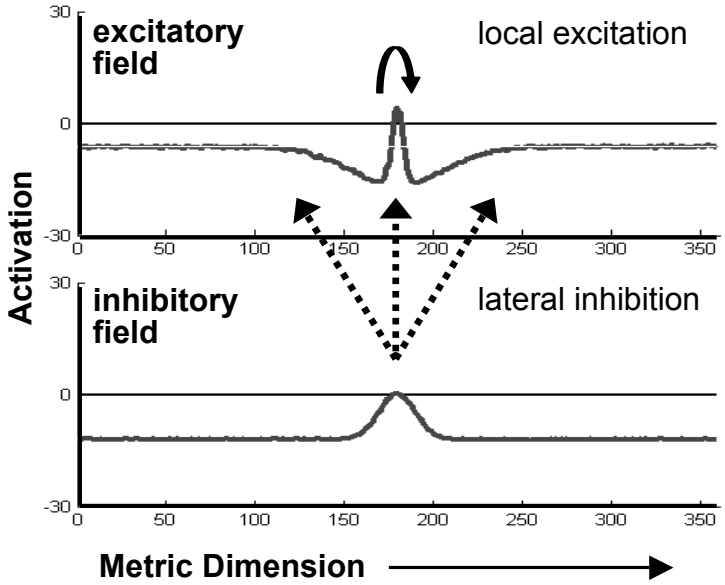

Figure 2. Two-layer neural field model of the type analyzed by Amari (1977). The model consists of a single layer of feature-selective excitatory neurons reciprocally coupled to a similarly-tuned layer of inhibitory interneurons. Neurons coding for similar values along the metric dimension in the excitatory 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). See text for additional details.

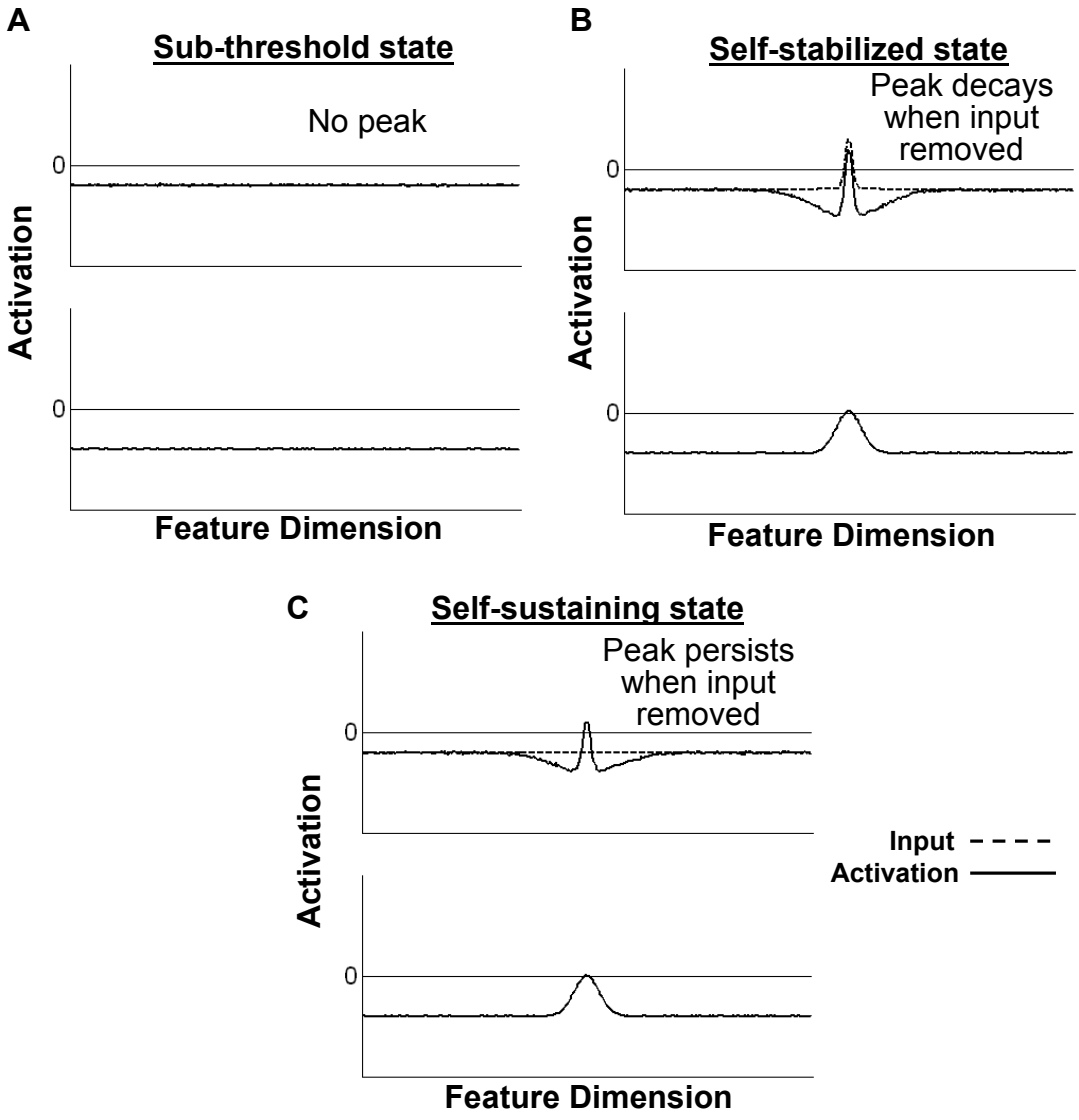

Figure 3. Three basic attractor states arising in the two-layer dynamic neural field model. (A) A stable *sub-threshold state*, where activation at all field sites is negative, reflecting the absence of information along the represented dimension. (B) A *self-stabilized state*, where localized peaks of activation, representing the detection of a particular feature (e.g., a certain color) in the task space, are formed in the presence of input but decay once input is removed. (C) A *self-sustaining state* where localized peaks of activation are maintained in the absence of input, implementing a form of working memory. In each panel, input to the model is represented by a dashed line, whereas the pattern of activation within the field is indicated by a solid line.

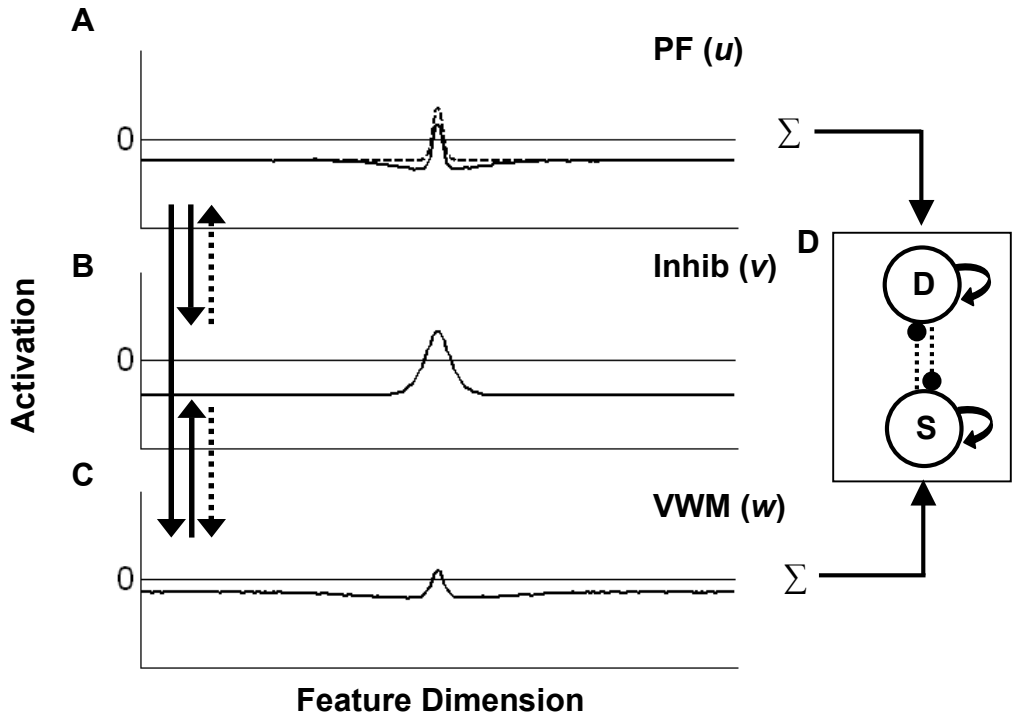

Figure 4. The three-layer dynamic neural field model of visual working memory and change detection. The model consists of an excitatory perceptual field, PF(*u*), (A), and an excitatory working memory field, WM(*w*), (C), which are reciprocally coupled to a single layer of inhibitory neurons, Inhib(*v*), (B). The model also includes a response layer (D) consisting of two nodes that receive summed excitatory activation from WM and PF and are responsible for generating 'same' and 'different' responses. Excitatory and inhibitory patterns of connectivity among the layers are indicated by solid and dashed arrows, respectively. In addition, neurons in both PF and WM engage in localized excitatory interactions.

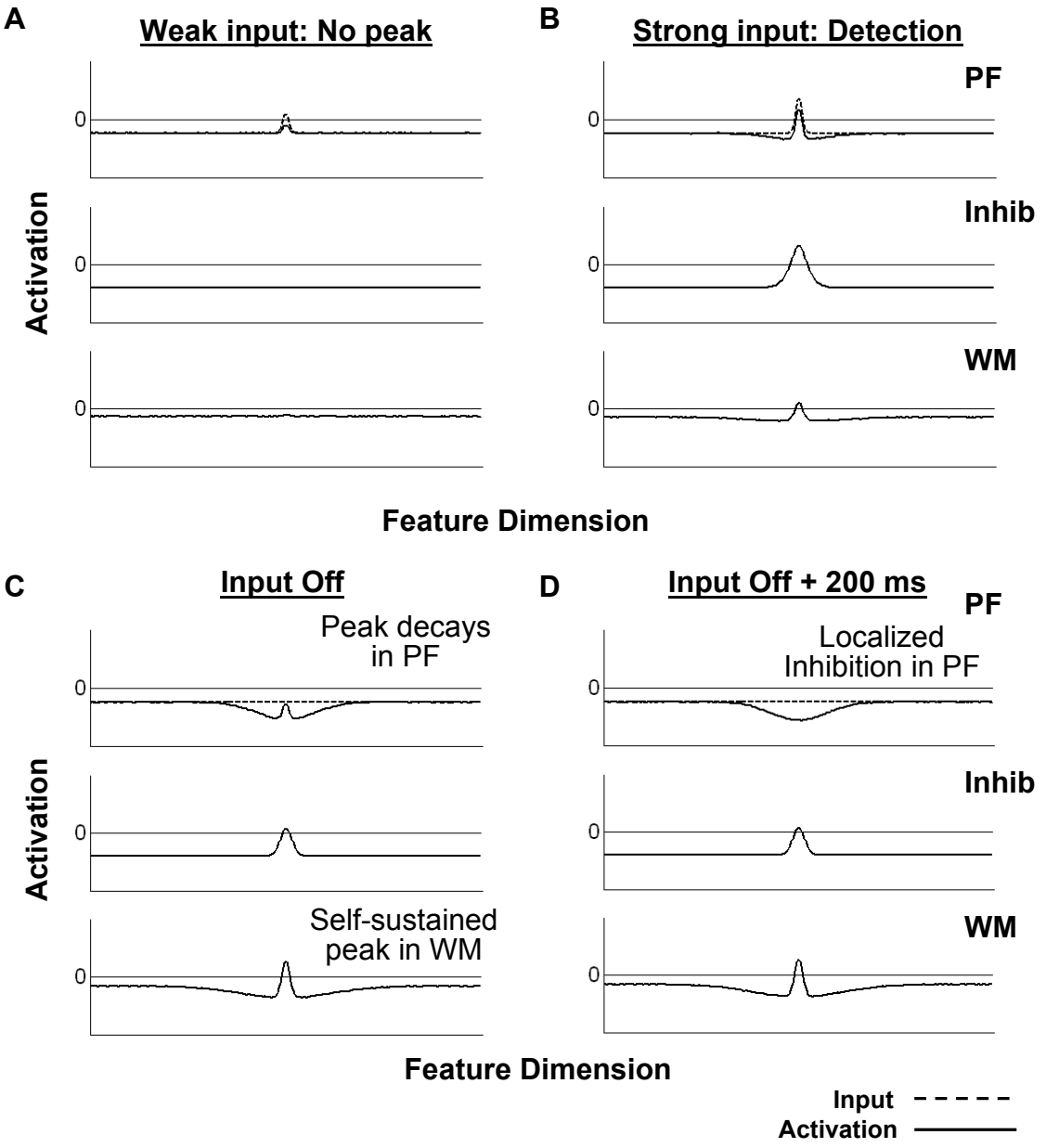

Figure 5. Simulations demonstrating the emergent functionality of the three-layer model in response to a single input. Dashed lines show the input to the model, whereas activation within each of the layers is represented by solid lines. (A) With weak input all field sites remain below threshold. (B) Detection of strong input in the task space (C) When input is removed, the peak decays in PF, but is sustained in the WM layer. (D) Formation of a localized region of inhibition in PF centered on value being maintained in WM.

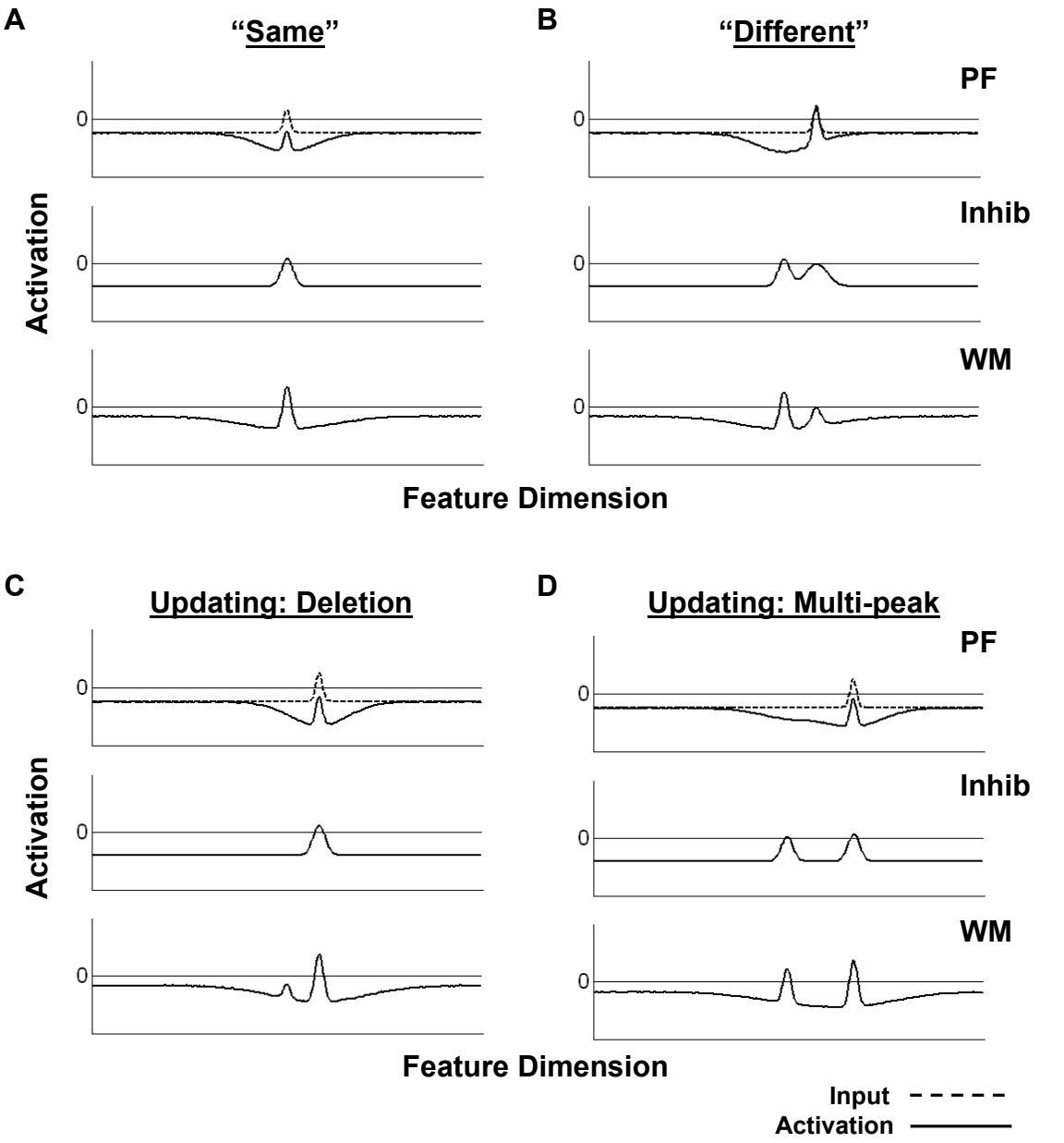

Figure 6. The generation of "same" (A) and "different" (B) responses in the model. A new peak is built in PF only when a new input that is sufficiently different than the value being held in WM is presented to the model. (C) and (D) show the updating of WM in response to new input. In the first case, the original peak is destabilized and replaced by a new peak, whereas, in the second case, a second peak is added to the field without destabilizing the original peak.

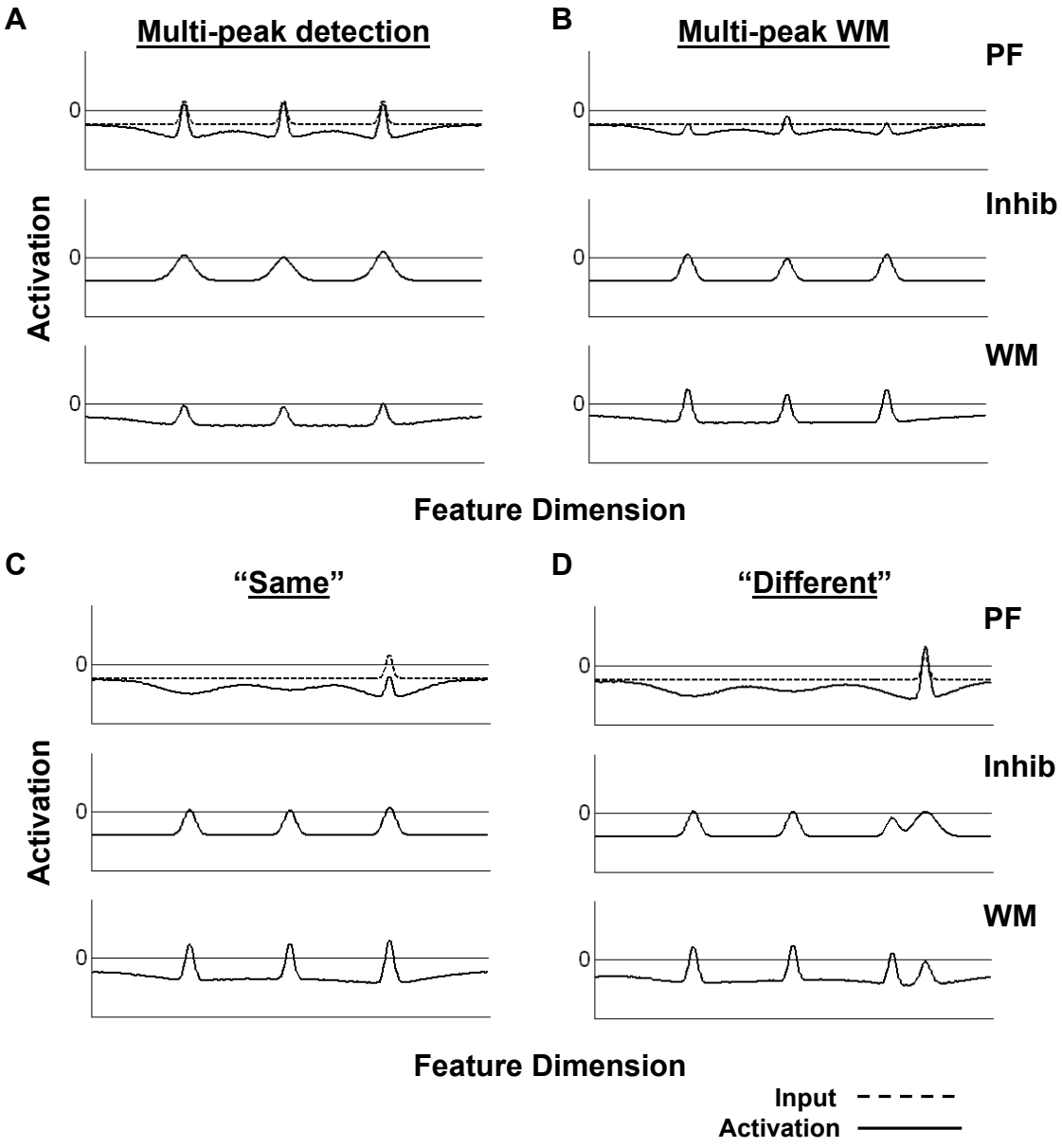

Figure 7. Simulations of the three-layer model in response to multiple simultaneous inputs. (A) and (B) show the detection of the inputs by PF, and their consolidation and maintenance in WM, respectively. The simulations in (C) and (D) show the generation of "same" and "different" responses in the multi-item case.

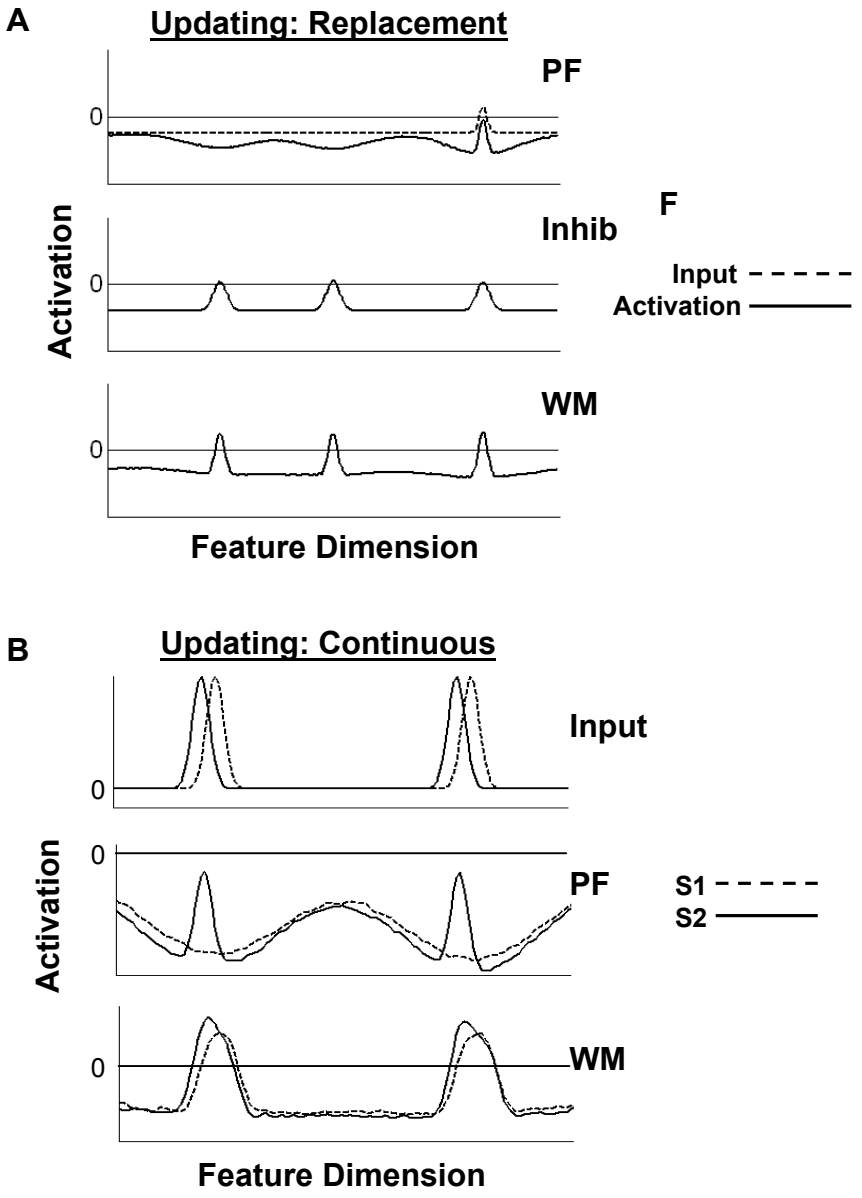

Figure 8. Two different forms of updating in the three-layer model. (A) The presentation of a changed input at test results in the destabilization of the original peak in WM and its replacement by the new value. In the second case (B), activation in WM is updated in a continuous fashion in response to new input. Note, that the scale of the simulations shown in (B) have been magnified, and the inputs to the model are now shown in the top panel. The first (S1) and second (S2) stimuli presented to the model are indicated by dashed and solid lines, respectively.

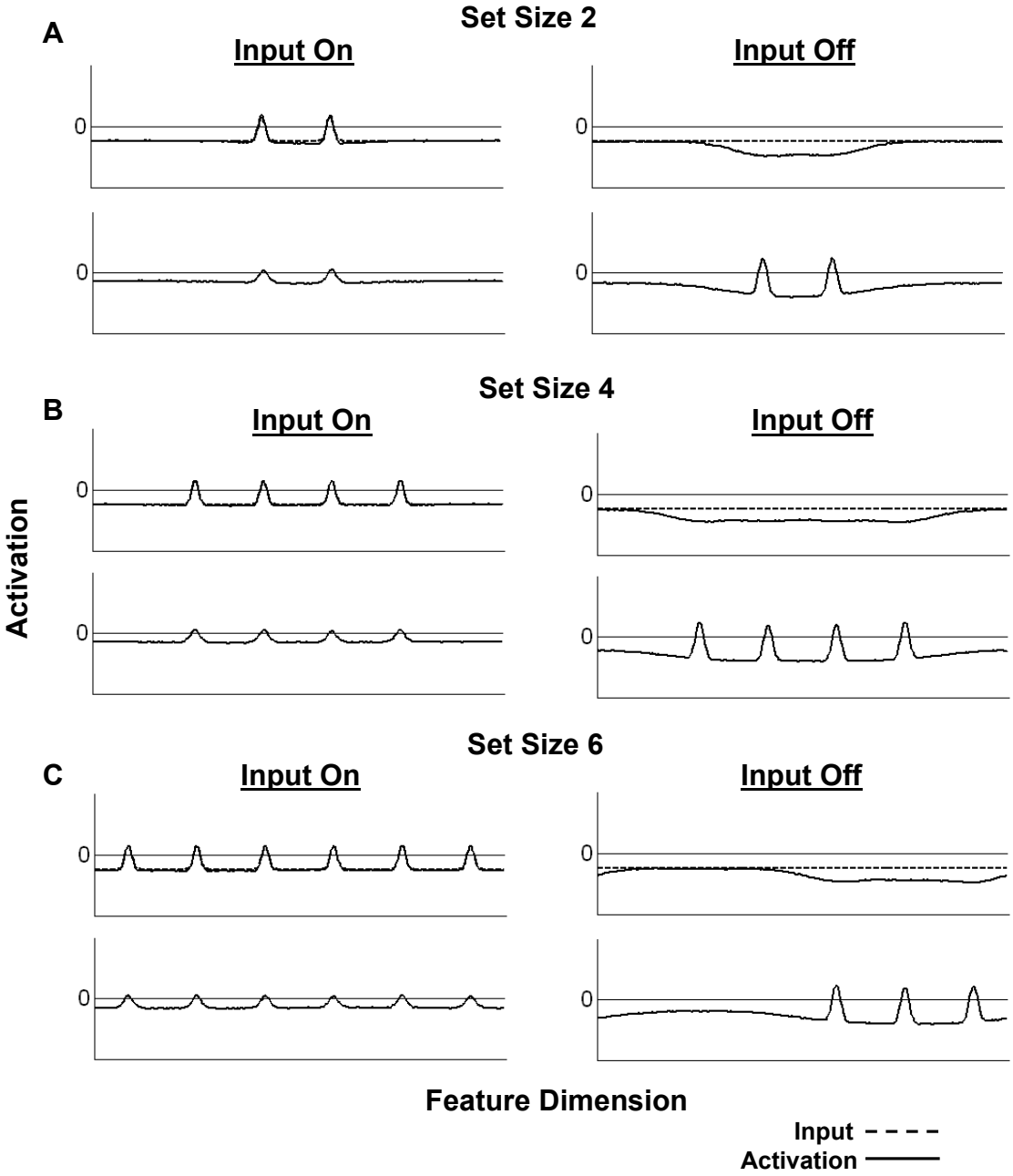

Figure 9. Simulations showing capacity limits in the three-layer neural field model. The simulations in (A) and (B) show the model successfully maintaining up to four peaks of activation simultaneously in WM. However, when two more items are added (C), inhibitory input to WM outweighs excitation, and three of the items fail to be sustained in the absence of input. Note that, for simplicity, the inhibitory layer is not shown here.

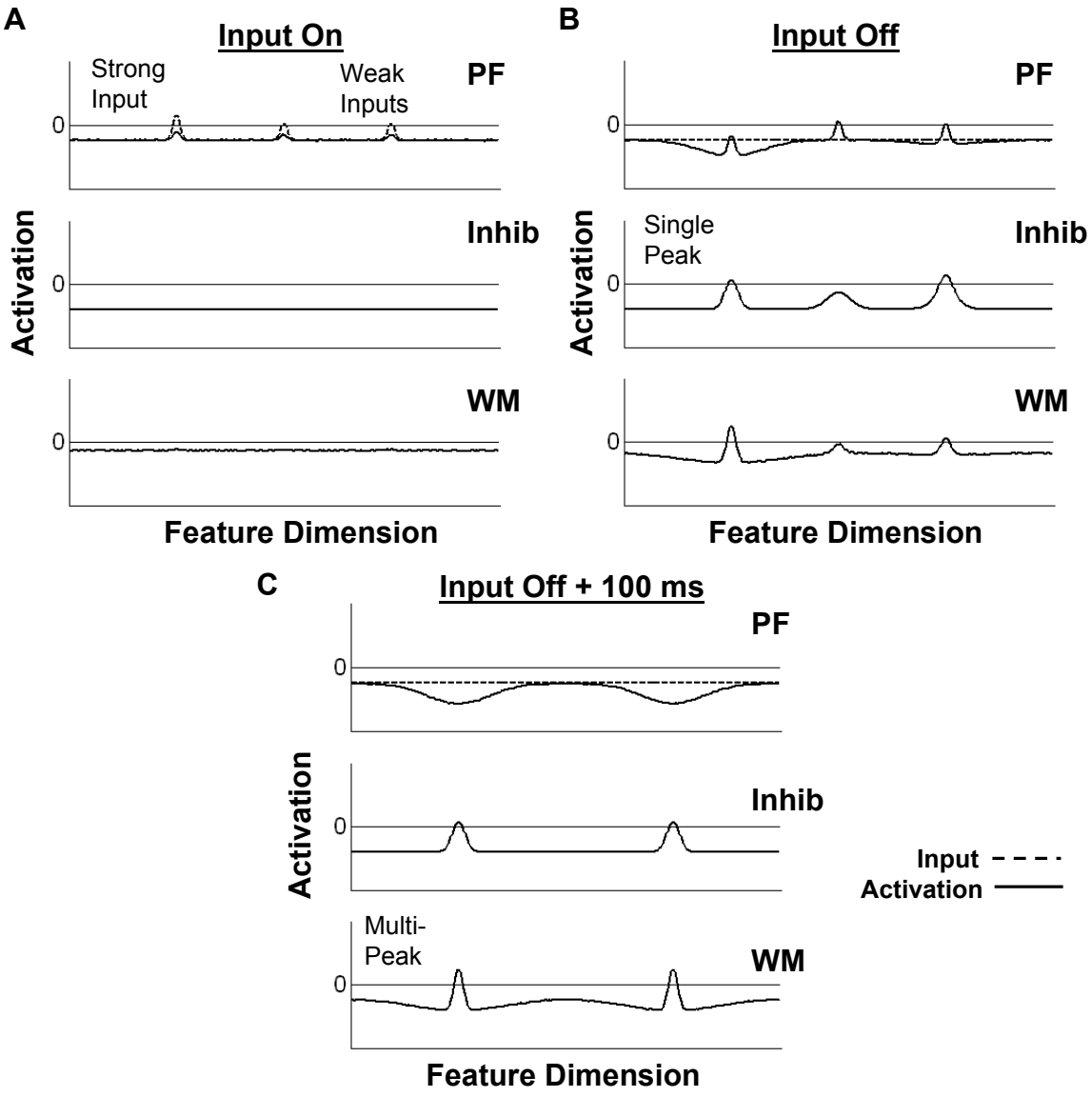

Figure 10. Simulations of the three-layer model with variable amplitude inputs. When one strong and two weak inputs are presented to the model (A), the strong input is successfully consolidated and maintained in WM (B). In addition, a short time later, one of the weaker inputs is also consolidated (C).

Table 1

Parameter Values Used in Chapter 2 Simulations

Layer	τ	*h*	Self-excitation	Excitatory projections	Inhibitory projections	Target input	Noise
u (PF)	80	-7	$c_{uu} = 2.0$		$c_{uv} = 1.05$	$c_{tar} = 30$	$c_{noise} = .04$
					$\sigma_{uv} = 24$	$\sigma_{tar} = 3$	$\sigma_{noise} = 1.0$
					$k_{uv} = .05$		
v (Inhib)	10	-12		$c_{vu} = 2.0$			
				$\sigma_{vu} = 10$			
				$c_{vw} = 1.95$			
				$\sigma_{vw} = 5$			
w (WM)	80	-4	$c_{ww} = 3.15$	$c_{wu} = 1.5$	$c_{wv} = .325$	$c_{tar} = 6$	
				$\sigma_{wu} = 5$	$\sigma_{wv} = 42$	$\sigma_{tar} = 3$	
					$k_{uv} = .02$		

Note: For the variable amplitude simulations shown in Figure 10, the strength of the strong input was set to $c_{tar} = 30$, and the weak inputs to $c_{tar} = 20$.

CHAPTER 3
ENHANCED CHANGE DETECTION WITH CLOSE METRICS

The model described in Chapter 2 represents the first neurally-plausible model to formalize the processes underlying the encoding and maintenance of multiple items in VWM *and* that specifies how same and different responses are generated in change detection tasks. In addition, the proposed model makes novel and counterintuitive predictions that can be empirically tested using the tools of vision science. In the present chapter, I test the prediction that *change detection will be enhanced when the items held in memory are highly similar*. This prediction arises as a result of metric-dependent interactions occurring among peaks in VWM. Specifically, when peaks are relatively far apart in feature space, reflecting, for instance, memory for multiple highly distinctive colors, they either do not interact, or they interact in a weakly inhibitory fashion. In contrast, when items are moved closer together in feature space they interact in a strongly inhibitory fashion.

The simulations depicted in Figure 11A-C illustrate one consequence of such interactions on the sustained patterns of activation produced in the WM layer of the model. Figure 11A shows the patterns of activation across the layers of the model when three items are being remembered: two ‘close’ colors that are very similar, and one ‘far’ color that is unique. In these simulations, shared lateral inhibition between close peaks produces a narrowing of the range of local excitation in WM. This can be seen if one compares the relative widths of the self-sustained peaks of activation formed in response to ‘close’ versus ‘far’ colors in the WM layer of Figure 11A. This in turn leads to a narrower and shallower inhibitory projection from WM to PF via the Inhibitory field (see PF in Figure 11A). Figure 11B-C show the consequences of this narrowing on change detection performance.

Figure 11B shows a simulation trial where one of the close colors changes at test. Because inhibition is relatively shallow in PF, when a changed item is presented, an above-threshold peak is formed and a 'different' response is generated. In contrast, when the far item changes by an identical amount (e.g., 30° in color space) at test, the input to PF meets with stronger inhibition, which makes it difficult to build a 'different' peak in response to the new input (see PF in Figure 11C).

Simulation Experiment 1: Enhanced change detection with close metrics

To quantify the consequences of the metric-dependent interactions described above for change detection performance, I conducted a simulation experiment where the model was run through a series of trials similar to what a human participant would experience. Specifically, on a given simulation trial, the model was presented with three inputs, two of which were close together and one that was far away. At test, either one of the close targets or the far target was probed with a single test input. The test input was presented either at the same location as the corresponding sample input, or at a different location 30 units away from this input. Thus, the only thing that varied across trials was the item that was probed at test (i.e., whether the probed item was a close or far target). It was expected that metric-dependent sharpening of nearby peaks would make it easier to detect changes when one of the close targets was probed at test, resulting in better performance for close versus far targets.

Methods

Model architecture

The model consisted of the three-layer model with response nodes shown in Figure 4, and described in Chapter 2. Each layer in the neural field model contained 361 feature-selective bistable neurons.

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 equation (model equations and parameters are given in Chapter 2) was integrated using the Euler method with one time step = 2 ms. For these simulations, the model was presented with three inputs on each trial, corresponding to different colors selected from a continuous 360° color space: Two close inputs that were 20, 30, or 40° apart in color space (1 unit in the model = 1° in color space), and a single far color that was always at least 150° away from the nearest close color.

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 500-ms presentation of the memory array, and a 1000-ms delay interval. Next, a test input was presented that was either identical to one of the initial inputs, or was centered at a different location 30° away from one of the memory array inputs. Upon the appearance of the test array, the resting level of the nodes in the response layer was increased, which allowed their activation to build beyond threshold with suitably strong inputs. Because the nodes in the response layer were coupled via strong inhibitory connections, only a single node surpassed threshold on each trial. The node whose activation reached threshold first, therefore, determined the model's response. For each close separation condition (20°, 30°, or 40°), the model completed 400 trials, 100 *same* and 100 *different* trials for both close and far targets, for a total of 1200 simulation trials.

Results and Discussion

Because the model makes a *same* or *different* response on each trial, I was able to calculate the signal detection sensitivity measure *d'*, as is done with human subjects. Simulation results are shown in Figure 13 (black) and demonstrate an advantage for close versus far targets. Thus, the model makes a clear prediction: change detection performance should be enhanced when similar versus distinct items are held in VWM.

The prediction that change detection should be enhanced with close metrics is rather striking considering decades of research that has shown just the opposite—that *similarity among representations in memory leads to impaired performance*. For example, studies of verbal working memory (Conrad, 1964) have shown that memory for phonetically similar words (e.g., "mad, cap, mat, pat, man") is usually worse than memory for phonetically distinct words (e.g., "get, tip, car, mad, buy"). Similarly, proactive interference has been shown to impair performance when the same materials are repeated across trials (Keppel & Underwood, 1962).

Do the close advantages predicted by the DFT also appear in human subjects? To test this prediction, I conducted a series of change detection experiments probing memory for both color (Experiments 1A and 1B) and orientation (Experiment 2).

Experiment 1A: Enhanced change detection for close colors with simultaneous presentation of sample array items

Methods

Participants

Ten University of Iowa undergraduate volunteers (9 women) participated in the present experiment. Participants received class credit or monetary compensation for their participation. Participants in this and all additional experiments reported normal or corrected-to-normal visual acuity and normal color vision.

Stimuli

Stimulus presentation and response recording was controlled by a Macintosh G4 computer running Matlab 5.2 using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli were presented against a gray background (28.73 cd/m^2) on a 17" CRT monitor at a viewing distance of 57 cm. Memory stimuli consisted of small (1.7° × 1.7°) colored squares. Individual colors were selected from a set of 180 colors equally distributed in CIELAB (1976) color space (centered at CIE L*a*b* coordinates: L=70, A=28, B=12), a perceptually uniform color space and color appearance model.

Equal distances in CIELAB color space correspond to approximately equal perceptual color differences as perceived by a human observer. All colors had equal luminance and varied mainly in hue and slightly in saturation.

Sample arrays consisted of three items presented simultaneously at least 90° apart on the circumference of an imaginary circle centered at fixation and with a radius of 4.25 cm. Two of the sample array items were close in color space (20°, 30°, or 40° apart), whereas the color of the third item was always at least 150° away from the nearest close color. The test array contained a single item appearing at one of the spatial locations previously occupied by a sample array item (see Figure 12A). Change magnitude was fixed at 30º in color space.

Procedure

Individual trials began with the appearance of a fixation cross at the center of the screen for 500 ms, followed by the 500-ms presentation of the sample array. This was followed by a 1000-ms delay interval and the appearance of a test array that remained visible until a response was generated. Participants were instructed to make an unspeeded response to the test array, indicating whether the color of the test item was the same as or different than the item appearing at that location originally. Participants completed a practice block of 24 trials, and a total of 240 experimental trials: 40 close and 40 far trials at each close color separation (probability of change = 0.5). Finally, to prevent verbal recoding of the sample array colors, participants repeated three randomly generated digits (e.g., "6,4,9") out loud at a regular pace throughout each trial.

Results and Discussion

An alpha level of .05 was used as the criterion for statistical significance in all experiments, and the signal detection measure *d´* was the primary dependent measure.

An analysis of variance with Target Type (close, far) and Close Separation (20°,30°,40°) as within-subjects factors revealed a significant main effect of Target Type, $F(1,9)=29.82$, $p < .001$. No other effects reached statistical significance, all Fs < 1. As

predicted by the DFT, change detection performance was improved when a close color was probed relative to performance when a far color was probed (see dark gray bars in Figure 13). This effect held whether the close targets were 20, 30, or 40° apart in color space.

These findings confirm the predictions of the model, suggesting that items in VWM interact in a metric-dependent fashion. However, before accepting this conclusion, it was necessary to rule out an alternative explanation for these results. Because all of the colors were simultaneously visible on the screen, and occupied relatively nearby spatial locations, it is possible that the enhancement effect reported here reflects differential perceptual encoding of close versus far colors, rather than interactions among similar items in VWM (see discussion in, Lin & Luck, in press). For instance, low-level contrast effects may have made the close colors more perceptually distinct. Or, alternately, participants may have remembered differences between the two close colors, rather than the colors themselves. To rule out these possibilities, I ran a second experiment where the sample array items were presented sequentially at widely separated spatial locations, rather than simultaneously at nearby locations.

Experiment 1B: Enhanced change detection for close colors with sequential presentation of sample array items

Methods

Participants

Twelve University of Iowa undergraduate volunteers (7 women) participated in the present experiment.

Stimuli

The stimuli used in this experiment were very similar to those used in Experiment 1, with the following exceptions: 1) Stimuli were presented on a 19” CRT monitor, 2) stimuli subtended 2.0° × 2.0° of visual angle, and 3) stimuli were presented on the

circumference of an imaginary circle with a radius of 7.5 cm, versus 4.5 cm in Experiment 1A.

Procedure

The procedures were identical to those described for Experiment 1A with the exception that sample items were presented sequentially, rather than simultaneously (see Figure 12B). Individual trials involved the presentation of three sample array items one at a time for 200 ms each, separated by a 500-ms blank interval. Presentation of the third item was followed by a 1000-ms delay interval, and the appearance of a single test item, as above. Close and far targets appeared with equal likelihood in each probe position (1st, 2nd, or 3rd), and in all possible orders. Close and far targets were probed equally often at each probe position and in each order. Participants completed 48 trials (24 change and 24 no-change) probing each target type (Close, Far), at each probe position (1, 2, 3), and for each close color separation (20°, 30°, 40°), for a total of 864 trials.

Results and Discussion

An ANOVA with Target Type, Probe Position, and Close Separation as within-subjects factors revealed a main effect of Probe Position, $F(2,22) = 14.60$, $p < .001$, showing better change detection performance for the most recently viewed item and worst performance for the first item viewed (Mean d': P1 = .88, P2 = 1.0, P3 = 1.47). In contrast to the results of Experiment 1A, there was also a main effect of Close Separation, $F(2,22) = 17.99$, $p < .001$—change detection performance was most accurate when the close colors were separated by 20°, and least accurate when they were separated by 40° (Mean d': 20° = 1.30, 30° = 1.12, 40° = 0.92). Most critically for the present purpose, however, there was also a main effect of Target Type, $F(1,11) = 24.01$, $p < .001$. As in Experiment 1A, change detection performance was enhanced when a close versus a far color item was probed (see white bars in Figure 13). These results, therefore, replicate Experiment 1A, confirming that the enhancement effect does not reflect differential perceptual encoding of close versus far colors.

Experiment 2: Enhanced change detection for close orientations

Results from Experiment 1A-B are consistent with the predictions of the DFT. In Experiment 2, I examined whether the enhancement effect is a general property of the neural mechanisms that underlie VWM or instead reflects some property peculiar to working memory for color. To do this, I used the sequential change detection paradigm from Experiment 1B to ask whether this effect generalizes to working memory for orientation.

Methods

Participants

Fifteen University of Iowa undergraduate volunteers (10 women) participated.

Stimuli

Stimulus presentation and response recording was as in Experiment 1B. Memory items consisted of thin, black rounded rectangles that spanned the interior of a small light-gray circular background (2.0° in diameter; see Figure 12C).

Sample arrays consisted of three items presented sequentially as in Experiment 1B. Target separations and change magnitudes were selected to match those used for color in Experiments 1A-B. Specifically, for each trial, two of the memory display items had similar orientations (20º, 30º, or 40º apart), whereas the orientation of the third item was at least 70º away from the nearest close orientation (due to the compression of the orientation feature space, we could not present the far item 150° away from the close items as in Experiment 1B). When a change occurred at test, the orientation of the test item was rotated by 30º. Note that, although we attempted to keep the metric details of the stimuli consistent across Experiments 1B and 2, the representation of metric structure is likely to differ substantially across separate feature dimensions.

Procedures

All procedures used in this experiment were identical to those reported for Experiment 1B.

Results and Discussion

An ANOVA with Target Type, Probe Position, and Close Separation as within-subjects factors revealed a main effect of Probe Position, $F(2,28) = 30.24$, $p < .001$. As in Experiment 1B, change detection performance was best for the most recently presented item and worst for the first item viewed (Mean d': P1 = 1.01, P2 = 1.12, P3 = 2.00). In addition, there was a main effect of Close Separation (20º,30º,40º), $F(2,28) = 3.40$, $p < .05$, which was subsumed by a significant Target Type × Close Separation interaction, $F(2,28) = 4.06$, $p < .03$. Tests of simple effects revealed a significant Target Type effect in the 20° close separation condition (see light gray bars in Figure 13), $F(1,2) = 7.67$, $p < .02$, but not in the 30° or 40° conditions, all F's < 1. This finding supports the proposal that enhanced change detection with close metrics is a general property of working memory for metric feature dimensions, rather than a peculiarity of the color system. However, not surprisingly, the metric separations at which this effect arises differ across the two dimensions probed here.

General Discussion

In the present chapter, I conducted three different experiments looking for specific *behavioral signatures* of the processes proposed to underlie working memory in the model described above. In particular, I proposed that metric-dependent laterally inhibitory interactions among similar items in VWM produces a narrowing of the range of local excitation associated with each item. Through the mechanism of change detection described in Chapter 2, I showed how such interactions lead to enhanced change detection when metrically similar features are remembered. Results from experiments probing memory for color and orientation were consistent with this novel prediction.

These findings are somewhat surprising given that research in the verbal domain has consistently revealed performance *decrements*, rather than improvements, for phonetically similar versus dissimilar words (Conrad, 1964). What can account for the differences across modalities? One possible source of these discrepant findings is differences in the complexity of verbal versus visual stimuli. For instance, the stimuli in the experiments reported here consisted of simple visual objects, where a single feature needed to be remembered (e.g., its color or orientation). Although the phenomenal appearance of color can be decomposed into separate components (e.g., hue, saturation, and luminance), the stimuli used here varied primarily in hue, and their similarity in hue space was carefully controlled. Visual cortical neurons have been found that are specifically selective for the hue component of color (Komatsu, Ideura, Kaji, & Yamane, 1992). In contrast, even relatively simple words are not so easily defined, and may vary along multiple auditory dimensions, including temporal duration, timbre, pitch, and loudness, which are represented in a complex and as yet poorly understood fashion by the brain (Young, 1998). Thus, a simple word may be more akin to a multi-feature object (e.g., a colored, tilted, rectangle) than a simple visual feature (e.g., the color blue). It is unclear whether the enhancement effects found here would arise in more complex multi-feature objects whose similarity varies along multiple dimensions.

Another possible source of differences has to do with the responses required in each task. In the change detection experiments reported here, participants saw one visual display, and generated a 2-AFC response at test, indicating whether a test item was the same as or different than a corresponding item in the memory display. Thus, the change detection task places relatively modest non-memory demands on participants. In the classic phonological similarity experiments of Conrad (1964), however, participants heard a list of words that they were required to remember, and were asked to recall as many of these words as they could at test. Although simple, the need to verbally recall each of the remembered words may have interfered with the storage of the other words in

the list. If more similar words are somewhat more prone to disruption, this would explain the difference.

The present study reveals a direct behavioral consequence of the dynamic neural processes supporting maintenance in the DFT. This finding is counterintuitive given previous research showing interference effects for similar items in a variety of memory tasks. It also runs counter to the predictions of other neural approaches to maintenance in VWM (see, e.g., Raffone & Wolters, 2001). Such models achieve independent memory representations via temporal synchrony. It is not clear how the metric-dependent effects reported here might arise from this neural mechanism. The present findings, therefore, place strong constraints on models addressing the nature of and mechanisms underlying visual working memory.

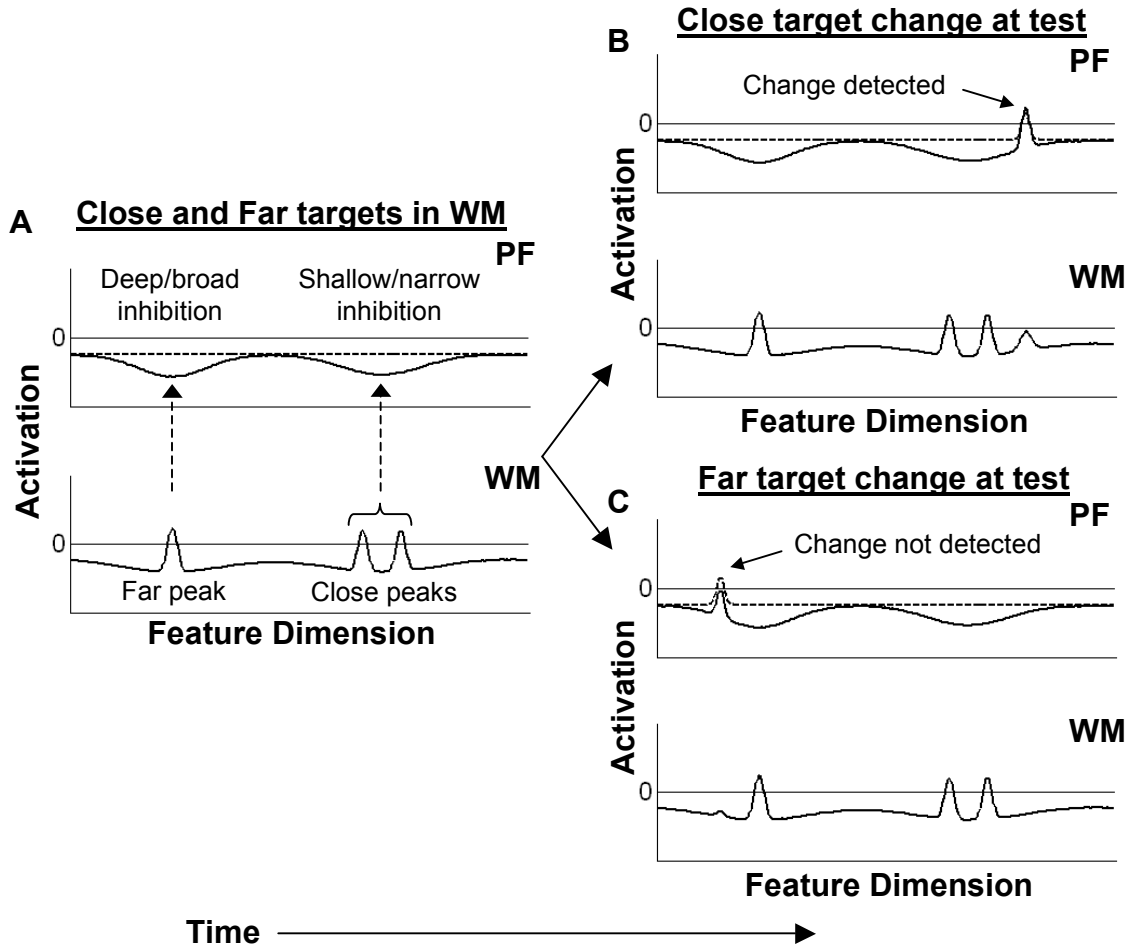

Figure 11. Simulations showing differences in patterns of activation in WM and PF as a function of metric similarity among items. Shared lateral inhibition between nearby items (i.e., similar colors) narrows the range of local excitation in WM, and, as a result, the strength and width of feedback inhibition to PF during the delay interval of a change detection task (A). Because the inhibition associated with the close peaks is weaker and narrower in PF, changes are easier to detect when one of the two close targets changes (B), versus the far target (C).

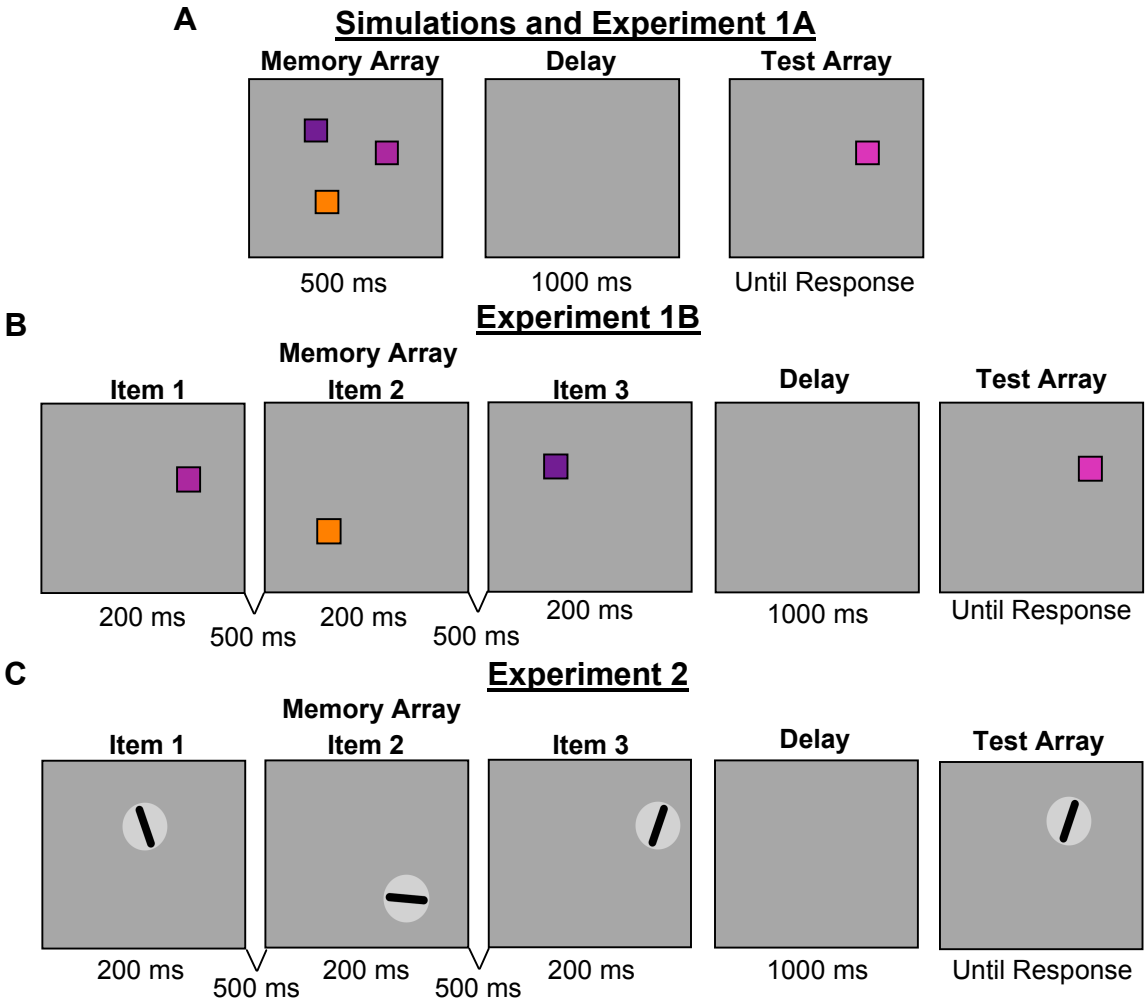

Figure 12. Stimuli and procedures from Simulation Experiment 1, Experiments 1A-B and Experiment 2. The sequence of events in a single trial in simulations and Experiments 1 and 2 using simultaneous (A) and sequential (B-C) change detection paradigms probing color and orientation, respectively.

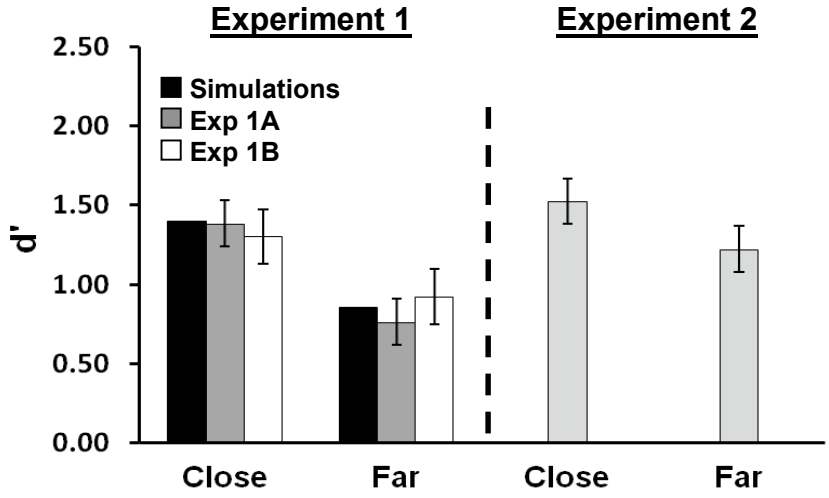

Figure 13. Results from Experiments 1A-B, Experiment 2, and model simulations showing enhanced change detection for metrically-similar colors and orientations. In each case, performance was reliably better for 'close' versus 'far' targets.

Table 2

Parameter Values Used in Simulation Experiment 1

Layer	τ	*h*	Self-excitation	Excitatory projections	Inhibitory projections	Target Input	Noise
u (PF)	80	-7	$c_{uu} = 2.0$		$c_{uv} = 1.05$	$c_{tar} = 30$	$c_{noise} = .04$
					$\sigma_{uv} = 24$	$\sigma_{tar} = 3$	$\sigma_{noise} = 1.0$
					$k_{uv} = .05$		
v (Inhib)	10	-12		$c_{vu} = 2.0$			$c_{noise} = .04$
				$\sigma_{vu} = 10$			$\sigma_{noise} = 1.0$
				$c_{vw} = 1.95$			
				$\sigma_{vw} = 5$			
w (WM)	80	-4	$c_{ww} = 3.15$	$c_{wu} = 1.5$	$c_{wv} = .325$	$c_{tar} = 6$	$c_{noise} = .04$
				$\sigma_{wu} = 5$	$\sigma_{wv} = 42$	$\sigma_{tar} = 3$	$\sigma_{noise} = 1.0$
					$k_{uv} = .02$		
Different	80	-5	$c_{dd} = 1.85$	$c_{du} = 1.0$	$c_{id} = 14$		$c_{noise} = .1$
Same	80	-5	$c_{ss} = 1.85$	$c_{sw} = .05$	$c_{di} = 14$		$c_{noise} = .1$

CHAPTER 4
MUTUAL REPULSION BETWEEN CLOSE FEATURES IN VISUAL WORKING MEMORY

The present chapter reports an experiment designed to test a second prediction arising as a result of close metrics in VWM: *that laterally inhibitory interactions among items will lead to mutual repulsion between nearby peaks*. This arises as a result of overlapping inhibition, which is stronger in-between nearby peaks than it is on the "outer side" of each peak (see close peaks in Figure 14A). As a result, it is easier for the excitation associated with each peak to grow in a direction away from the other peak (i.e., away from the other item in memory) across the delay interval. This leads to the prediction that when similar features are held in VWM, recalled estimates of these features will be systematically biased away from each other over delays. In contrast, memory for unique (i.e., "Far") features should not be systematically biased, and recalled estimates should be comparable to performance when only a single item is remembered.

Simulation Experiment 2: Mutual repulsion between close features in VWM

As in Chapter 3, a simulation experiment was run to quantify the magnitude of repulsion induced by close metrics in the model. I began by measuring repulsion using the same model and parameters as used in Simulation Experiment 1. These simulations revealed a small but reliable repulsion effect of approximately 1° in each direction, a rather modest effect. Further simulations revealed that to get robust repulsion, the width of the self-excitatory component of the interaction function for the WM field (σ_{ww}) needed to be wider than it was in Simulation Experiment 1. Recall that repulsion effects emerge in the model when inhibition is stronger on one side of a peak versus the other, allowing excitation to build more easily in that direction. When the width of excitatory interactions within WM is very small, there is little difference in the excitatory/inhibitory

gradient on either edge of a peak because the peak effectively samples neural interactions across a very limited range. As a result, peaks in the model show very little repulsion. Thus, to explore repulsion in the model in greater detail, I modified the network in two ways. First, I increased σ_{ww} from 3 to 5. Second, to prevent the close peaks from fusing together on every trial, I also increased the size of the field to 540 neurons and re-scaled the mapping from neuronal units in the model to degrees in color space such that 1.5 units in the model equaled 1 degree in color space. With these modifications, the model produced robust repulsion.

Methods

Model architecture

The model architecture was identical to the three-layer model used in Simulation Experiment 1 and described in Chapter 2, with the exception that no response nodes were used, and the field size was increased from 361 to 541 units. Thus, each 1.5 units in the model represented 1° of color space, rather than the one-to-one mapping of units to degrees used in the simulations reported in Chapter 2. Parameter modifications for these simulations are shown in Table 2.

Simulation details

For these simulations, the model was presented with either a single input (SS1 condition) or three inputs (SS3 condition) on each trial: two close inputs presented at the same two locations 20° apart in color space, and a single far input that was selected randomly with the constraint that it was always at least 150° away from the nearest close input. Each trial consisted of a 200-ms pre-relaxation period, the presentation of inputs for 800-ms, a 1000-ms delay interval, and a 3000-ms response interval, after which the model's estimate was determined. The model was run through 200 SS1 trials, and 600 SS3 trials: 400 trials probing the close inputs (200 CW and 200 CCW), and 200 probing the far inputs, for a total of 800 trials.

Results and Discussion

Recall simulations

To determine estimation errors in the model, I calculated the distance (in degrees) between the actual location of the input, and the location of the mean of the cluster of above-threshold activation defining the peak in the WM field associated with the cued item. Although this is clearly a simplification of the processes that are likely to be involved in making such a response, this mapping to recalled estimates has been used extensively in the spatial working memory literature (see, e.g., Schutte, Spencer, & Schöner, 2003; Simmering et al., 2008).

Results are shown in Figure 14B. As can be seen, color estimates in the SS1 condition were very close to zero, as were estimates of the Far target in the SS3 condition. In contrast, estimates of the CW and CCW targets were biased away from each other by ~5° in each direction. A sample repulsion trial in the SS3 condition is illustrated in Figure 14A.

Change detection simulations

The modified model shows a robust repulsion effect. However, to obtain this effect I needed to modify the model in several ways (see Table 2). Can the enhancement effect found in the change detection simulations reported in Simulation Experiment 1 also be achieved with the new model and new parameter set? To determine this, I re-ran the change detection simulations described above using the modified model. The primary difference between the change detection and recall simulations was the length of time that the model was run. For the recall simulations, I added on a 3s period at the end of the delay interval as a stand-in for the time needed to make a recall response. In contrast, for the change detection simulations, a test input was presented at the end of the delay interval, which led to the generation of a same or different response by the model. Thus, for recall, peaks were free to keep drifting for some period of time after the delay interval was over, whereas for change detection, new inputs came into the field precisely at the

end of the delay. Results revealed a 4% improvement in performance for close versus far targets, which was comparable to the enhancement effect found in Simulation Experiment 1. Thus, the model used in this chapter was able to produce both repulsion between close peaks in the context of recall *and* an enhancement effect for close colors in change detection.

Experiment 3A: Mutual repulsion between close colors in VWM

The simulations reported above make the prediction that close features in WM interact in an inhibitory fashion, and that these interactions can lead to mutual repulsion. To test this rather subtle prediction, a measure of VWM that is more sensitive than the change detection task used in Experiments 1A-B and Experiment 2 is needed. For this purpose, I developed the cued color recall task depicted in Figure 15. The basic structure of each trial was very similar to a typical trial in the change detection experiments reported above, with the exception that participants made a cued color recall response following the delay, rather than a 2-AFC same/different response. Specifically, for each trial, participants were shown a sample array that contained two similar colors and one distinct color presented at widely separated spatial locations. After a brief delay, a continuous color wheel was presented around one of the three target locations, cueing the participant to estimate the color that had appeared at that location originally. Participants estimated the remembered color by moving a mouse cursor to the color on the color wheel that matched the color being held in working memory.

Methods

Participants

A total of 12 undergraduate students (7 Women) participated in this experiment for course credit. All participants reported normal or corrected-to-normal visual acuity and normal color vision.

Stimuli

All stimuli were presented on a 30" LCD monitor[1] with a gray background (28.73 cd/m2), using custom software developed in Visual Studio 2005. The viewing distance was 57 cm. Sample arrays consisted of either 1 or 3 small (1.9° x 1.9°) colored circles. As in Experiments 1A-B, individual colors were selected from a set of 180 colors equally distributed in CIELAB (1976) color space (centered at CIE L*a*b* coordinates: L=70, A=28, B=12). Each stimulus was centered on an invisible circle with a radius of 7.5° from the center of the screen, with a minimum distance between targets of 80 degrees of angular rotation (see Figure 15).

For set size 3 (SS3) trials, two of the sample colors were always close in color—one clockwise (CW) target, and one counter-clockwise (CCW) target 20º apart in color space—whereas the third color was always at least 150º away from the nearest close color. For set size 1 (SS1) trials, a single color was randomly chosen from the 180 possible colors making up the color space. However, to prevent possible carry-over effects from trial to trial, the choice of sample array colors in both conditions was constrained such that none of the sample colors used on trial N were within +/- 40° in color space of the color that was estimated on trial N-1.

Each trial began with the 800-ms presentation of a sample array. This was followed by a 1000-ms delay interval and the appearance of a test array that remained present until a response was made. The test array contained a color wheel with an outlined white circle at its center whose size and location matched one of the original sample array items. The color wheel contained each of the 180 possible sample colors

---

[1] The monitor used here is both larger (30" versus 17" in Experiment 1A, and 19" in Experiments 1B and 2), and of a different type (LCD versus CRT) than those used in the change detection experiments reported above. Differences in color reproduction characteristics across monitors could lead to a different pattern of results depending on the monitor used. However, a somewhat different study testing the same idea has been run using the 17" CRT monitor used in Experiment 1A. The results replicated the findings reported below.

equally distributed along a "color wheel" in CIELAB (1976) color space. The appearance of the color wheel at a particular location cued the participant to estimate the color that had appeared there originally. Estimates were made by moving a set of crosshairs over the color wheel using a computer mouse, and left-clicking on the color that most closely matched the color being held in memory. After the response, the outlined circle was replaced by the original sample array color, and a second set of crosshairs appeared over the correct location on the color wheel. Accuracy was stressed rather than speed, and estimation responses were not timed. To prevent spatial strategies from affecting color estimates, the orientation of the color wheel was randomized across trials. In addition, the color wheel was presented in a standard and a mirror reversed fashion, where the direction of colors within the wheel went in the opposite direction.

At the beginning of the experiment, participants read a written description of the task, and the experimenter demonstrated the stimuli and procedures. Participants then completed 24 practice trials followed by 240 experimental trials: 80 SS1 trials and 160 SS3 trials (80 close color trials and 80 far color trials). Because I was specifically interested in how estimation performance would differ for the two close color targets, half of the close color trials probed the CW close color and the other half probed the CCW close color. All trial types were randomly intermixed. Participants were offered a short break after every 24 trials, and the entire experimental session lasted ~30-40 minutes.

Finally, to prevent verbal recoding of the sample array colors, participants repeated three randomly generated digits (e.g., "6,4,9") out loud at a regular pace throughout each trial.

Results and Discussion

The primary dependent measure in the present experiment was mean directional error in color space (in degrees). To arrive at this value, for each trial, I computed the angular difference between the position of the cued color and the position on the color

wheel that was selected by the participant. Errors that were clockwise from the cued color in color space were assigned positive values, whereas errors that were counter-clockwise were assigned negative values. For trials where the direction of colors in the color wheel were mirror reversed, the sign of the errors was flipped to maintain consistency.

Recall that the model predicts specific patterns of errors depending on whether the cued color was a close or a far target, and, for the close targets, on whether the close target was a CW or CCW target. Specifically, because the far target is presumed to be outside the range of strong lateral interaction in VWM, there should be no systematic biases to this target. That is, mean error for the far target should not be significantly different from mean error found in the SS1 baseline condition, where only a single color was remembered on each trial. In contrast, errors to the two close targets should be biased in opposite directions (i.e. repelled from each other), depending on their relative positions in color space. Specifically, mean error to the CW targets should be positive (i.e., further clockwise in color space), whereas errors to the CCW target should be negative (i.e., further counter-clockwise in color space).

Overall, mean directional errors to targets in the SS1 condition were very small and slightly negative (Mean error = -2.14º). Nonetheless, a one-sample t-test revealed that this bias was significantly different from zero error, $t(11) = 2.345$, $p < .04$. Although the origin of this slight CCW error is unclear, the systematic deviations with set size 1 suggest that there is some estimation bias in the recall task, even when only a single target is held in memory across a short delay.

Figure 15 shows mean directional errors for the Far, CW, and CCW targets in the SS3 condition. Given the small but systematic CCW bias found in the SS1 condition, we normalized the errors shown here with respect to the -2.14 baseline error estimate. An ANOVA comparing performance in the CW, CCW, FAR, and SS1 conditions revealed a significant main effect of target, $F(3,33) = 8.51$, $p < .001$. Errors to the Far target in the SS3 condition were nearly identical to errors in the SS1 condition (Mean difference =

.37º). This was confirmed in a paired-samples t-test, $t(11) = .312$, $p = .761$. Thus, estimates of the Far target do not appear to be strongly influenced in a systematic way by the other two colors being held in VWM, in keeping with the first prediction discussed above.

By contrast, errors to the CW and CCW targets differed substantially from errors in the SS1 condition, and were biased in opposite directions. Specifically, estimates of the CW target were biased in a positive direction, whereas estimates of the CCW target were biased in a negative direction, as predicted. This was confirmed in a paired-samples t-test comparing the CW and CCW conditions, $t(11) = 5.05$, $p < .001$.

The results reported here support the second prediction of the DFT, suggesting that laterally inhibitory interactions can lead to mutual repulsion between metrically-similar items in VWM. This finding runs counter to what would be expected on the basis of existing theories of VWM. For instance, a prevalent idea in the literature is that VWM stores a fixed number of discrete, slot-like representations (Zhang & Luck, in press). If representations in VWM were truly discrete, one would not expect the kinds of interactive effects demonstrated here.

A similar idea is embodied in the neural synchrony model of Raffone and Wolters (2001) discussed above. According to this model, the features of objects are represented by distinct subpopulations of neurons, and are maintained separately in VWM. As a result, there is no basis in such models for the type of interactions found here.

Experiment 3B: Mutual repulsion between close colors in VWM with dense sampling

Experiment 3B was identical to Experiment 3A with the exception that participants completed many more trials. This allowed the data to be fit using the Resolution, Accuracy, and Capacity (RAC) model proposed by Zhang (2007). This model makes it possible to separately measure the resolution and accuracy of the representations stored in VWM, as well as the number of items that were stored. The

basic logic of this approach is that, if the observer stores the probed item in memory, estimation responses will be near the original value. Conversely, if the item was not stored in memory, responses should be random and uniformly distributed throughout the 360º color space. These two types of trials wind up being mixed together in the data, making it difficult to separate out variance reflecting the resolution of the VWM representations themselves, from variance arising as a result of random guesses. The model of Zhang proposes a method whereby the components of the response distribution representing these different types of trials can be recovered using standard estimation procedures. This procedure decomposes the observed data into three parameters: the width (quantified as the standard deviation [SD], σ) and central tendency (quantified as the mean, μ) of the distribution, representing trials where the item was stored; and the height of the tails of the mixed distribution, representing the probability that the probed item was stored in memory (probability of encoding, P_e).

Thus far, this framework has been used as a means of clarifying the nature of the capacity limits in VWM. Towards this end, the majority of work done within this framework has focused on the number of representations that can be stored, and their precision (see Zhang, 2007). Results have shown that precision (i.e., σ) is high (σ is low) when only one item must be remembered, and decreases as the set size is increased toward the capacity of VWM. As set size increases beyond capacity, precision plateaus and the probability of encoding decreases, providing support for the notion that VWM "slots" are fixed in number and have a fixed resolution/precision. Note that Zhang (2007) only probed memory for distinctive items; this might explain why mean estimates were always near 0 error. In this context, evidence of mean shifts for CW and CCW targets with dense sampling and using a robust response estimation procedure would constitute a significant test of the robustness of the DFT's predictions. Furthermore, the use of the estimation procedure developed by Zhang allows us to probe whether close vs. far items are represented with the same precision.

Methods

Participants

A total of 9 students (6 Women) participated in this experiment in exchange for monetary compensation. All participants reported normal or corrected-to-normal visual acuity and normal color vision.

Stimuli and Procedures

Stimuli and procedures for this experiment were identical to Experiment 3A, with the exception that many more trials were completed. The increase in trials was needed for robust estimation of distribution parameters using the method developed by Zhang (2007). Participants completed 160 SS1 trials, 320 Close trials (160 CW, 160 CCW), and 320 Far trials. In addition, participants completed 32 practice trials at the beginning of the experiment, for a total of 832 trials. The experimental session lasted approximately 1.5 hours, and was completed in a single session. Participants were offered breaks at regular intervals, and were encouraged to leave the testing room for a longer break half way through the experiment to prevent fatigue.

Results and Discussion

The parameters of the response distributions for each participant and each target type (SS1, Far, CW, CCW) were estimated using the method developed by Zhang (2007). Overall, seven out of nine participants produced the expected repulsion effect. An additional two participants showed the opposite pattern of results, showing substantial attraction toward the other close color in memory (Mean CW error = -7.85º; Mean CCW error = 3.91º). It isn't clear what led to these differences, but this may reflect the fusion limit case in the DFT, where peaks of activation that are close together and relatively unstable merge into one peak as a result of strong mutually-excitatory interactions. Interestingly, the pattern of results for SD and P_e for these participants was similar to the group showing repulsion. Nevertheless, because these participants showed a qualitatively different pattern of bias, they are not considered further here.

Average obtained parameter estimates from the remaining seven participants are plotted in Figure 16. The full distribution of responses obtained in Experiment 3B and in the simulations that follow can be seen in Figure 18. As in Experiment 3A, mean directional errors in the SS1 condition were small and biased in a slightly counterclockwise direction (M = -1.19º). A one-sample t-test revealed that errors in this condition were again significantly different from zero, $t(6)=-2.99$, $p < 0.03$. Thus, in each experiment there was a small counterclockwise bias for targets in the SS1 condition.

Figure 16A shows mean directional errors in the Far, CW, and CCW conditions. As before, mean errors were normalized with respect to the -1.19º bias found for SS1. An ANOVA comparing performance in the SS1, Far, CW, and CCW conditions once again revealed a significant main effect of target, $F(3,18)=16.98$, $p < .01$ Follow-up paired-samples t-tests revealed that SS1 and Far did not differ significantly, $t(6)=1.84$, $p > 0.10$, although mean directional errors were slightly more negative for Far (M = -2.69º) versus SS1. Additionally, mean errors to the CW and CCW targets differed substantially from mean errors in the SS1 condition, and were biased in opposite directions. Once again, estimates of the CW target were biased in a positive direction, whereas estimates of the CCW target were biased in a negative direction. This was confirmed in a paired-samples t-test comparing the CW and CCW conditions, $t(6)=4.91$, $p < .01$. Thus, results for this group of participants replicate the pattern of performance found in Experiment 3A, in keeping with the predictions of the DFT.

Average estimates of the standard deviation of the response distributions across each condition are shown in Figure 16B. As with the mean directional errors, an ANOVA revealed a significant main effect of target (SS1, Far, CW, CCW), $F(3,18)=22.50$, $p < .01$. As can be seen, responses were less variable (smaller SD) when a single item was held in memory, versus three items, in keeping with Zhang (2007). This was confirmed in a paired-samples t-test comparing SS1 to SS3, $t(6)=-8.79$, $p < .001$.

To test whether average resolution differed for close versus far targets in the SS3 condition, a separate ANOVA was conducted. This analysis revealed a significant main effect of target (Far, CW, CCW), $F(2,12)=4.97$, $p < .03$. Follow-up paired comparisons revealed that the SDs for Far and CW differed significantly, $t(6)=2.48$, $p < .05$, but no other comparison reached significance, all ps $> .07$. These results suggest that the resolution of items stored in WM varies depending on metric similarity, in contrast to the fixed-resolution hypothesis.

Finally, with respect to the parameter P_e (Figure 16C), which represents the probability that the probed item was encoded in memory, an ANOVA revealed a significant main effect of target (SS1, Far, CW, CCW), $F(3,18)=12.20$, $p < .01$. As with SD, clear differences emerged between the SS1 and SS3 conditions. Although encoding likelihood was rather high in each case, the probability that the probed item was encoded in memory was higher when one item was to be remembered. This was confirmed in a paired-samples t-test comparing SS1 to the average obtained P_e in the SS3 condition $t(6)=4.51$, $p < .01$.

An ANOVA was also run to determine if close and far targets varied in terms of P_e. This analysis revealed a significant main effect of target (Far, CW, CCW), $F(2,12)=5.46$, $p < .03$. Follow-up t-tests revealed that the Far target was less likely to be encoded than the CW target, $t(6)=-2.71$, $p < .04$, but not the CCW target, $t(6)=-2.15$, $p > .07$. P_e did not differ for CW and CCW targets, $t(6)=1.09$, $p > .3$.

The results of Experiment 3B replicate the primary findings of Experiment 3A, showing clear evidence of repulsion between nearby colors in VWM, and equivalent, small counterclockwise biases in the SS1 condition and for far colors in the SS3 condition. In keeping with the findings of Zhang (2007), resolution was found to be better in the SS1 versus SS3 condition, and P_e was lower when one versus three items needed to be encoded. Interestingly, P_e was also found to vary for close versus far targets in the SS3 condition, with a higher likelihood of encoding the close colors. This may

have reflected differential attention to close versus far targets due to the simultaneous presentation of the memory display elements. In addition, there was a significant difference in the precision of response estimates to the Far vs. the CW target. This is not consistent with the proposal that the resolution of items in VWM is fixed. Rather, this finding suggests that the precision of items can vary depending on their metric similarity. It is important to note, however, that the precision of response estimates to the Far and CCW targets did not differ significantly, suggesting that this similarity effect is quite subtle in nature.

Experiment 3B Simulations

The simulations reported here follow up on the simulations reported in Simulation Experiment 2. Those simulations accurately predicted that memory for close items in VWM would show a repulsion effect in the cued recall task. The question I probe here is whether the DFT can also account for other aspects of the data obtained in Experiment 3, such as the pattern of SDs and P_e. To determine this, I ran another series of simulations and obtained fits of the model data using the method developed by Zhang (2007). That is, I ran the model through large numbers of trials, generated response distributions for each condition, and then estimated parameters of these response distributions using the same method used with the empirical data. The question was whether I could quantitatively fit all parameters of the empirical distributions. Such fits would suggest that the DFT not only captures the shift in mean directional errors, but also captures effects related to the precision of VWM representations as well as the likelihood that such representations are successfully consolidated and maintained in VWM.

Methods

Model Architecture

The model architecture used here was identical to the model used in Simulation Experiment 2. However, several parameters were modified to provide precise quantitative fits to the empirical results (see Table 3 notes).

Simulation details

The simulations reported here were identical to those conducted in Simulation Experiment 2, with the exception that the model was run through 500 trials per condition, rather than 200.

Results and Discussion

Obtained fits to the simulation data are shown in Figure 17, and the full distribution of responses made by the model in each condition are shown in Figure 18 (right column). Obtaining model fits for the model data was achieved in several steps. First, the raw data from the simulations was inspected, and missing values where one or more peaks were lost (due to stochastic fluctuations in the field which can de-stabilize peaks) were replaced with a number between -180 and 180 drawn from a uniform random distribution. This was used as a stand-in for guessing trials, where participants have no information about the cued target in memory, and contributed to the estimation of P_e.

Second, simulations revealed that variance was relatively low in the model data (SDs were < 3° in all cases). Although SDs < 3° are typical for behavioral performance in spatial recall tasks (see, e.g., Spencer & Hund, 2002), these values are quite low relative to the empirical data from the color recall task. It is likely that the higher variance in the empirical data reflects both noisy VWM representations as well as noise introduced by the color estimation procedure where participants had to map a remembered color onto a randomly oriented color wheel.

To capture this second source of noise, which was not contained in the DFT, I implemented a procedure that added a random number to the model's estimation responses. Critically, the same random process was used across all conditions. The procedure involved taking the obtained estimation response on a given trial and adding to it a random number drawn from a Gaussian distribution with a given standard deviation. The Gaussian had a mean of zero, and a standard deviation was chosen that increased the overall variance into the range obtained in Experiment 3. Thus, the procedure was meant

to increase the overall variance in the model data, so that RAC fits could be obtained, without changing the relative amounts of variance across each condition.

Estimates of the mean directional errors from obtained fits of the model response distributions are shown in Figure 17A. Not surprisingly, the model captured the pattern of little to no systematic error in the SS1 and Far conditions, and substantial error in the Close condition, with CW and CCW targets moving in opposite directions across the delay.

RAC estimates of SDs can be seen in Figure 17B. The model produced small differences in SD for SS1 versus SS3 targets, in keeping with the findings of Zhang (2007) and Experiment 3B. However, the difference between SS1 and SS3 was much smaller than was seen in the data, and variance for the Far, CW, and CCW targets did not separate out to the same extent.

Finally, estimates of P_e from the model data can be seen in Figure 17C. The model accurately captured the finding of high P_e in SS1 and lower P_e for the close targets in the SS3 condition. However, the model substantially overestimated P_e for far targets. This was due to the fact that, in the SS3 simulations, when a peak was lost, it was always one of the two close peaks. This was caused by the same mechanism that produced the shift in the mean for these targets—strong lateral inhibition from a nearby target in WM. In the absence of strong competition from nearby targets, the far peak was never destabilized during the delay or the response interval.

In summary, the model was able to capture several aspects of the data reported in Experiment 3B. Mean directional errors followed the predicted pattern across each condition. Additionally, SDs were somewhat higher overall when three items versus one item was held in memory, and P_e was lower overall for targets in the SS3 condition. However, overall variation in obtained errors was much lower than was found in the experiments reported here, and differences in P_e for far versus close targets was not reproduced by the model.

There are several possible explanations for the model's divergence from the data. First, lower estimates of variance by the model were likely due in part to the procedure used to calculate errors. Recall that estimation errors in the model were determined by calculating the difference between the actual target location and the centroid of the peak in WM at the end of the simulation. As noted above, this procedure only accounts for variance arising as a result of differences in the position of the peak in WM across trials, and ignores variance that may arise at other points in the trial. For instance, the process of mapping the color in memory onto the color wheel may serve as an additional source of variation affecting responses. Given the relatively small size of the color wheel, small variations in the exact placement of the mouse cursor on the color wheel can lead to large increases in variable error across trials.

The model's underestimation of P_e may reflect another simplification implemented here. For these simulations, all three items were presented to the model simultaneously with equal strength on every trial. However, it seems reasonable to assume that variations in the amplitude of inputs across trials could produce cases where the far target is lost, either at the point of encoding, or shortly after a peak is established in WM, as seen in the variable amplitude simulations reported in Chapter 2. Additionally, variations in attention from trial-to-trial may also increase the likelihood of losing the far target on a given trial. This would occur if participants tend to focus more of their attention on the two similar items, in an effort to differentiate and accurately encode them in memory. If this is the case, one would expect that differences in P_e for close versus far targets would disappear in a sequential presentation variant of the task.

General Discussion

The results reported here provide strong support for the predictions of the DFT. In two separate experiments, I found clear evidence of metric interactions in WM leading to mutual repulsion between close colors. In contrast, single color estimates and estimates of a far color in the SS3 condition were not systematically biased. This suggests that items

in working memory interact as a function of how similar they are along the represented dimension. When items are very far apart along the represented dimension, they interact only weakly or not at all. However, when items are more similar, they interact through strong locally excitatory and laterally inhibitory connections. Such interactions can systematically distort the representations of remembered stimuli, leading to mutual repulsion in the context of recall, or improved performance in the context of change detection, as demonstrated in Chapter 3.

In addition to influencing the accuracy of memory for metric information, interactions among items in WM may also affect the resolution of WM representations. This was suggested by the finding of improved resolution (i.e., lower SDs) in the SS1 versus SS3 condition. These findings run counter to several recent proposals regarding the nature of storage in WM (see, eg., Raffone & Wolters, 2001; Zhang, 2007; Zhang & Luck, in press). For example, Zhang and Luck have proposed a model wherein VWM stores a small number of fixed resolution representations. Thus, the finding of improved resolution with one versus several items in memory would seem to run counter to this proposal. However, they argue that when VWM is not filled to capacity, multiple samples of the remembered item can be stored and averaged together when making an estimate, thereby reducing estimation variance. In Experiment 3B, however, I also found significant differences in SD across targets in the SS3 condition. Recall that the data fitting procedure used here allows variance to be estimated without being influenced by random guessing on trials where no information about the cued value was retained in memory. As a result, any differences in variance cannot be attributed to failures of encoding or maintenance. Thus, these results are clearly at odds with the proposal of Zhang and Luck.

In conclusion, the findings reported in the present chapter provide strong support for the predictions of the DFT. Although the model simulations of Experiment 3 did not capture all aspects of the obtained data with high fidelity, I was able to demonstrate that

metric dependent interactions among items in memory can lead to systematic changes in the nature of the representations stored. Such changes can include distortions in the feature values that are remembered, changes in the resolution of the stored representations, and changes in the likelihood that a given item will be maintained in memory throughout the delay.

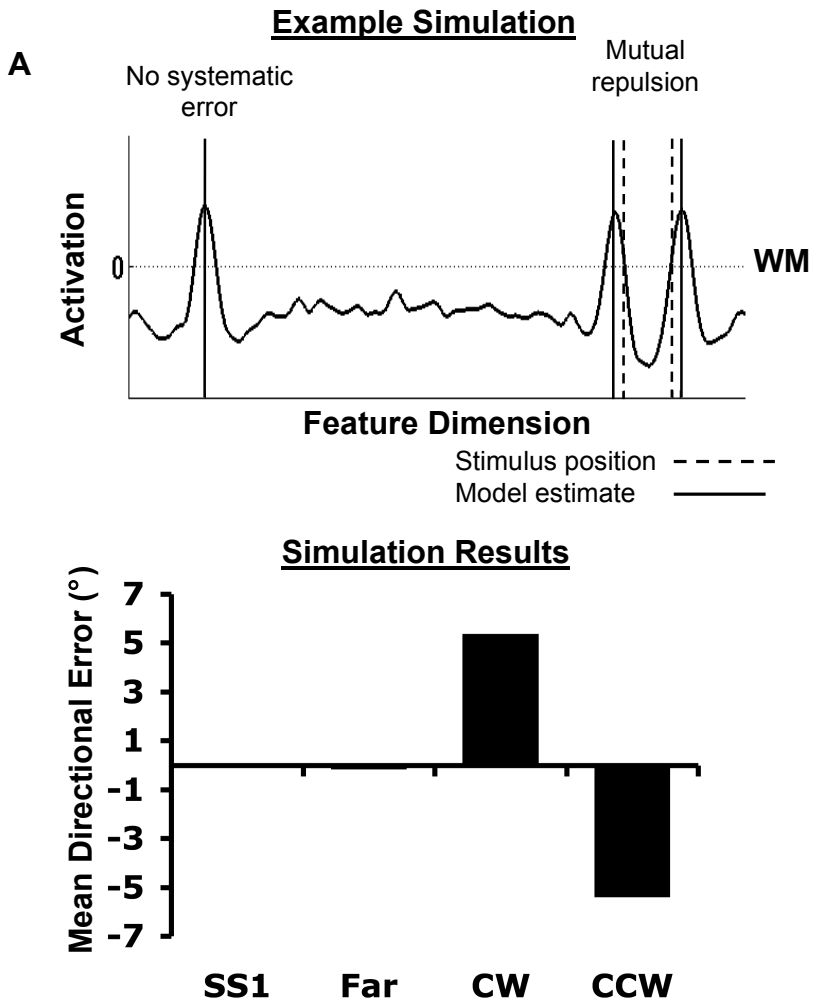

Figure 14. Example simulation and results from Simulation Experiment 2. (A) The model performing a single trial in the color estimation task. In this example, the model incorrectly estimates the two close colors to be further away than they really are, whereas estimation of the far color remains accurate. (B) Results of a full simulation experiment showing the predicted repulsion effect and no estimation errors in the SS1 and Far conditions.

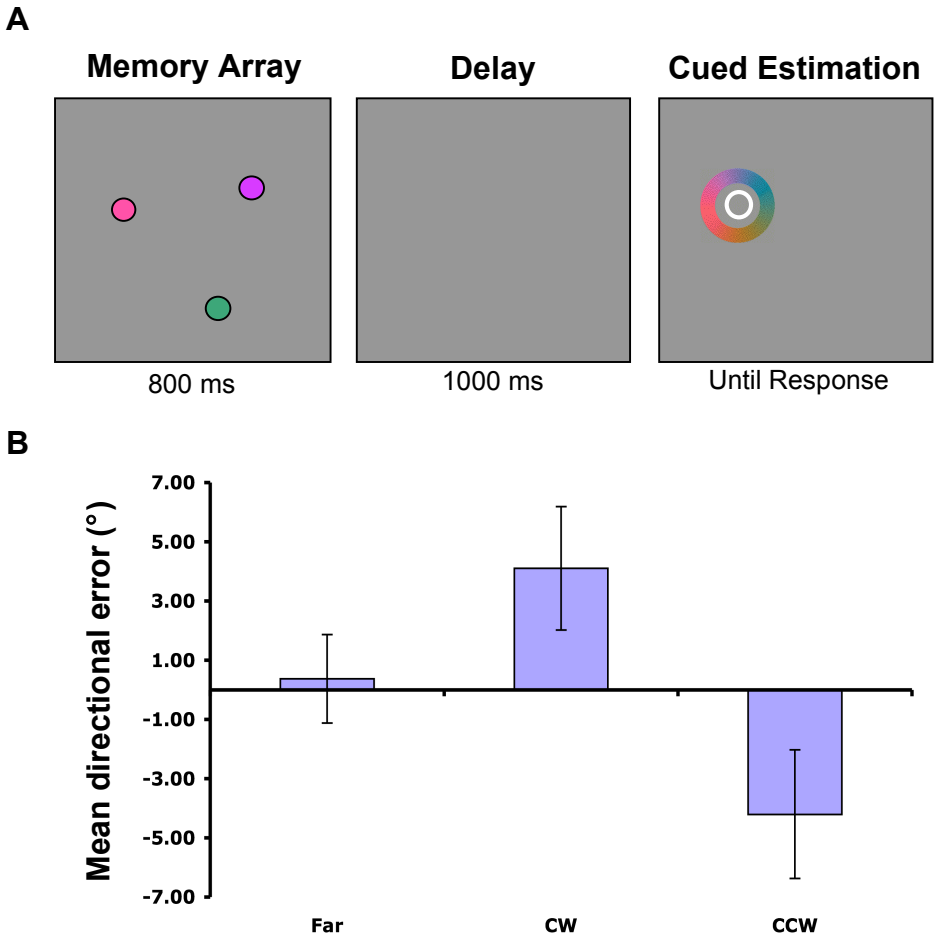

Figure 15. Stimuli, procedures and results for Experiment 3A.

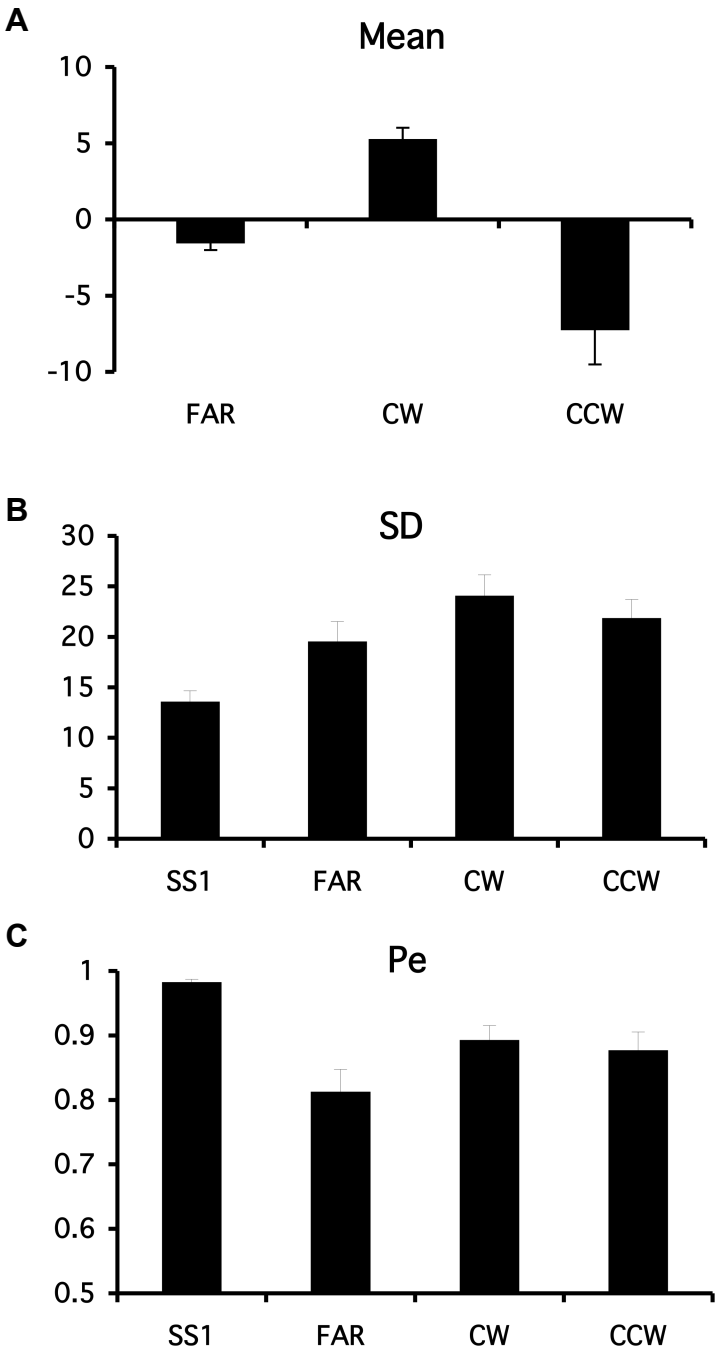

Figure 16. RAC estimates of the three parameters, Mean, SD, and P_e obtained in Experiment 3B.

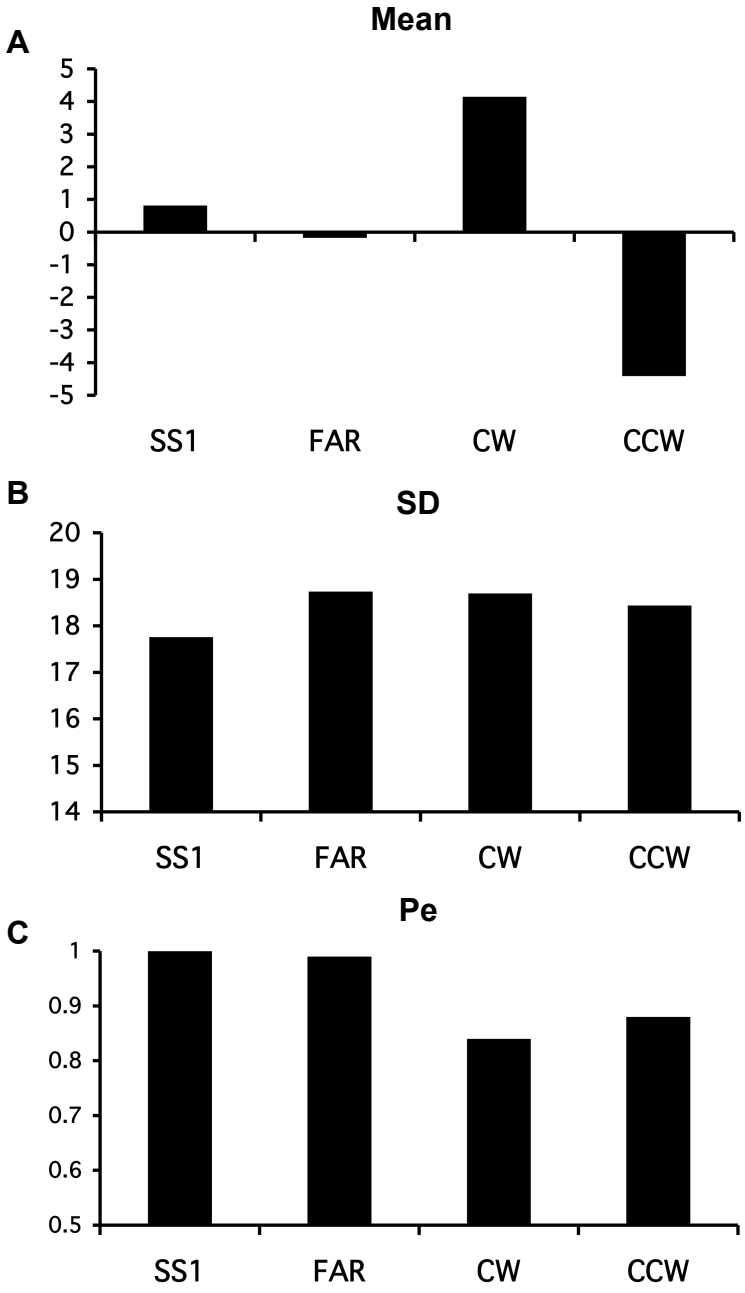

Figure 17. RAC estimates of the three parameters, Mean, SD, and P_e obtained in Experiment 3B simulations.

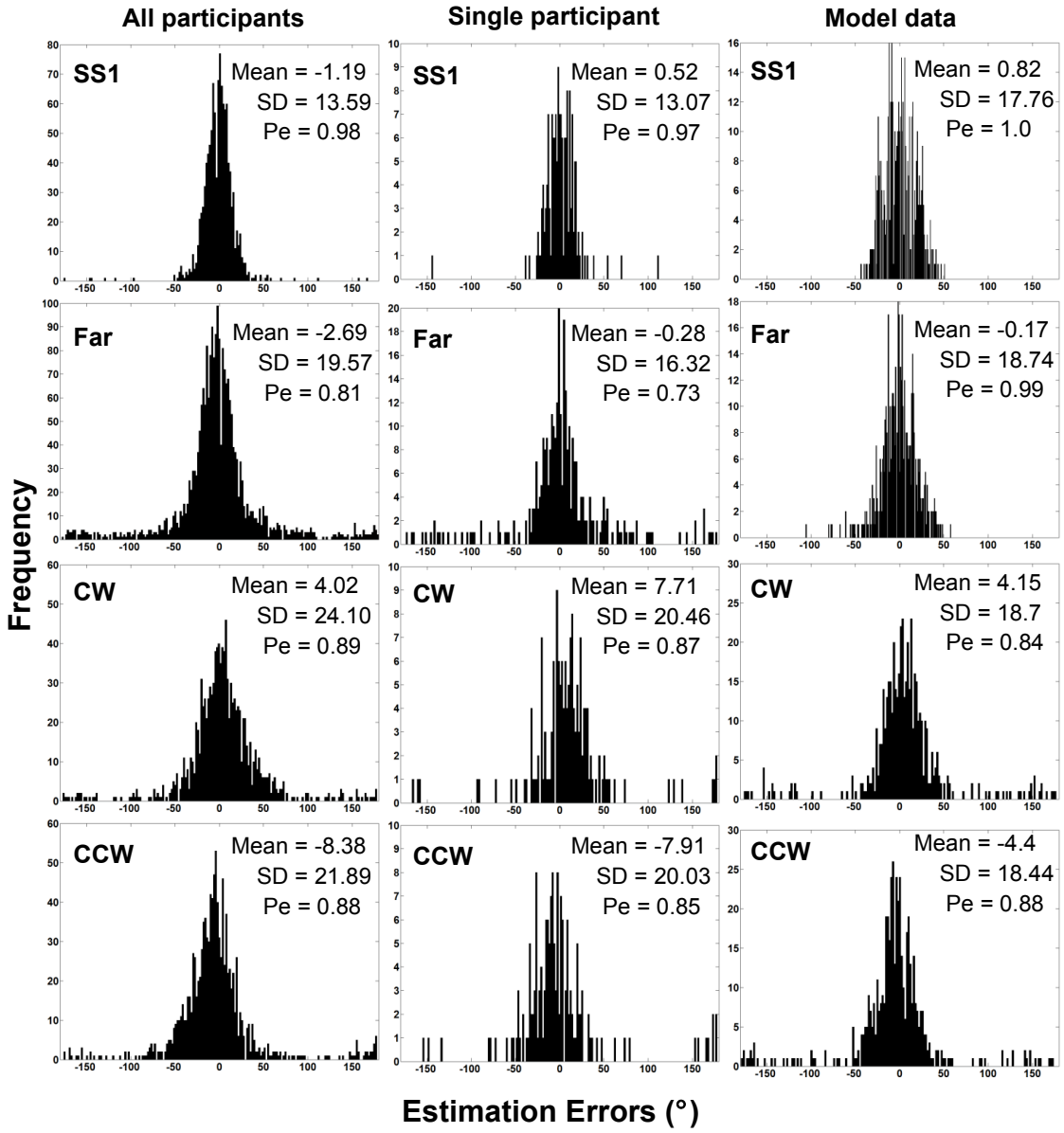

Figure 18. Frequency histograms of obtained errors in Experiment 3B and simulations. The leftmost column shows the distribution of responses across all participants showing repulsion in Experiment 3, the middle column shows distributions for a single subject from Experiment 3, and the rightmost column shows error distributions obtained from the model. Means, SDs, and P_e values are inset in each plot. In each case, errors in the SS1 and Far conditions are near zero. In contrast, estimates of the CW target are shifted slightly to the right (positive), whereas CCW target estimates are shifted to the left (negative).

Table 3

Parameter Values Used in Simulation Experiment 2

and Experiment 3B Simulations

Layer	τ	*h*	Self-excitation	Excitatory projections	Inhibitory projections	Target input	Noise
u (PF)	80	-7	$c_{uu} = 2.0$		$c_{uv} = 1.15$	$c_{tar} = 30$	$c_{noise} = .10$
			$\sigma_{uu} = 3$		$\sigma_{uv} = 24$	$\sigma_{tar} = 3$	$\sigma_{noise} = 1.5$
					$k_{uv} = .05$		
v (Inhib)	10	-12		$c_{vu} = 2.0$			$c_{noise} = .10$
				$\sigma_{vu} = 10$			$\sigma_{noise} = 1.5$
				$c_{vw} = 1.5$			
				$\sigma_{vw} = 13$			
w (WM)	80	-4	$c_{ww} = 3.15$	$c_{wu} = 1.5$	$c_{wv} = .325$	$c_{tar} = 6$	$c_{noise} = .10$
			$\sigma_{ww} = 5$	$\sigma_{wu} = 3$	$\sigma_{wv} = 28$	$\sigma_{tar} = 3$	$\sigma_{noise} = 1.5$
					$k_{wv} = .02$		

Note: For Experiment 3B simulations, c_{wv}=.20, k_{wv} = .06, c_{noise}= .08, and σ_{noise}= 2.75

CHAPTER 5
CONCLUSIONS AND GENERAL DISCUSSION

Summary and findings

The present project developed a new approach to visual working memory and change detection based on the principles of the Dynamic Field Theory that bridges the gap between neural and behavioral levels of theorizing. Although theories at the cognitive level have maintained a tight back and forth with behavioral research looking at VWM, such theories have not been formulated at the level of neural processes, and have not addressed the comparison process in change detection. Conversely, theories at the neural level have shed light on the biophysical and neurodynamical properties of the neural systems supporting maintenance in working memory, but have made little contact with the human behavioral literature looking at change detection. In addition, such approaches have focused primarily on single-item memory for spatial information, and have not addressed multi-item maintenance and the process of change detection in a single, integrated neural system. In the present work, I described a new neural process model of VWM and change detection that addresses these limitations. In addition, I reported a series of experiments that test novel predictions of the proposed model.

The basic functionality of the model was demonstrated in a series of exemplary simulations showing how the model can be used to capture different aspects of the change detection task. Recall that the first step in change detection involves the detection of items in the memory array and their successful consolidation in working memory. When inputs of sufficient strength and duration are presented to the model, the system goes through a *detection instability*, marked by an abrupt transition from a stable *sub-threshold state* of activation to an above-threshold state where localized peaks of activation can be formed and, in some cases, maintained once the input has been

removed. Whether peaks of activation are maintained in the absence of input depends on the balance of excitatory and inhibitory inputs to the field.

Excitatory recurrence in PF and laterally inhibitory input from Inhib can move PF into a *self-stabilized state* in response to input, where one or more peaks of activation are maintained as long as input is present. Once input is removed, however, PF quickly transitions back to the sub-threshold state, where activation at all field sites remains at baseline. In contrast, neurons in the WM layer, where excitatory recurrence is stronger, enter a *self-sustaining state* in response to input, where localized peaks of activation are maintained after the input is removed. Once a stable self-sustaining peak of activation is formed in the WM layer, the model can be said to have consolidated the information in WM.

The self-sustained state persists unless systematic or random perturbations (e.g., noise) destabilize the peak. One source of systematic perturbation of particular interest here is the case where new inputs are presented to the model some time after the original inputs are consolidated in WM. Under these circumstances, several functions emerge, depending on the nature of the new inputs. When the new inputs are the same as the original inputs, neurons in PF coding for those features fail to enter the self-stabilized state. The failure to build a peak in PF occurs as a result of strong inhibitory recurrence from Inhib, whose activation is driven by the presence of self-sustaining peaks in WM. Thus, in this case, the self-sustained state is maintained in WM, and PF remains in the sub-threshold state.

In contrast, when a new input that is different than the original is presented to the model, patterns of activation in the model are changed depending on the metric similarity of the new and old inputs. When the new inputs are only slightly different than the original inputs, no new peak is built in PF, but the peaks in WM may be updated in a continuous fashion to reflect the new values. In contrast, when a new input is presented that is sufficiently metrically different from the original inputs, PF transitions to the self-

stabilized state, signaling that a change has occurred, and working memory is updated to reflect the new value. If the new inputs are metrically similar to the original inputs, the original peak in WM is destabilized, and is replaced by the new value. However, when the new input is metrically very different than the original input, the new item may be added to the WM field without destabilizing the old peak.

Thus, the proposed model realizes each of the basic components needed to capture performance in change detection tasks, from encoding and consolidation, to maintenance, to comparison and updating in response to changed inputs. Importantly, the model shows how these different functions can arise within an integrated, dynamic neural system. In addition, the model makes novel and counterintuitive predictions that were tested in a series of behavioral experiments.

The first series of experiments used a change detection paradigm to test the prediction that metric-dependent interactions among nearby items in WM would lead to improved performance. This prediction arises as a result of laterally inhibitory interactions among nearby peaks in WM, which was shown to produce a sharpening of the range of local excitation associate with each peak. This, in turn, leads to a narrower and weaker inhibitory projection to PF. The consequence is that changes are easier to detect when one of the close items changes at test. This prediction was tested in three different experiments, two of which looked at WM for color, and one that looked at WM for orientation. In each case, the predicted enhancement effect was found. That is, performance in the change detection task was better for close versus far features.

In a second series of experiments, I tested another prediction arising as a result of close metrics—that laterally inhibitory interactions will lead to mutual repulsion between nearby peaks in WM. This arises as a result of the fact that inhibition is stronger in-between nearby peaks than it is on the "outer side" of each peak, which allows activation to grow more easily in a direction away from the other peak. Mutually repulsive interactions in the model produce a form of systematic bias in color memory that were

predicted to yield systematic distortions in color recall estimates for close colors. To test this possibility, I developed a cued color recall task where participants were shown one or more to-be-remembered colors, and after a short delay were asked to estimate one of the remembered colors by selecting it from a continuous color wheel. In the critical condition, participants were shown three colors, two of which were close in color and one that was far (i.e, unique). I predicted that estimates of the far color would not be systematically biased, and would not differ significantly from estimates of single colors. In contrast, I predicted that estimates of the two close colors would be systematically biased in opposite directions across the delay interval. This is exactly what was found. Thus, taken together, the obtained results suggest that metric-dependent interactions among items in VWM can have a measurable impact on performance in tasks used to probe working memory.

Theoretical implications of the proposed model and empirical findings

What are the implications of these findings for current models of VWM and change detection? Recall from my discussion of behavioral models, that there is currently some disagreement within the literature regarding the nature of storage in VWM. On one hand, Zhang and Luck (in press) have argued that working memory stores a small number of discrete, fixed-resolution representations. A corollary of this view is the idea that items in VWM do not interact. On the other, Alvarez and Cavanagh (2004) have argued that working memory consists of a pool of resources that can be allocated flexibly depending on the demands of the task. Thus, when presented with more than a few items, observers can store either a small number of high-resolution representations, or a larger number of low-resolution representations. Are the model simulations and data reported here consistent with either of these proposals?

The model and data are consistent with the proposal of Zhang & Luck (in press) in that working memory representations are realized in the model as individual localized

peaks of activation. In this sense, they are discrete entities, and the dynamical properties underlying their formation and maintenance could be thought of as the neural implementation of a slot-like form of working memory (see discussion in Luck, in press). Additionally, the simulations of Chapter 2 demonstrate clear capacity limits in the model. That is, on any given trial, the model builds and maintains a small number of peaks. When the number of items that need to be remembered increases beyond some point, one or more peaks fail to be maintained in WM.

Other aspects of the model and the data, however, are at odds with this view. For instance, in the model, peaks interact with one another in a metric-dependent fashion. Across multiple experiments, such interactions were shown to have a measurable impact on performance in tasks used to probe VWM. Thus, although storage may be thought of as more-or-less discrete in the limit where the items being maintained are highly distinct, this is clearly not the case when similar items are simultaneously maintained. These findings do not support the notion of independence among representations in working memory, as proposed by Zhang and Luck (in press). Moreover, results showing lower response variance when participants estimated the Far versus Close targets in Experiment 3B are not consistent with the proposal that slots in working memory have fixed resolution.

The neural field model and empirical findings are also consistent with some aspects of the resource model of Alvarez and Cavanagh (2004). For instance, similarity among the items stored was shown to have an impact on performance, and, importantly, on the resolution of WM representations in Experiment 3B. This is consistent with the proposal that WM representations may vary in resolution. However, the basic "unit of storage" in the model consists of localized peaks of activation that can be compared to the notion of a discrete "slot" in working memory. That is, WM as implemented here does not store *either* a small number of high-resolution representations *or* a large number

of lower resolution representations, but stores a small number of representations that may vary in resolution as a function of their similarity to the other items in memory.

At another level, the DFT is generally consistent with the view that working memory is an emergent property arising from interactions among neural populations (see, e.g., D'Esposito, 2005; Fuster, 2002; Postle, 2006). Additionally, the proposed model is consistent with neurodynamical accounts of maintenance in WM introduced earlier (e.g., Camperi & Wang, 2001). However, the present formulation moves to another level by providing a framework wherein such ideas can be directly tied to behavior in tasks used to explore WM. The basic strategy of this approach is to search for behavioral signatures of the neurodynamical processes proposed to underlie performance in such tasks. The metric-dependent sharpening and mutual repulsion among items in WM explored here represent two examples of this strategy. This close interplay between theoretical and empirical work makes it possible to connect a rigorous analysis of behavior with established neural principles.

Neural plausibility of the proposed model

I have claimed that the model proposed here represents a neurally-plausible approach to VWM and change detection. In what sense is the neural field model grounded in neural principles? First of all, there is a demonstrated link between a population dynamics approach to cortical activation and patterns of activation in neural fields, as well as clear methods that can be used to map single-unit recordings onto dynamic population representations that can be directly compared to dynamic field models. For instance, Bastian et al. (Bastian et al., 1998; Bastian, Schöner, & Riehle, 2003; Erlhagen et al., 1999) have used population coding techniques to compare single-unit neural activity in motor cortex to time-dependent changes in neural activation in a dynamic field model of motor planning. The first step in making this comparison was to map the responses of neurons in motor cortex to basic stimuli and create a continuous field by ordering the neurons based on their "preferred" stimulus. This was followed by a

behavioral precuing task that probed predictions of a dynamic field theory of movement preparation (Erlhagen & Schöner, 2002). A similar procedure was used in studies of neuronal interaction in the cat visual cortex (Jancke et al., 1999). In both cases, the reported results suggested a robust relationship between predictions of dynamic field models and neural measures.

Additionally, there is strong evidence from studies of cortical neurons for the basic local excitation / lateral inhibition form of neural interaction used in dynamic field models (Durstewitz et al., 2000). Moreover, because cortical neurons never project both excitatorily and inhibitorily onto targets, the inhibitory lateral interaction must be mediated through an ensemble of interneurons. We used a generic, two-layer formulation (Amari & Arbib, 1977) to realize this interaction where an inhibitory activation field receives input from an excitatory activation field and in turn inhibits that field.

Finally, the model presented here incorporates additional insights gained from studies of the layered structure of cortex. Specifically, the three-layered architecture was inspired by a cortical circuit model of the neocortex that was 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 the layers of the model do not map directly on to specific cortical layers, the basic structure and patterns of connectivity within the model are consistent with know principles of cortical organization.

Future directions

The work presented here extends the DFT of spatial cognition to address working memory for metric feature dimensions and the process of changed detection. An important future direction for this approach will involve linking the model described here back to the rich, embodied spatial system it grew out of. In the present work, working memory for metric features was realized using neural fields where activation was defined

along a single dimension (e.g., color). That is, the spatial locations of the targets were not included in the neural representation of the target features. Although the non-spatial nature of neural selectivity in the ventral stream is often highlighted, cortical feature maps in these areas are typically spatially organized in addition to being organized according to their preferred features (Op De Beeck & Vogels, 2000). Thus, a straightforward way to bring space into the picture sketched thus far is to simply add a spatial dimension to the model, forming a two-dimensional feature-space field. With this addition, localized peaks of activation within a field now represent both the feature value present in the environment and its spatial location. In the context of the model of change detection described above, this has the added benefit that when a change is detected at test, the resulting peak in PF signals the changed feature value as well as its spatial location (see Johnson, Spencer, & Schöner, in press). Such information could be used to direct spatial attention or the eyes to the location of the change, in keeping with behavioral and neural findings (Hyun et al., in press; Pessoa & Ungerleider, 2004).

In addition to representing information about particular feature-space pairings, simple distributed object representations can be achieved via spatial coupling among feature-space fields tuned to different features. In this case, features belonging to the same object enter into mutually-supportive interactions as a consequence of their shared location, and whole objects are represented as distributed patterns of activation across multiple linked feature maps. For example, the neural representation of a red, horizontally oriented bar would be represented by peaks of activation in color-space and orientation-space fields that are positioned at the same location along the spatial dimension of the field (Buss & Spencer, 2008; Johnson, Spencer et al., in press).

The use of feature-space fields to represent particular pairings of features and locations represents a neurally-plausible means of achieving simple distributed object representations. However, another property of spatial coding in the ventral stream can lead to problems when more than one multi-feature object needs to be represented. As

one progresses through the ventral stream, there is a steady, sometimes dramatic increase in receptive field sizes and an accompanying decrease in the spatial resolution of receptive fields for individual neurons (Desimone & Gross, 1979; Gross, Rocha-Miranda, & Bender, 1972). Additionally, although feature selectivity becomes more complex at higher levels of the visual system, the individual features of complex objects (e.g., color, form, size, and direction of motion) are coded in a distributed manner through the parallel activation of large numbers of neurons (Fujita, Tanaka, Ito, & Cheng, 1992; Komatsu & Ideura, 1993; Llinás & Paré, 1996). Although this type of encoding can be quite efficient, it can also lead to substantial problems when multiple objects are presented simultaneously, as in visual search experiments and most real-world visual tasks. For example, when two or more objects are presented simultaneously, the individual features making up the objects are coded in partially independent neural populations with overlapping spatial receptive fields. As a result, it can be difficult to determine which features belong together as attributes of a single object, an example of the *binding problem* in vision (Damasio, 1989; von der Malsburg, 1981, 1995).

One of the strongest sources of evidence suggesting that this is a real problem for the visual system is the finding that, under certain circumstances, the features of objects can be miscombined in normal perception. For example, in a series of experiments reported by Treisman and Schmidt (1982), participants were briefly shown multi-element displays of colored letters. In some cases, the colors and letters were perceived correctly but in the wrong combination. For example, a participant could be presented with a display containing a red horizontal line and a green vertical line, and incorrectly report seeing a red vertical and a green horizontal line (see also, Ashby, Prinzmetal, Ivry, & Maddox, 1996; Cohen & Ivry, 1989; Prinzmetal, 1981).

A likely cause of such failures is uncertainty about the spatial locations of the individual features making up objects. In the framework described here, this problem can be solved by coupling separate feature-space fields to a single spatial working memory

(SWM) field that maintains precise information about spatial location. The presence of additional spatial input serves to "pull" activation into the right position along the spatial dimension, keeping peaks aligned in each of the feature-space fields and clarifying which features belong to which objects (Johnson, Spencer et al., in press). The peaks of activation in SWM, then, indicate the number of items actually present in the display.

Conclusion

In conclusion, I contend that the three-layer neural field architecture described here provides a useful framework for thinking about how elementary perceptual and cognitive functions can emerge from the coordinated activity of an integrated, dynamic neural system. The proposed model captures each of the primary components required in simple visual comparison tasks such as change detection and is consistent with general principles of neural function. In addition, the proposed model makes novel and counterintuitive predictions that were tested in behavioral experiments. Thus, I was able to connect a rigorous analysis of behavior to established neural principles. Importantly, this thesis shows that we do not have to forge this link using the concepts of information processing. Rather than partitioning the continuous flow of time into discrete stages, I showed how a single system could capture behavioral performance *and* generate novel behavioral predictions that run counter to existing theories. This retains the fundamental strengths of cognitive psychology but within a dynamical systems framework.

An important future direction for this approach will be to integrate the framework described here with the robust, embodied theory of spatial cognition that it grew out of.

REFERENCES

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by total 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.

Amit, D. J. (1995). The hebbian paradigm reintegrated: local reverberations as internal representations *Behavioral and Brain Sciences, 18*, 617-626.

Amit, D. J., & Mongillo, G. (2003). Selective delay activity in the cortex: Phenomena and interpretation. *Cerebral Cortex, 13*, 1139-1150.

Andersen, R. A., Bracewell, R. M., Barash, S., Gnadt, J. W., & Fogassi, L. (1990). Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. *Journal of Neuroscience, 10*, 1176-1196.

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review, 103*, 165-192.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation, Vol. 2*. London: Academic Press.

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *The Bell System Technical Journal, 40*, 309-328.

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. H. Bower (Ed.), *The psychology of Learning and Motivation* (Vol. VIII, pp. 47-90). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory*. Cambridge, UK: Cambridge University Press.

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.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436.

Braun, M. (1994). *Differential equations and their applications* (4th ed.). New York: Springer Verlag.

Buerle, R. L. (1956). Properties of a mass of cells capable of regenerating pulses. *Transactions of the Royal Society of London B, 240*, 55-94.

Buss, A., & Spencer, J. P. (2008). The emergence of rule-use: A dynamic neural field model of the DCCS. *Paper to appear in the Twenty-ninth Annual Conference of the Cognitive Science Society*.

Camperi, M., & Wang, X.-J. (1998). A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *Journal of Computational Neuroscience, 5*, 383-405.

Cohen, A., & Ivry, R. B. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception & Performance, 15*, 650-662.

Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex, 10*, 910-923.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*(1), 75-84.

Constantinidis, C., & Steinmetz, M. A. (1996). Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *Journal of Neurophysiology, 76*(1352-1355).

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104*, 163-171.

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control*. Cambridge, UK: Cambridge University Press.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*, 761-772.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation, 1*, 123-132.

De Renzi, E., & Nichelli, P. (1975). Verbal and nonverbal short-term memory impairment following hemispheric damage. *Cortex, 11*, 341-354.

Desimone, R., & Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Research, 178*, 363-380.

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, 3*, 1184-1191.

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.

Farell, B. (1985). "same"-"different" judgements: A review of current controversies in perceptual comparison. *Psychological Bulletin, 98*, 419-456.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature, 360*, 343-346.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 61*, 331-349.

Fuster, J. M. (2003). *Cortex and Mind: Unifying Cognition*. Oxford: Oxford University Press.

Fuster, J. M., & Alexander, G. (1971). Neuron activity related to short-term memory. *Science, 173*, 652-654.

Fuster, J. M., & Jervey, J. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science, 212*, 952-955.

Gegenfurtner, K. R., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception & Performance, 19*, 845-866.

Georgopoulos, A. P. (1995). Motor cortex and cognitive processing. In M. Gazzaniga (Ed.), *The cogntive neurosciences*. Cambridge: MIT Press.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum (Ed.), *Handbook of Physiology: Nervous System* (Vol. 5, pp. 373-417). Bethesda: American Physiological Society.

Griffith, J. S. (1963). A field theory of neural nets. I. *Bulletin of Mathematical Biophysics, 25*, 111-120.

Griffith, J. S. (1965). A field theory of neural nets. II. *Bulletin of Mathematical Biophysics, 27*, 187-195.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of of neurons in inferotetmporal cortex of the macaque. *Journal of Neurophysiology, 35*, 96-111.

Grossberg, S. (1978). A theory of human memory: self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 5). New York: Academic Press.

Grossberg, S. (1980). Biological competition: Decision rules, pattern formation and oscillations. *Proceedings of the National Academy of Sciences, 77*, 2338-2342.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243-271.

Hollingworth, A., & Henderson, J. M. (2004). Sustained change blindness to incremental scene rotation: A dissociation between explicit change detection and visual memory. *Perception and Psychophysics, 66*, 800-807.

Howard, M. W., Rizzuto, D. S., Caplan, J. B., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., et al. (2003). Gamma oscillations correlate with working memory load in humans. *Cerebral Cortex, 13*, 1369-1374.

Hyun, J.-S. (2006). *How are visual working memory representations compared with perceptual inputs?* , University of Iowa, Iowa City, Iowa.

Hyun, J.-S., Woodman, G. F., Vogel, E. K., Hollingworth, A., & Luck, S. J. (in press). The comparison of visual working memory representations with perceptual inputs. *Journal of Experimental Psychology: Human Perception & Performance*.

Irwin, D. E. (1992). Visual memory within and across fixations. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading.* (pp. 146-165). New York: Springer-Verlag.

Irwin, D. E. (1993). Perceiving an integrated visual world. In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance 14: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 121-142). Cambridge, MA: MIT Press.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics, 64*, 882-895.

James, W. (1890). *The principles of psychology*. New York: Holt.

Jancke, D., Erlhagen, W., Dinse, H. R., Akhavan, A. C., Giese, M. A., 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.

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

Johnson, J. S., Spencer, J. P., & Schöner, G. (in press). Moving to higher ground: The dynamic field theory and the dynamics of visual cognition. In F. Garzón, A. Laakso & T. Gomila (Eds.), *Dynamics and Psychology [special issue]. New Ideas in Psychology*.

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge: MIT Press.

Kelso, J. A. S., Scholz, J. P., & Schöner, G. (1986). Non-equilibrium phase transitions in coordinated biological motion: critical fluctuations. *Physics Letters A, 118*, 279-284.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning & Verbal Behavior, 1*(3), 153-161.

Komatsu, H., & Ideura, Y. (1993). Relationship between color, shape, and pattern selectivities in the inferior cortex of the monkey. *Journal of Neurophysiology, 73*(218-226).

Komatsu, H., Ideura, Y., Kaji, S., & Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *Journal of Neuroscience, 12*, 408-424.

Lin, P.-H., & Luck, S. J. (in press). The influence of similarity on visual working memory representations. *Visual Cognition*.

Llinás, R., & Paré, D. (1996). The brain as a closed system modulated by the senses. In R. Llinás & P. S. Churchland (Eds.), *The mind-brain continuum*. Cambridge: MIT Press.

Logie, R. H., & Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial or central executive? *Advances in Psychology, 80*, 105-115.

Luck, S. J. (in press). Visual short-term memory. In S. J. Luck & A. Hollingworth (Eds.), *Visual Memory*. New York: Oxford University Press.

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

Machens, C. K., Romo, R., & Brody, C. D. (2005). Flexible control of mutual inhibition: A neural model of two-interval discrimination. *Science, 3307*, 1121-1124.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in the prefrontal cortex of the macaque. *Journal of Neuroscience, 16*, 5154-5167.

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience, 13*, 1460-1478.

Miller, G. A. (1960). *Plans and the Structure of Behavior*. New York: Holt, Rinehart, & Winston.

Miller, P., & Wang, X.-J. (2006). Inhibitory control by an integral feedback signal in prefrontal cortex: A model of discrimination between sequential stimuli. *Proceedings of the National Academy of Sciences, 103*, 201-206.

Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Computational Neurology, 426*, 505-518.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Pessoa, L., Gutierrez, E., Bandettini, P. B., & Ungerleider, L. G. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron, 35*, 975-987.

Pessoa, L., & Ungerleider, L. G. (2004). Neural correlates of change detection and change blindness in a working memory task. *Cerebral Cortex, 14*, 511-520.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics, 16*, 283-290.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 130*, 23-38.

Potter, M. C. (1976). Short-term conceptual memory for pictcures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 509-522.

Prinzmetal, W. (1981). Principles of feature integration in visual perception. *Perception & Psychophysics, 30*, 330-340.

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

Rao, S. G., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276*, 821-824.

Rao, S. G., Williams, G. V., & Goldman-Rakic, P. S. (1999). Isodirectionl tuning of adjacent interneurons and pyramidal cells during working memory: Evidence for microcolumnar organization in PFC. *Journal of Neurophysiology, 81*, 1903-1916.

Romo, R., Brody, C. D., Hernández, A., & Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature, 399*, 470-473.

Salinas, E., Hernández, A., Zainos, A., & Romo, R. (2000). Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. *Journal of Neuroscience, 20*, 5503-5515.

Scarborough, D. L. (1972). Memory for brief visual displays of symbols. *Cognitive Psychology, 3*, 408-429.

Schmidt, B. K., Vogel, E. K., Woodman, G. F., & Luck, S. J. (2002). Voluntary and automatic attentional control of visual working memory. *Perception & Psychophysics, 64*, 754-763.

Schöner, G., & Kelso, J. A. S. (1988a). A dynamic theory of behavioral change. *Journal of Theoretical Biology, 135*, 501-524.

Schöner, G., & Kelso, J. A. S. (1988b). Dynamic pattern generation in behavioral and neural systems. *Science, 239*, 1513-1520.

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*(5), 1393-1417.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5*, 195-200.

Shadlen, M. N. (2002). Rate versus temporal coding of information in the cerebral cortex. In *Encyclopedia of Cognitive Science*: Macmillan.

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron, 24*, 67-77.

Shibuya, H., & Bundeson, C. (1988). Visual selection from multielement displays: Measure and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception & Performance, 14*, 591-600.

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 section]. Brain Research* (Vol. 1202, pp. 68-86).

Simmering, V. R., Spencer, J. P., & Schöner, G. (2006). Reference-related inhibition produces enhanced position discrimination and fast repulsion near axes of symmetry. *Perception & Psychophysics, 68*, 1027-1046.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*, 261-267.

Spencer, J. P., Clearfield, M., Corbetta, D., Ulrich, B. D., Buchanan, P., & Schöner, G. (2006). Moving toward a grand theory of development: In memory of Esther Thelen. . *Child Development, 77*, 1521-1538.

Spencer, J. P., & Hund, A. M. (2002). Prototypes and particulars: Spatial categories are formed using geometric and experience-dependent information *Journal of Experimental Psychology: General, 131*, 16-37.

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamical systems approach to development. *Developmental Science, 6*, 392-412.

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.), *Emerging landscapes of mind: Mapping the nature of change in spatial cognitive development.* New York, NY: Oxford University Press.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*, (Whole No. 498).

Tallon-Baudry, C., Bertrand, O., Peronnet, F., & Pernier, J. (1998). Induced γ-band activity during the delay of a short-term memory task in humans. *Journal of Neuroscience, 18*, 4244-4254.

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

Thelen, E., & Ulrich, B. D. (1991). Hidden skills: a dynamic systems analysis of treadmill stepping during the first year. *Monographs of the Society for Research in Child Development, 56*, 1-104.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*, 751-754.

Trappenberg, T. P. (2003). Why is our capacity of working memory so large? *Neural Information Processing-Letters and Reviews, 1*, 97-101.

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*, 107-141.

Vogel, E. K., & Machizawa, M. G. (2005). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*, 748-751.

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.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance, 32*, 1436-1451.

von der Malsburg, C. (1981). The correlation theory of brain function. *MPI Biophysical Chemistry, Internal Report 81-2*.

von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology, 5*, 520-526.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neuroscience, 24*, 455-463.

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

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal, 12*, 1-24.

Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin & Review, 11*, 269-274.

Woodman, G. F., & Vogel, E. K. (2005). Fractionating working memory: Consolidation and maintenance are independent processes. *Psychological Science, 16*, 106-113.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science, 12*, 219-224.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440*, 91-95.

Young, E. D. (1998). What's the best sound? *Science, 29*, 1402-1403.

Zhang, W. (2007). *Resolution and capacity limitations in visual working memory: A new approach.* University of Iowa, Iowa City.

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