UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Serial Dependence in Visual Working Memory: Time Course and Neural Mechanisms

Permalink
https://escholarship.org/uc/item/1bf6q544

Author
Bliss, Daniel

Publication Date
2017

Peer reviewed|Thesis/dissertation

Serial Dependence in Visual Working Memory: Time Course and Neural Mechanisms

by

Daniel P. Bliss

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Neuroscience

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mark D'Esposito, Chair
Professor David Whitney
Professor Michael A. Silver
Professor William Prinzmetal

Summer 2017

Serial Dependence in Visual Working Memory: Time Course and Neural Mechanisms

Abstract

Serial Dependence in Visual Working Memory: Time Course and Neural Mechanisms

by

Daniel P. Bliss

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Mark D'Esposito, Chair

Visual cognition applies temporal smoothing to its inputs, which creates a serial dependence between successive representations at the focus of attention. This is thought to promote perceptual stability. While the benefits of serial dependence have been assumed, evidence that perception itself is altered has been limited. In the first chapter of this dissertation, I vary the delay between stimulus and response in a spatial delayed response task to investigate whether serial dependence occurs at the time of perception or later in working memory. I find that behavioral responses made immediately after viewing a stimulus are on average veridical. Only as memory demands increase is a blending of past and present information apparent in behavior, reaching its maximum with a memory delay of six seconds.

In the second chapter, I explore potential neural-circuit mechanisms of serial dependence. I consider two possible substrates of the effect: stable persistent activity during the memory delay and dynamic "activity-silent" synaptic plasticity. I find that networks endowed with both strong reverberation to support persistent activity and dynamic synapses can closely reproduce behavioral serial dependence. Specifically, elevated activity drives synaptic augmentation, which biases activity on the upcoming trial, giving rise to a spatiotemporally tuned shift in the population response. My hybrid neural model is a theoretical advance beyond abstract mathematical characterizations of working memory and demonstrates the power of biological insights to provide a quantitative explanation of human behavior.

The model developed in Chapter 2 proposes that serial dependence is due in part to synaptic augmentation, which is especially prominent in prefrontal cortex. In the third chapter, I investigate whether the bias in behavior depends on activity in three separate nodes of prefrontal cortex (PFC) – the frontal eye fields, the dorsolateral PFC, and the anterior PFC near the frontal pole. I find that transcranial magnetic stimulation (TMS) to these nodes causes reductions in serial dependence consistent with the model's predictions. In contrast, TMS to posterior sites – either primary somatosensory cortex or posterior parietal cortex – fails to alter the magnitude of the behavioral effect. This general result holds across TMS protocols (online vs. offline) and tasks with different stimulus and response types.

Acknowledgments

The members of my dissertation committee – Mark D'Esposito, David Whitney, Michael Silver, and Bill Prinzmetal – gave me the combination of freedom and support I needed to see this project to its conclusion. I am grateful to all of them. Special thanks go to Mark, my advisor, whose mentorship for seven years (starting before I applied to graduate school, when I was a tech in his lab) has helped me grow independence as a scientist. David deserves extra thanks as well. My dissertation would be without its moorings were it not for the Whitney Lab's pioneering research on serial dependence. David was generous to champion my efforts even when my interpretation of this phenomenon differed from his.

Chapter 1

Serial dependence is absent at the time of perception but increases in visual working memory

Introduction

Even in contexts where visual input varies randomly from trial to trial, human observers tend to blend stimuli from previous trials into their representation of the current one, leading to a bias in behavioral reports[1–14]. This smoothing of representations – termed "serial dependence" – is a function of how close successive stimuli are in space[1, 15] and time[1–7, 12, 14, 15]. It is also sensitive to their featural similarity[1–3, 7–10, 12, 14, 16, 17]. Serial dependence has been observed in judgments of orientation[1, 8, 9] and location[16, 17], as well as more complex attributes like the identity[2] and attractiveness[3, 5, 11] of human faces. That the bias is observed for such disparate features suggests it may be a universal principle of visual processing, and recent work has sought to demonstrate its adaptiveness[8]: In natural environments – where the input to our eyes is generally very similar from moment to moment[18] – temporal smoothing would be expected to stabilize perception in the face of noise and occlusion[1, 8].

While the benefits of perceptual stability seem obvious, it is important to note that serial dependence impedes another goal of perception, which is to be sensitive to change. A classic example of how visual perception prioritizes change detection is the tilt after-effect[19]. This illusion (also termed adaptation[20]) is the quantitative opposite of serial dependence: Perception of the current moment is repelled away from, rather than merged with, recently processed stimuli – exaggerating differences. Like serial dependence, adaptation spans stimulus features[20–24]. However, unlike adaptation, the attractive bias depends on attention: the observer must attend to each stimulus for serial dependence to occur[1]. Attention is thought to rely on the same neural and psychological mechanisms as working memory[25–31]. Hence, it is possible that whereas adaptation is a phenomenon of visual perception[20–24]

serial dependence arises instead from post-perceptual visual working memory[9, 32]. If this were true, stability would operate in parallel with (rather than compete against) change detection, as these functions would be relegated to distinct cognitive systems[9].

Preliminary efforts have been made to resolve whether serial dependence is perceptual or mnemonic in nature, with mixed results[1, 9]. Using a comparison task that minimizes memory demands, one group identified positive serial dependence in a small number of individuals[1] – in favor of the perceptual account. However, an attempt to replicate this effect with a larger sample size only revealed repulsive adaptation[9]. That is, no attractive serial dependence was observed when memory demands were removed from the same comparison task in the follow-up study, in support of the idea that serial dependence requires working memory[9]. A complementary strategy for clarifying this issue has been to boost memory demands – by increasing the delay between stimulus and response in delayed-estimation tasks – to determine whether this potentiates the attractive bias[9, 32]. Traditionally, errors that scale with delay length are interpreted as mnemonic in origin, whereas those that are constant over time are assumed to be tied to the perceptual or motor demands that are also fixed[33]. Over a limited range, the magnitude of the serial dependence effect increases the longer that working memory is active[9, 16, 17]. Yet, despite this potential connection to working memory, serial dependence has yet to be incorporated into the many mathematical models of memory storage that have been developed in recent years[34–43].

In the present study, I investigate temporal smoothing in visual cognition over a wider range of memory delays than has been used in the past. I use a spatial delayed response task, which has been shown to produce serial dependence in non-human primates[16, 17]. Previous experiments using delayed response tasks to measure serial dependence have included a visual mask after the stimulus presentation period[1, 2, 8, 9], as well as a delay period of at least several hundred milliseconds before a response is permitted[1, 2, 8, 9, 16, 17], which encourages encoding into working memory and cannot cleanly measure more fragile perceptual representations[44, 45]. In my shortest delay condition, I allow participants to respond immediately after stimulus offset, with no mask. From this 0-s baseline, I parametrically increase the delay length up to 10 s. In a separate experiment, I parametrically manipulate the length of the inter-trial interval (ITI), which has never been done before in the study of serial dependence. This permits us to assess the decay rate of the trial-history effect in the absence of intervening trials – clarifying its potential functional and biological implementation. Finally, I pursue a novel formal unification of the serial dependence phenomenon with mathematical models of working memory[34, 35, 37–41]. This sets the stage for future experiments to dissect the neural mechanisms of serial dependence in the context of ongoing research into the organization of the working memory system[32].

Results

Experiment 1: Manipulation of visual working memory delay

Forty one adult participants (16 male, 25 female) completed a spatial delayed response task depicted in Figure 3.1 (adapted from [46]). Each trial began with the presentation of a circle at a random angle from fixation (with eccentricity fixed across trials). Participants were instructed to remember the location of this circle across a delay period that varied randomly from trial to trial (0, 1, 3, 6, or 10 s). At the end of the delay period, participants reported the location of the cue using a computer mouse, whose pointer appeared at the center of the screen after the disappearance of the fixation square. Error was measured in degrees of polar angle. Three participants were excluded due to poor performance (mean absolute error greater than 10° polar angle). Each of these participants completed 1,000 trials (200 per delay) divided into 40 blocks over the course of one or two experimental sessions.

Studies that have modeled the tuning of serial dependence to featural differences between past and current visual stimuli have used either the Gabor function [16, 17] or the derivative of Gaussian [1, 2, 8, 9]. However, neither of these functions accounts for the negative tails of serial dependence – the observation that when previous visual input is close to maximally different from the input on the current trial, repulsion rather than attraction occurs[9]. A model developed by Clifford and colleagues to describe the tuning of adaptation[20] does account for this reversal of the response bias and provides an explicit functional explanation for it. (In practice, the Clifford model applies to serial dependence as well as it does to adaptation, though it has only ever been employed in the study of the latter.) Within the Clifford model, the neural/psychological representational space into which visual stimuli are encoded is subjected to re-centering and scaling as a function of recent experience. This warping serves (1) to recalibrate the sensitivity of the visual system to match fluctuations in the environment (re-centering) and (2) to decorrelate neural responses across trials (scaling)[20]. The attractive bias of serial dependence arises as the combined consequence of re-centering and scaling, whereas only scaling contributes to the repulsive bias.

Collapsing across the five delays tested in this experiment, I identified serial dependence in the group dataset significantly greater than zero ($p < 10^{-4}$, group permutation test). To do this analysis, I measured the magnitude of serial dependence as the peak-to-peak of the Clifford model fit to behavioral responses. The peak-to-peak is a measure of the maximal pull of responses away from the correct stimulus feature value as a result of this trial-history bias. Previous studies have used similar measures of amplitude to quantify serial dependence[1, 2, 8, 9, 16]. No bias is present in the data in the direction of the stimulus on the upcoming trial (*n.s.*, group permutation test), which supports the conclusion that the dependence of behavior on the previous trial is not due to spurious correlations in the particular randomized sequences of stimuli generated for the subjects[10, 14].

Next I examined each of the delays individually. The magnitude of serial dependence across memory delays from 0-10 s is plotted in Figure 3.2A. When participants reported the location of the stimulus immediately after viewing it, presumably relying at least in part on

residual neural activity associated with perception, their responses bore no relationship to the previous trial's stimulus (*n.s.*, group permutation test, Bonferroni-corrected $\alpha = 0.01$; Fig. 3.2B). In contrast, for every other delay tested, serial dependence was significantly greater than zero (all $p < 10^{-3}$, group permutation tests). Moreover, the magnitude of serial dependence increased from 0-3 s ($p < 10^{-3}$, group permutation test, Bonferroni-corrected $\alpha = 0.005$) and again from 3-6 s ($p < 10^{-3}$) before asymptoting between 6 and 10 s (*n.s.*). Serial dependence was numerically strongest in the 6-s condition, shown in Figure 3.2C. Here, the peak-to-peak (the distance along the y-axis between the maximal and minimal values of the function fit) is 3.53°. Note that the asymptote in serial dependence between 6 and 10 s does not correspond to an asymptote in the accumulation of noise in working memory. Consistent with previous reports[33, 47] and computational theory[48], I observed a continued linear increase in the variance of responses up to a 10 s delay (Fig. 3.3).

This observed time course of serial dependence over the current trial's delay period was not observed when trials were sorted based on the previous trial's delay period. For each individual preceding delay period, serial dependence in the current trial's response was significantly greater than zero (all $p < 0.01$, group permutation tests, Bonferroni-corrected $\alpha = 0.01$). However, no delay length in the preceding trial potentiated serial dependence relative to the other delay lengths tested (all comparisons *n.s.*, group permutation tests, Bonferroni-corrected $\alpha = 0.005$). This finding is consistent with results from another study that tested a narrower range of delays in non-human primates[16]. I conclude that if a subject attends to a stimulus in the previous trial (or, perhaps equivalently, the stimulus in the previous trial is encoded into working memory), the duration of its maintenance in working memory does not influence the subsequent trial. However, the delay period on this subsequent trial is the period over which residual information from the preceding trial interacts with the newly encoded information. The length of this period does influence the magnitude of the serial dependence effect.

In line with past work on the sources of error in working memory[9, 33], my findings suggest that the increasing serial dependence over longer delays I observed reflects its association with mnemonic processes rather than perceptual and motor processes that were held constant in my experiment. Research over the last decade has yielded several mathematical models designed to isolate distinct sources of error in working memory[34, 35, 37–41]. None of these include parameters for the proactive interference that serial dependence represents[32]. Also, the only substantive difference between the models is their characterization of noise in the distribution of behavioral responses. As a form of systematic error, serial dependence is separable from noise, and so can be incorporated into any of these models without changing their definitions or differences. The simplest model (sometimes called the "equal precision" model[37, 39]) fits random error with a von Mises distribution [34]. In contrast, the "variable precision" model assumes the standard deviation parameter of the von Mises varies from trial to trial according to a gamma distribution[37, 38]. A third model explicitly regards the precision of working memory as arising from noise in Poisson-distributed spike trains of individual neurons[40, 41]. Errors in this model are distributed according to a von Mises random walk[41]. I will refer to these three working memory models as EP (equal

precision), VP (variable precision), and VMRW (von Mises random walk).

As a first pass, I fit each of these models to the behavioral data from Experiment 1. Model comparison on the basis of the corrected Akaike Information Criterion (AICc) revealed that the VMRW model fit the data better than the VP model (ΔAICc = 826), which fit the data better than the EP model (ΔAICc = 700). This finding is consistent with published comparisons of these models using behavioral data from other working memory tasks[37–40]. Next, I created a hybrid model that uses the Clifford model of serial dependence to set the mean of the VMRW distribution on each trial. This hybrid model significantly outperforms the base VMRW model (ΔAICc = 902). An alternative hybrid "memory confusion"[2] model – in which, on a subset of trials, subjects mix up which stimulus belongs to the current trial and report the previous trial's location when probed (analogous to a "swap" [36, 43] over time rather than space) – provides no benefit above the base VMRW model and makes it worse, due to the addition of parameters that capture little variance (ΔAICc = −291). The addition of the Clifford model to all three of the base models does not change the order of performance among them: the improved VMRW model outperforms the improved VP model (ΔAICc = 457), which outperforms the improved VM model (ΔAICc = 551).

Experiment 2: Manipulation of baseline interval between trials

It is possible that the delay manipulation in Experiment 1 confounded two variables: (1) the time for which subjects must hold the current item in memory and (2) the time that has elapsed since the behavioral response on the previous trial, before the current trial's response. To demonstrate that the time course of serial dependence I observed (Fig. 3.2A) corresponds to mnemonic processes and not the simple passage of time, I conducted a second experiment in which the inter-trial interval (ITI) varied randomly among 1, 3, 6, and 10 s. The delay in this new task was held constant at 3 s. In all other respects, the tasks for the two experiments were identical. Twenty participants (8 male, 12 female), 6 of whom also completed Experiment 1 (and all of whom were recruited from the same subject pool), completed Experiment 2. Two participants were excluded due to poor performance (mean absolute error greater than 10° polar angle), which left data from 18 participants (7 male, 6 from Experiment 1) for analysis. All but two of these participants completed 1,008 trials (252 per ITI). The remaining two participants completed 999 and 1,017 trials, respectively.

Collapsing across ITIs, I identified serial dependence in the group dataset significantly greater than zero ($p < 10^{-4}$, group permutation test). As for Experiment 1, there was no bias in the data in the direction of the stimulus on the upcoming trial (*n.s.*, group permutation test), an important control[10, 14]. This pair of results replicates my finding from Experiment 1 of serial dependence in this spatial delayed response task, using an independent dataset.

Next I examined each of the ITIs individually. The magnitude of serial dependence across ITIs from 1-10 s is plotted in Figure 3.4A. The magnitude of serial dependence appears to decrease gradually during the interval between trials, decreasing from 3-6 s ($p < 10^{-3}$, group permutation test, Bonferroni-corrected $\alpha = 0.005$) and again from 6-10 s ($p < 10^{-4}$). The difference in serial dependence between the 1-s (Fig. 3.4B) and 3-s ITIs is statistically

non-significant. The slope of this time course is opposite that obtained in Experiment 1, strengthening my conclusion that increased serial dependence with increased delay length is due to the prolongation of memory demands rather than the mere passage of time. With the largest ITI (10 s) participants' responses on the trial after the ITI bear no relationship to the preceding trial's stimulus (*n.s.*, group permutation test, Bonferroni-corrected $\alpha = 0.01$; Fig. 3.4C). In contrast, for every other ITI tested, serial dependence is significantly greater than zero (all $p < 10^{-3}$, group permutation tests).

Discussion

In everyday visual experience, humans rely not just on moment-to-moment perception but also on continued maintenance of information in working memory to navigate their environments and accomplish tasks. While there is much evidence to suggest that working memory recruits the same cortical areas active during sensory perception[49–57], remembered visual content differs in quality[33, 47, 58] – and potentially representational format[59, 60] – from feedforward signals driven by the presence of an external stimulus. Both behavioral data[33, 47, 58] and computational theory[48] have implied that passage of visual percepts into memory makes them less precise. This past work has also claimed that mnemonic processes do not attach to percepts any accumulating systematic bias – just random noise due to drift and/or decay[33, 47, 48, 58]. With the experiments reported here, I provide new evidence to disconfirm this view. Serial dependence – a systematic bias in the direction of the preceding trial's stimulus – is absent from percepts until the working memory system is engaged. This null result does not require a perceptual comparison between two simultaneously presented stimuli[9], but occurs in the context of the same delayed response task that yields serial dependence when memory demands are increased.

By testing a wider range of delays between stimulus and response than used in previous studies[9, 16, 17], I was able to chart the time course of serial dependence in visual working memory. This technique – of probing participants to report the contents of memory at variable time points after stimulus offset – is common in visual psychophysics[61]. It has revealed how information passing through the visual system progresses from a rich perceptual code to a more impoverished mnemonic one. For a few hundred milliseconds after visual input ceases, a great deal of perceptual detail is still accessible to the observer in iconic memory – a form of storage intermediate between perception and working memory[62]. After that, within one second of delay, capacity-limited, distraction-resistant working memory comes online in parallel with a larger-capacity system that is vulnerable to distraction – fragile memory[44, 45, 63–66]. My experiments demonstrate that the residual sensory trace associated with iconic memory is free of serial dependence. The bias arises slowly in the later shorter-term memory systems, but asymptotes before longer-term storage processes are engaged (at approximately 20 seconds of delay[47]). Future research may resolve with finer resolution the exact moment at which serial dependence appears and whether it is most strongly associated with fragile stages of working memory. (Consistent with most work in this area[67, 68], I

have tended to use the term "working memory" as a shorthand for both of these systems.)

The Clifford model[20] I used to measure serial dependence is equally capable of detecting adaptation. Given that adaptation is commonly observed in perceptual judgment tasks[20–24], one might question why I observed no history effect in my zero-second delay condition, as opposed to a distance-dependent repulsion. There are several possibilities worth considering. First, it should be noted that a few individual participants in this study did evince significant repulsion (data available upon request), though the effect was not significant at the group level. Adaptation seems to be strongest when stimulus exposure times are longer than one second[1] or when the task involves a perceptual comparison[1, 9], rather than the delayed estimation procedure used here. Furthermore, whereas serial dependence has been observed in a spatial delayed response task[16, 17], adaptation – to my knowledge – has not. Hence, the spatial task employed in this study was optimized to produce serial dependence, not adaptation. This makes it even more striking that the attractive bias was absent in the shortest delay condition tested.

Beyond demonstrating that serial dependence accumulates for longer in working memory than previous studies have indicated[9, 16, 17], I have taken strides to integrate this phenomenon into the study of working memory in ways it has not been before[32]. Specifically, I have made concrete, formal improvements to prominent mathematical models designed to characterize the psychological architecture of working memory. The provision of terms for serial dependence to these models both (1) allows them to capture more variance in behavioral data and (2) ensures that the variance associated with temporal smoothing does not distort estimates of the models' other parameters. Claims that have been made about the nature of decay rates in working memory without consideration of trial-history biases must now be reëvaluated[42]. Furthermore, I have determined the approximate duration for which serial dependence persists between trials. At least in spatial working memory, the attractive bias disappears within ten seconds after the end of each trial. This constrains possible neural theories of serial dependence – viable mechanisms must have time constants on the order of 10 s, which rules out especially short-term (e.g., synaptic facilitation) and long-term (e.g., long-term potentiation) forms of plasticity. Previous attempts to measure the washout period of serial dependence have used a short, fixed ITI, preventing the measurement of pure time in the absence of intervening trials[1, 14].

Reframing serial dependence as a phenomenon of working memory rather than perception does not change the theories that have been put forth about its functional importance[32]. Thus, it remains an important mechanism for stabilizing representations against interruptions in visibility[1, 8]. The contents of working memory track the focus of attention[25–31], which, during the execution of a single goal, can remain the same for several seconds, even as the raw visual input that impinges on the retina fluctuates due to saccades, occlusion, and changes in lighting. Hence, temporal autocorrelation in visual working memory is potentially even higher than it is in visual scenes (and perception). If true, this would explain why serial dependence may have evolved in working memory as opposed to perceptual circuits – more autocorrelation enhances the ability of temporal smoothing to limit the influence of noise and boost signal. Moreover, the offloading of attractive serial dependence to memory systems

may accord perceptual systems enhanced capacity to specialize in novelty detection, in part via adaptation. More research is needed to elucidate the ways in which serial dependence and adaptation interact, and to reveal the ecologically-valid situations in which one or the other (perhaps both at the same time) enhance visual performance[32].

Methods

Participants

Fifty-five adults from the UC Berkeley community were recruited to participate in this study. Thirty-five of these individuals completed Experiment 1 only, fourteen completed Experiment 2 only and six completed both experiments. All aspects of data collection and analysis were conducted in accordance with guidelines approved by the Committee for the Protection of Human Subjects at UC Berkeley. Informed consent was obtained from all subjects, and they were compensated monetarily for their time.

Experimental Procedures

Participants completed the protocol in a soundproof, dimly lit testing room. The tasks were programmed in MATLAB using the Psychophysics Toolbox[69] (version 3) and run on a Mac mini (OSX El Capitan 10.11). For eight subjects in Experiment 1, a 17-in monitor was used with a screen resolution of 1280 X 1024 pixels. The remaining sessions were run with a 23-in monitor, 1920 X 1080 pixels. All participants were seated such that their eyes were approximately 60 cm from the center of the testing display.

The stages of the generic task used for both experiments are depicted in Figure 3.1 (adapted from[46]). All stimuli were displayed against a gray background. In the description that follows, all angle measurements are reported in degrees of visual angle. Participants were instructed to fixate the central black square – which spanned 0.5° X 0.5° – whenever it was on the screen (all stages of the task aside from the response period). Each trial began with the presentation of a black circle for 1 s at a random angle from fixation, with eccentricity fixed at 12°. The circle's diameter was 1°. In Experiment 1, participants remembered the location of this circle for a delay that varied randomly from trial to trial (0, 1, 3, 6, or 10 s). The delay was always 3 s in Experiment 2. At the end of the delay, the fixation square was replaced with the mouse cursor (at the exact center of the screen), and participants indicated the location in mind by moving the cursor to that location and clicking once. No feedback was given. Errors were measured in degrees of polar angle. In Experiment 1, a 1-s ITI followed the response period, before the start of the next trial. The ITI varied randomly from trial to trial in Experiment 2 (1, 3, 6, or 10 s). Each participant completed 1,000 trials (200 per delay) in Experiment 1 and all but two participants completed 1,008 trials (252 per ITI) in Experiment 2. The remaining two participants completed 999 and 1,017 trials, respectively.

Data Analysis

The data were analyzed using Python, MATLAB, and shell scripts. All code written for this study is available upon request.

Before model fitting, the data were submitted to preprocessing. First, trials with responses that were within 5° of visual angle of the origin were dropped (0.2% of all trials, across subjects). Next, I computed systematic directional error as the mean response for each stimulus location. This mean was then subtracted from the response on each individual trial (ignoring the location of the previous trial) to obtain the residual error that was used to characterize serial dependence. Replicating the procedure in[16], I computed the systematic error by spatially low-pass filtering the responses as a function of stimulus location using the MATLAB function `loess`. Finally, to ensure that my analyses were restricted to data from participants who performed the task correctly, I removed those with noticeably poor performance (mean absolute error greater than 10° of polar angle). Only three subjects in Experiment 1 failed to pass this criterion, two in Experiment 2.

I used the Clifford model[20] to characterize the tuning of serial dependence across all possible differences between past and current visual input. In this study, differences ranged between -180 and 180° of polar angle (a complete circle). The Clifford model applies most readily to feature spaces that are circular, and such spaces are commonly used in the study of serial dependence[1, 2, 8, 9, 16, 17]. Although the Clifford model was originally designed to account for adaptation, it was trivial for us to extend it to fit, flexibly, either serial dependence or adaptation (which differ only in sign). The model is stated as follows:

$$\sin(\theta_A) = \frac{\sin(\theta_0)}{\sqrt{(s\cos(\theta_0) - c)^2 + \sin^2(\theta_0)}}, \tag{1.1}$$

where θ_A is the perceived location (in polar angle) of the current stimulus, θ_0 is its true feature value (in the absence of any trial-history effects), s is the scaling parameter, and c is the centering parameter. Both θ_A and θ_0 are expressed relative to the previous trial's true location. I used the scipy[70] function `least_squares` (in the `optimize` module) to find the values of c and s that minimized the difference, for each θ_0, between the estimated θ_A and the subjects' errors. Across all values of θ_0, I take the magnitude of serial dependence (or adaptation) to be the peak-to-peak of $\theta_A - \theta_0$, with the sign adjusted to match the direction of the effect. As explained in the Results, this peak-to-peak depends on both the scaling and centering parameters.

To determine whether the magnitude of serial dependence was significantly greater than zero, or greater in one condition than in another, I submitted the data to permutation testing at the group level[2, 9]. Specifically, I shuffled the values of θ_0 (current trial's location relative to the previous trial's) while leaving in place the corresponding errors. I then fit the Clifford model to the shuffled dataset. This process was repeated 10,000 times. As p-values I report the proportion of permutations that led to equal or higher values for the peak-to-peak of the function fit than the one estimated for the unshuffled data. In the case of a comparison between conditions, I subtracted the null peak-to-peaks for one condition from those for

the other, and report the proportion of these differences that had equal or higher values
than the empirical difference. (I tested the data for both adaptation and serial dependence,
but found no evidence of adaptation in any condition.) The criterion for significance was
Bonferroni-corrected for each family of tests.

Some additional statistical methods were employed just for data visualization. Before
computing the variance for Figure 3.3, I removed trials with errors beyond six standard
deviations above or below the mean (never as much as 0.4% of trials for any individual delay).
This correction was necessary to obtain a clean measure of variance, which is notoriously
vulnerable to outliers. For Figures 3.2 and 3.4, I computed bootstrapped confidence intervals
as follows[2, 9]: I resampled the data with replacement 10,000 times. To each resampled
dataset, I fit the Clifford model. This yielded a distribution of peak-to-peak values from
which I selected the boundaries of the 95% confidence interval – separately for each delay
and ITI condition.

Three base mathematical models of working memory – EP[34], VP[37, 38], and VMRW[40,
41] – were fit to my behavioral data, as described in the Results. Model fitting was done
using the MATLAB function `fminsearch`, separately for each delay condition. EP is defined
as

$$p(\hat{s}|s;\kappa) = \frac{e^{\kappa\cos(\hat{s}-s)}}{2\pi I_0(\kappa)}, \tag{1.2}$$

the von Mises probability density function. Here, $\hat{s}$ is each trial's response, s the corre-
sponding stimulus, and I_0 the modified Bessel function of the first kind, order zero. The
concentration parameter, κ, is a measure of response precision, spans all trials, and is the
model's one free parameter for fitting.

In VP, precision is drawn anew for each trial from a gamma (γ) distribution with mean
$\bar{J}$ and scale parameter τ (the model's free parameters). Built from EP, this gives

$$p(\hat{s}|s;\bar{J},\tau) = \int EP(\hat{s};s,\Phi(J))\gamma(J;\bar{J},\tau)dJ, \tag{1.3}$$

where $\kappa = \Phi(J)$, a relation that can be computed numerically[37]. The integral similarly
has no analytical expression and so is approximated using Monte Carlo simulations[37, 39].

Finally, in VMRW, noise in working memory is distributed according to a von Mises ran-
dom walk, as derived from a population coding model of cortex[41]. Specifically, behavioral
errors for a random walk of length r are von Mises distributed:

$$p(\hat{s}|s;r,\kappa) = \frac{e^{\kappa r\cos(\hat{s}-s)}}{2\pi I_0(\kappa r)}, \tag{1.4}$$

where the distribution of r for m walk steps is

$$p(r|m,\kappa) = \frac{I_0(\kappa r)}{I_0(\kappa)^m}r\psi_m(r). \tag{1.5}$$

Here, $r\psi_m(r)$ is the probability density function for a uniform random walk of length r and number of steps m. The variable m is itself Poisson-distributed with expected value ξ. For additional equations and a full derivation, including the neural interpretation of these variables, see[41]. In order to fit this model to data, I approximated the density $\psi_m(r)$ via Monte Carlo simulation. The free variables for fitting are κ (the concentration parameter) and ξ, which corresponds to gain.

To these base models I added terms to capture temporal smoothing in the data in the form of serial dependence (or adaptation). In particular, I allowed the mean of each model's probability density function to vary on a trial-by-trial basis, as a function of the location of the previous trial's stimulus. Given a particular difference in location between the current and previous trial's stimuli, the mean shift was set to be the value of the Clifford model fit to the data at that point. (That is, in visual terms, the input to the model was a point on the x-axis in Figure 3.2, and the output mean shift was the Clifford function's value on the y-axis.) This procedure added two additional variables to each of the base models – c and s.

As an alternative to the base models with the Clifford expansion, I made alternative models that account for trial history by assuming that participants, on a subset of trials, confuse which stimulus was presented most recently and report the wrong item when probed. This alternative similarly allowed the mean of the base probability density functions to shift, depending on the difference between the previous trial's location and the current one, without altering their shape or width. This "swap over time" model is defined as[36]

$$p(\hat{s}|s) = (1 - \alpha)BM(\hat{s} - s) + \alpha BM(\hat{s} - s^*), \tag{1.6}$$

where BM is a base model, α (an additional free parameter) sets the frequency of swaps, and s^* is the stimulus location for the previous trial.

I combined likelihoods for different delays and subjects, and formally compared the fits of different models using the Akaike Information Criterion (as recommended in[39]), with the standard correction for finite sample sizes.
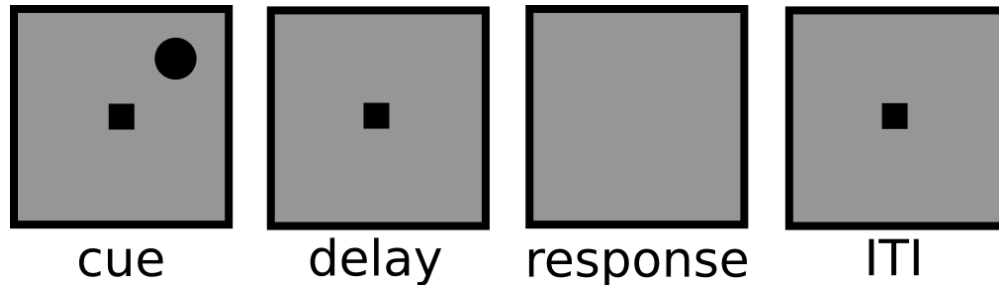
Figure 1.1: The events in each trial of the generic spatial judgment task used for Experiments 1 and 2 (not to scale, see Methods for exact dimensions). Stimuli were presented in black against a gray background. Participants maintained fixation at the central square whenever it was on the screen (all task stages aside from the response period). Each trial started with the presentation of the cue whose location needed to be remembered for a variable (Experiment 1) or fixed (Experiment 2) delay. Upon the disappearance of the central square at the end of the delay, the mouse cursor appeared at the exact center of the screen (not shown), and subjects used the mouse to make their response. Responses were not timed. Immediately after the response was made, the fixation square returned for a fixed (Experiment 1) or variable (Experiment 2) inter-trial interval.
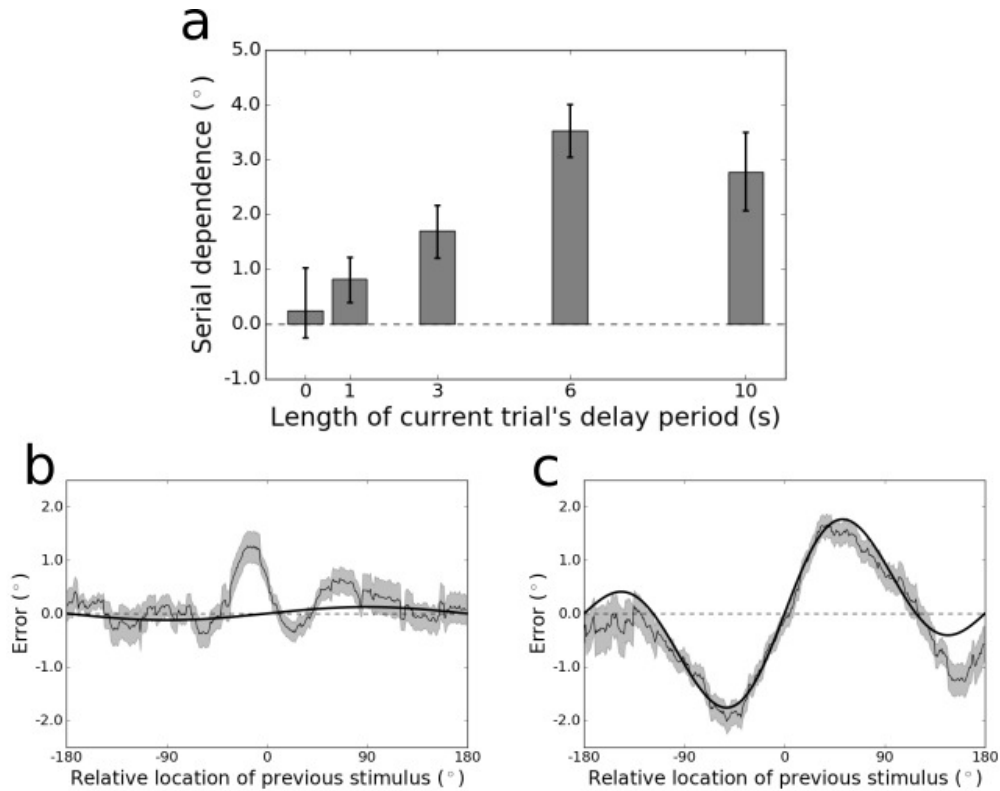
Figure 1.2: (A) Peak-to-peak of the Clifford model fit to the group data for each delay
period tested in Experiment 1. Error bars represent bootstrapped 95% confidence intervals.
The magnitude of the serial dependence increases during the first 6 s of the delay period.
(B) Tuning of serial dependence across all possible angular differences between the current
and previous stimulus, for the 0-s delay condition. The thin black line represents the group
moving average of response errors, with the standard error in gray shading, and the thick
black line is the best-fitting Clifford model curve. Serial dependence is not significantly
greater than chance in this condition, and the peak-to-peak of the Clifford fit is just 0.25°
of polar angle. (C) Tuning of serial dependence for the 6-s delay condition. Here, serial
dependence is significantly greater than chance, and the peak-to-peak of the Clifford fit is
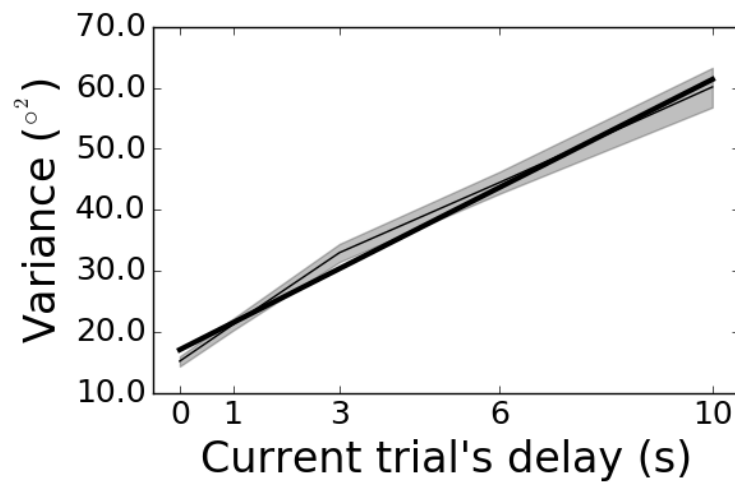3.53°.

Figure 1.3: Variance of response errors as a function of the current trial's delay in Experiment
1. The thin black line depicts the group mean, with bootstrapped 95% confidence intervals
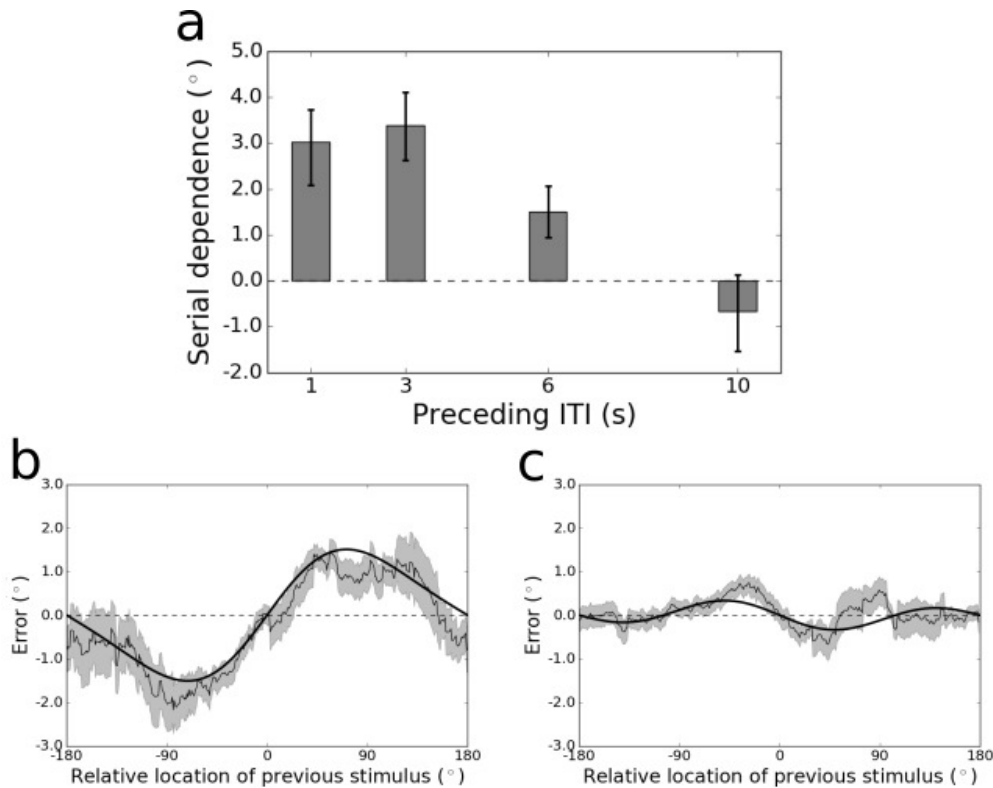in gray shading. In thick black is the linear best fit.

Figure 1.4: (A) Peak-to-peak of the Clifford model fit to the group data for each ITI tested in Experiment 1. Error bars represent bootstrapped 95% confidence intervals. Serial dependence decreases in magnitude as the ITI lengthens, falling to chance levels with an ITI of 10 s. (B) Tuning of serial dependence across all possible angular differences between the current and previous stimulus, for the 1-s ITI condition. The thin black line represents the group moving average of response errors, with the standard error in gray shading, and the thick black line is the best-fitting Clifford model curve (3.02° peak-to-peak). (C) Tuning of serial dependence for the 10-s ITI condition. Here, the peak-to-peak of the Clifford fit is -0.67° (non-significant adaptation).

Chapter 2

Synaptic augmentation in a cortical circuit model reproduces serial dependence in visual working memory

Introduction

Standard paradigms for measuring the contents of visual working memory have revealed that human observers tend to merge features of stimuli from previous trials into their representation of the current one, leading to a systematic bias in behavioral reports[1–14]. This smoothing of representations – termed "serial dependence" – is spatiotemporally tuned[1–7, 12, 14, 15] and sensitive to the featural similarity between items[1–3, 7–10, 12, 14, 16, 17]. I replicated this pattern of the behavioral bias in human participants in Chapter 1 of this dissertation. While it is proposed that serial dependence directly alters stimulus perception[1, 8] – and precedes the onset of memory or decision making[1, 8] – more recent studies (such as that in Chapter 1) have demonstrated that the trial-history bias is absent at the time of perception[9] and evolves slowly during the subsequent delay period of a working memory task[9, 16, 17], reaching an asymptote around six seconds after the most recent item has been encoded into memory and removed from view. The working memory system tends to maintain preferentially the objects in visual scenes toward which observers direct their attention[71]. In natural environments, the focus of attention tends to track the same object for several seconds at a time[72], which implies that the contents of visual working memory do not shift radically from moment to moment. Hence, a temporal smoothing operation like serial dependence would seem to be a useful mechanism for maintaining the stability of the current representation[9, 32].

The evidence that serial dependence in behavior is more strongly associated with the contents of working memory than immediate perception provides a starting point for the investigation of its neural mechanisms. Current theories of working memory representation in cortex have not sufficiently accounted for serial dependence in their models. In general,

the neural code that supports working memory has been the subject of considerable debate in the past few years[31, 55, 73–76]. A classical view[74, 75, 77], inspired by pioneering work in the Goldman-Rakic Lab[78], has been that persistent activity in cortical neurons tuned to stimulus features sustains representations across memory delay periods between perception and action. Reverberatory mechanisms have been posited to support persistent activity[48, 79, 80]. Alternative theoretical considerations[81], as well as occasional failures to identify persistent activity in physiological experiments[82–85], have called this account into question. A newer "activity-silent"[76] model proposes that a brief burst of neuronal firing may be sufficient to drive synaptic changes in cortical networks that can then maintain the information in memory in the absence of continued elevated activity[31, 55, 73, 81]. This theory relies on the finding that the capacity for short-term synaptic plasticity – in particular, synaptic augmentation – is enriched in cortical areas associated with working memory[86, 87].

Preliminary attempts to demonstrate how existing (or revised) neural models of working memory would give rise to serial dependence have focused on an activity code, rather than changes in synaptic weights[16]. It has since been argued that accounting for dynamic synapses might reconcile discrepant trial-history effects in neural recordings and behavior[17], but the biophysical details of how this would work have not been modeled. Perceptual studies have shown how short-term plasticity could explain a shift in responses from one trial to the next[88, 89], but these models have focused on the repulsive shift associated with sensory adaptation, not the attractive shift of (mnemonic) serial dependence. Adaptation and serial dependence manifest as mirror-opposite biases in behavior, but, unlike serial dependence, adaptation appears to be a phenomenon of passive perception that is independent of attention[1, 9]. Plasticity occurs at many time scales in the brain, from synaptic facilitation, which returns to baseline within hundreds of milliseconds, to homeostatic scaling that can last for hours or days[90]. The time scale of the rise and fall of serial dependence in behavior (~ 10 s per Chapter 1) is similar to a form of plasticity observed in neural circuits important for working memory – synaptic augmentation – which is the seconds long increase in synaptic vesicle release that occurs when calcium accumulates in presynaptic terminals[91] (in direct proportion to the presynaptic firing rate[86]).

In the present study, I investigate whether biophysically-detailed models of working memory storage in cortical circuitry can account quantitatively for the spatiotemporal tuning of serial dependence in memory-guided behavior. I initially test the traditional bump attractor model[48, 92], which proposed that active firing sustains memories. The reverberation in this network depends on NMDA receptors, balanced excitation and inhibition, and tuning to stimulus features that is wider for inhibitory than excitatory neurons[79]. Numerous empirical studies support the importance of these network features for working memory function[78, 80, 93–95]. However, I demonstrate that active firing alone in a bump attractor network is insufficient to account for the temporal dynamics of serial dependence. In contrast, endowing synapses in a bump attractor network with the capacity for realistic augmentation[86, 87, 96] causes a gradual potentiation of serial dependence over the memory delay period that, consistent with human psychophysics (Chapter 1), asymptotes within 10 s. Furthermore,

between trials, in the absence of new visual input, serial dependence in the simulated circuit decays within 10 s, again matching human behavior (from Chapter 1). Importantly, my final model represents a hybrid of principles from traditional and "activity-silent" theories of working memory, demonstrating with computational precision how these competing accounts can co-exist in the same circuit and make dissociable contributions to memory-guided behavior.

Results

Model 1: Original bump attractor model (with fixed synapses)

First, I confirmed that the traditional bump attractor model[48, 92] does not produce a signal that corresponds to serial dependence in human memory-guided behavior. That is, I determined whether there was any dependence between each behavioral response decoded from the network and the stimulus from the previous trial. The bump attractor model was developed as an explicit mechanistic theory of sustained, elevated firing among neurons in association cortex during the delay period of working memory tasks[48], a signature of mnemonic maintenance[77, 79]. In the network, synaptic reverberation supports a self-sustaining bump of activity, centered at the visual feature value encoded at the start of the trial. (This feature is generally considered to be a spatial location[48] – to match the stimulus type tested in early physiological experiments[78] – but it could also be an orientation[97], or any other property of the stimulus that falls along a continuous, circular dimension.) Several properties of the network underlie its capacity for reverberation: signaling through NMDA receptors, balanced excitation and inhibition, broader (or absent) tuning to the task-relevant stimulus feature among inhibitory neurons than excitatory ones, and stronger connectivity among neurons with similar preferred feature values[48, 79]. These characteristics of the network and its resultant population activity during simulated working memory task performance are remarkably consistent with neurophysiological data recorded from monkey cortex[78, 80, 93–95].

I implemented the reduced firing-rate version of the model[92], given that it recapitulates the behavior of the full spiking model[48] while consuming far fewer computer processing resources[98]. Simulations of a working memory task were run through the network (Fig. 3.1A). The generic task structure I used is common in studies of serial dependence[1, 2, 8, 9, 16, 17]. On each trial, a stimulus is presented briefly and then removed from view. A feature of this stimulus must be remembered over a blank delay period (usually a few seconds in length). Finally, the end of the delay is signaled, and a response is made to indicate the feature value in memory. Each simulation comprised a single pair of trials. Thirty-two combinations of stimuli were used, with the difference between the current and previous trial's stimulus varied uniformly around the circular feature space. For each of these combinations, 100 simulations with different random seeds were performed. Population activity for an example trial pair is displayed in Figure 3.1B. I extracted the behavioral response for each

trial from the network using population vector decoding[99], which identifies the approximate center of the activity bump[48].

As expected, despite realistic levels of noisy firing in the network, the behavioral performance of the model was on average veridical, with no systematic bias in responses with regard to the stimulus from the previous trial (Fig. 3.1C). This result is not consistent with human[1, 2, 9] and non-human[16, 17] primate performance on this task (including that from Chapter 1). The absence of serial dependence in the bump attractor model is not specific to any particular moment in the delay period. When the network was probed to make a response at variable time points after stimulus offset, there was no detectable increase or decrease in the trial-history bias – serial dependence was near zero throughout (Fig. 3.1D).

Model 2: Activity leak model (with fixed synapses)

The null results presented above provided an impetus for making specific adjustments to the bump attractor model to allow serial dependence to be observed. It has been proposed – within the traditional rate code model of working memory – that persistent activity may extend beyond the end of a given trial and leak into the next one, causing a disruption of the encoding of the subsequent stimulus that depends on the distance between the successive activity bumps[16, 17, 32]. There is evidence from data collected from the frontal eye fields of non-human primates that such an activity leak does occur, at least to some extent[17]. However, whether this leak biases neural responses in the way that can explain the pattern of serial dependence in behavior is unknown[17].

I implemented activity leak in the bump attractor model by weakening the mechanism that resets the network during the response period of each trial. Specifically, in the default version of the model, the start of the response period commences widespread inhibition in the network that shuts off the bump of activity and returns the neurons to unstructured, baseline levels of firing. (This reset may be a proxy for corollary discharge from motor areas in cortex[48, 100].) To obtain levels of leak firing into the inter-trial interval (ITI) commensurate with observations in monkey cortex[17] (Fig. 3.2A), I reduced the inhibitory current passed into the network at the response period by 88.4% (see Methods).

Again, 100 simulations of the task with different initial conditions were executed for each of 32 combinations of trial pairs, with the differences between successive stimuli uniformly covering the range of possible differences. The leak of elevated firing into the ITI is evident in the example population response shown in Figure 3.2B. Unlike the original model, the activity leak model did give rise to serial dependence. The pattern of behavioral responses decoded from the network bore a striking resemblance to what has been observed in the behavior of humans[1, 2, 9] and monkeys[16, 17] (Fig. 3.2C). The same function that I used to fit the tuning of serial dependence in human behavior in Chapter 1 (the Clifford function[20]) provides an excellent fit to the tuning over stimulus differences from the leak model. Furthermore, the amplitude of the fit (2.53° peak-to-peak) is on the order of that for the human data from Chapter 1 (3.53° peak-to-peak for the same timing of task events). However, over the full range of working memory delay periods tested – from extraction of the

behavioral response immediately after perceptual encoding to a delay before the response of 10 s – serial dependence in the activity leak model remained constant (Fig. 3.2D). This is in stark contrast with human performance, which shows a gradual rise in the magnitude of serial dependence as the memory delay advances – for up to 6 s per Chapter 1. The failure of the model to show any temporal dynamics may be explained by the notion that residual firing from the previous trial interacts with the newly forming bump at just one moment in time – the start of each trial (Fig. 3.2B). After that, the center of the new bump is established (shifted slightly off the stimulus' true feature value), the residual activity is quashed (as in primate cortex[17]) by lateral inhibition, and no mechanism remains to systematically alter the bump's drift as the delay proceeds.

Model 3: Bump attractor model with plastic synapses

Neither the original bump attractor model nor the activity leak version could reproduce the temporal dynamics of serial dependence in human memory-guided behavior. Hence, I turned to a property of neural circuits that has been assigned a great deal of importance in newer theories of working memory storage: short-term synaptic plasticity[31, 55, 73, 76, 81]. In particular, I focused on synaptic augmentation[91], both because its rate of accumulation and decay[86, 87, 91] matches the time scale of serial dependence measured in Chapter 1, and because it is a prominent synaptic dynamic in cortical circuits associated with working memory[86, 87]. The magnitude of augmentation is a function of calcium dynamics in presynaptic terminals[91]. Each new action potential passing into the terminal increases the likelihood that calcium channels are open, and, in the absence of spiking, intracellular calcium levels decay to baseline exponentially[91]. The dynamics of synaptic weight updates due to these processes has been implemented successfully in phenomenological models that do not include explicit variables for voltage-gated calcium channels or individual action potentials[96, 101]. Consistent with these theoretical models, as well as empirical work[86, 87, 91], I modeled the rise of synaptic augmentation for a given synapse as being proportional to the presynaptic firing rate (up to a saturation point) and assumed exponential decay. The capacity for augmentation was instantiated for every synapse in the original bump attractor network.

A third time, I ran 100 simulations of the task for each of 32 trial pairs, uniformly sampling the range of possible trial-to-trial stimulus differences. Figure 3.3A shows the subtle effects of augmentation on firing rates in the network, for a sample trial pair. Like the leak model, the augmentation model produced behavioral responses that tracked the pattern over stimulus differences seen in humans and monkeys (Fig. 3.3B, 2.66° peak-to-peak). I observed that the width of the tuning of the trial-history effect was sensitive to the width of the tuning of individual neurons in the network (see Methods). Wider tuning curves widened the tuning of serial dependence – as this caused augmented synapses from the previous trial and the bump of activity on the new trial to interact over greater distances in the network – whereas narrower tuning of individual cells narrowed the spread on the serial dependence plot (Fig. 3.3B, supplementary figures available upon request).

Most important with regard to the key shortcoming of the activity leak model, the augmentation model gives rise to a slowly evolving time course of serial dependence over the working memory delay that is quantitatively in line with psychophysics data (Fig. 3.3C). When responses are decoded from the network immediately after stimulus offset, at the end of the perceptual period (0 s of delay), the magnitude of serial dependence is near zero – within the confidence bounds of the human data from Chapter 1, which were non-significantly different from zero. From that starting point, I measured serial dependence at time points within the delay period 1, 3, 6, and 10 s after stimulus offset, as for the other models. The amplitude of the history effect at each of these delays exceeded the confidence bounds of the next-shortest delay, except at 10 s. Consistent with the human data, serial dependence reached an asymptote in the network between 3 and 10 s. Furthermore, its final amplitude at 10 s falls within the confidence bounds of the human data for this delay length (Chapter 1) – making the model a tight fit to human behavior across the full temporal range of serial dependence.

I confirmed that this time course in the model is specific to the memory period, and not an artifact of the mere passage of time, regardless of maintenance demands. To do this, I extended the ITI between the trials within each pair, keeping the delay length constant. I ran 100 simulations with different random seeds for each of 32 stimulus differences and four ITIs. Serial dependence decreased the longer the ITI was extended – approximating zero with an ITI of 10 s (Fig. 3.3D). Again, this result precisely matches human behavior (Chapter 1).

Discussion

In both the psychology[35, 67, 68, 102] and neuroscience[31, 73–76] literatures, proposals for the organization of working memory have tended to be contentious rather than compatibilist. For example, cognitive scientists have debated for years whether distinct items in working memory are arranged in discrete slots or assigned continuous resources[35, 67, 68, 102]. In this instance, modeling work has determined how neural circuits incorporate properties of both of these extremes[103], recasting the debate that pits the two theories against each other. The present study had a similar aim but in a different context. My final model contains explicit computational elements of two major competing theories of working memory – synaptic reverberation to support elevated delay-period activity and synaptic plasticity to support "activity-silent" memory storage – and demonstrates how the underlying mechanisms in these two models can operate synergistically to guide behavior. Persistent activity drives synaptic weight changes, which in turn can bias the drift of persistent activity. Although these processes occur in tandem, their contributions to behavior can be dissociated. Noise in the inputs to the network (from other cortical areas and local neurons without tuning to the task-relevant stimulus feature) cause random drift of the population response that accumulates linearly with time[48, 92]. This is consistent with the linear increase in the variance of human memory-guided behavior over the delay period (Chapter 1). Thus, the

random error in behavioral responses during working memory is due to the stochasticity of active neural firing[40, 41, 104]. Serial dependence, in contrast, originates not from neuronal spiking but from plasticity due to synaptic augmentation.

It is noteworthy that my synaptic augmentation model provides a quantitative (rather than merely qualitative) fit to the human behavioral data – matching both the time scale and amplitude of serial dependence measured in Chapter 1. In recent years, variations on the bump attractor model have settled for qualitative approximations of results from human psychophysics[46, 103], leading some to argue that, as a simplified model of cortex, it cannot make precise predictions about behavior[46]. Here, I show that even the reduced rate version of the model can be quite precise in its account of cognition. Other prominent working memory models that have yielded similarly close fits to behavior are invariably agnostic about biophysical implementation[34–39, 43]. An exception is recent work by Paul Bays[40, 41, 104] that has delineated how Poisson-distributed spiking in a model of visual cortex implies particular deviations from normality in error distributions from continuous-report working memory tasks (like the one in Fig. 3.1). However, the equations of this model require dramatic violations of biophysical realism: synapses are ignored, for example, such that the spike trains among different neurons are completely uncorrelated[40, 41, 104].

Combining neural modeling with behavioral and physiological experimentation in non-human primates, Papadimitriou and colleagues have attempted to uncover the neural mechanisms of serial dependence in working memory[16, 17]. Similar to my approach, they constructed their neural model iteratively, establishing the insufficiency of simpler, alternative versions to account for the full range of empirical findings. However, as a result, their final model attains explanatory power at the expense of parsimony. It requires that multiple independent memory storage sites with different decay rates combine to guide behavior[16], that receptive fields shift at a particular magnitude as a function of recent experience[17], and that Hebbian plasticity in projections to a downstream "readout" circuit causes a reversal of the population response before the response is executed[17]. While the model is a good fit to monkey behavior, these assumptions go beyond what physiological data have demonstrated is plausible, and the biophysics in their final model is incompletely specified: the receptive field shifts and Hebbian plasticity are applied instantaneously via hard-coding of values. My model, in contrast, achieves a comparable fit to human behavior with a single population of neurons, undergoing synaptic updates with precisely the same temporal dynamics that have been measured in association cortex[86, 87].

One potential criticism of my augmentation network model is that it assumes all synapses in the circuit experience the same kind of plasticity, which is demonstrably false in real cortical networks[86, 87]. However, it is at this level of analysis that the realism of the bump attractor model as a whole breaks down. The reduced version of the model I used as my starting point contains just 256 neurons[92], far fewer than would be expected to participate in the performance of a working memory task *in vivo*. Hence, these neurons should be viewed as abstractions of a larger collection of cells with heterogeneous synaptic dynamics, but in which the augmentation signal is prominent[86, 87]. From this perspective, the small amplitude of augmentation in individual synapses in my network (see Methods) should be

interpreted as the average over many synapses, only some of which are reliably plastic. This consistent scaling of magnitudes solidifies, rather than detracts from, the biophysical accuracy of my model. In the past, the bump attractor model has been found to perform equivalently regardless of whether heterogeneity among cells is explicitly coded, as long as homeostatic processes keep connections balanced[105].

Methods

General Description of the Model

All model simulations were performed with a working memory task that I depict as testing memory for the stimulus' angle from fixation (Fig. 3.1). All of the code used to perform the simulations was written using the Python package `brian2`[106] and is available upon request. I used the reduced firing-rate version of the bump attractor model that has been demonstrated to recapitulate the activity patterns of the full, spiking version[48, 92, 98]. The details of this general framework have been reported previously[92, 100]; I summarize them here. Excitatory signaling for each neuron is defined with a single variable s that represents NMDA conductance (specifically, the fraction of open channels). The differential equation for s is

$$\frac{ds}{dt} = -s/\tau_s + (1-s)\gamma f(I), \tag{2.1}$$

with $\gamma = 0.641$ and $\tau_s = 60$ ms. The firing rate r for each neuron is computed as a function of the total synaptic current I:

$$r = \frac{aI - b}{1 - \exp[-d(aI - b)]}, \tag{2.2}$$

with $a = 270$ Hz/nA, $b = 108$ Hz, and $d = 0.154$ s. Synaptic current comes from three sources: recurrent signaling, sensory drive, and random noise ($I = I_r + I_s + I_n$). Recurrent input to a neuron i from each other neuron j in the network is summed as follows:

$$I_{r,i} = \sum_j g_{ij}s_j, \tag{2.3}$$

where g_{ij} is the synaptic coupling from j to i.

Neurons are tuned to the task-relevant stimulus feature and uniformly tile the range of possible preferred stimulus angles (from 0-360°). The circuit contains 256 neurons total. The synaptic couplings g_{ij} have a Gaussian profile over all possible differences in tuning between neurons with preferred angles θ_i and θ_j:

$$g_{ij}(\theta_i - \theta_j) = J_- + J_+ \exp\left(-(\theta_i - \theta_j)^2/2\sigma^2\right). \tag{2.4}$$

For Models 1 and 2, $\sigma = 43.2°$. For Model 3, σ was varied between 30 and 70°. To generate the results displayed in Figure 3.3, I used $\sigma = 50°$. Parameters J_- and J_+ determine the levels of recurrent inhibition and excitation in the network, respectively. These levels are balanced such that the population can sustain persistent firing. For Models 1 and 2, I used $J_+ = 2.2$ nA and $J_- = -0.5$ nA. For the implementation of Model 3 used to generate Figure 3.3, I used $J_+ = 1.52$ nA and $J_- = -0.5$ nA.

When a stimulus at a particular angle θ_s is presented at the start of each trial of the task, neurons receive sensory currents that depend on their preferred angle θ:

$$I_s = g_s \exp\big(- (\theta_s - \theta)^2/2\sigma_s^2 \big), \tag{2.5}$$

where $\sigma_s = 43.2°$ and $g_s = 0.02$ nA. In all simulations, stimuli were presented for 1 s (as in Chapter 1).

Random noise passed into the network represents background activity in cortex unrelated to the task. (To be clear, the model does not assume that all neurons in cortex are tuned to the task stimuli – rather, neurons without tuning are represented implicitly rather than explicitly to conserve computational resources.) Noise varies over time as

$$\tau_n dI_n/dt = -(I_n - I_0) + \sqrt{\tau_n}\sigma_n\eta(t), \tag{2.6}$$

where $\eta(t)$ is white Gaussian noise, $I_0 = 0.3297$ nA, $\tau_n = 2$ ms, and $\sigma_n = 0.009$ nA.

I decoded the center of the activity bump to compute a behavioral response from the network during the delay period of each trial using the population vector method[48, 99].

Activity Leak Mechanism for Model 2

In the general model, the activity bump is reset during a response period that lasts 300 ms, throughout which unstructured inhibitory current (-0.08 nA) is passed into the circuit[100]. In order to allow a residual bump of activity to persist into the ITI, at a magnitude consistent with what has been observed in monkey cortex[17], I changed the reset signal to be -0.00925 nA.

Plasticity Rule for Model 3

My implementation of synaptic augmentation is based on[96]. I define the synaptic vesicle release probability F for each synapse as

$$\frac{dF}{dt} = \alpha(x - F)r - \frac{F}{\tau_F}, \tag{2.7}$$

where $\alpha = 0.015$, $x = 0.008$, and $\tau_F = 4.2$ s – which matches the best-fitting time constant for measurements of synaptic augmentation in prefrontal cortex[87]. In addition, the equation for the NMDA conductance was updated to

$$\frac{ds}{dt} = -s/\tau_s + (1 - s)\gamma(y + F)f(I), \tag{2.8}$$

where $y = 0.992$.

Simulations

To measure serial dependence in responses from the three models, I ran trials back-to-back
in pairs. The connections of the network are radially symmetric, allowing us to place the
first stimulus in each trial pair at 180° without loss of generality. Thirty-two angles were
tested for the second stimulus of each pair, evenly spaced between 0 and 360°. For each of
these 32 pairs, 100 simulations were run using different random seeds.

The timing of task events matched the protocol used in Chapter 1 to characterize the time
course of serial dependence in human behavior. Specifically, each stimulus was presented for
1 s, and a 1-s ITI was used. For the first trial in each pair, I tested a range of delay-
period lengths (up to 10 s). Consistent with what was observed in human behavior, I found
that the length of the delay period on this trial – the one that precedes the trial on which
serial dependence is measured – had little impact on the magnitude of the response bias
(a maximum difference of $\sim 0.6°$ peak-to-peak in the Clifford fit among all delays tested).
Hence, to conserve computational resources, the first trial in each pair for all simulations
used to generate my final results had a delay period of 1 s. The second trial in each pair had
a delay period of 10 s, and the behavioral response was decoded at several time points within
the delay (0, 1, 3, 6, and 10 s). Decoding took as input a window of population activity
starting 100 ms before the time point of interest.

For Model 3, I ran an additional battery of simulations in which the ITI was varied.
Again, 100 simulations (with different random seeds) were run for each of 32 trial pairs and
for each of four ITIs (1, 3, 6, and 10 s).

Characterization of Serial Dependence

I used the Clifford model[20] to characterize the tuning of serial dependence across all possible
differences between past and current visual input. The Clifford model applies most readily
to feature spaces that are circular, and such spaces are commonly used in the study of serial
dependence[1, 2, 8, 9, 16, 17]. The model is stated as

$$\sin(\theta_A) = \frac{\sin(\theta_0)}{\sqrt{(s\cos(\theta_0) - c)^2 + \sin^2(\theta_0)}},\tag{2.9}$$

where θ_A is the reported angle of the current stimulus, θ_0 is its true feature value (in the
absence of any trial-history effects), s is a scaling parameter, and c is a centering parameter.
Both θ_A and θ_0 are expressed relative to the previous trial's true location. I used the
scipy[70] function `least_squares` (in the `optimize` module) to find the values of c and s
that minimized the difference, for each θ_0, between the estimated θ_A and each model's errors.
Across all values of θ_0, I take the magnitude of serial dependence to be the peak-to-peak of
$\theta_A - \theta_0$.

I computed bootstrapped confidence intervals for the serial dependence magnitude in
each condition as follows[2, 9]: I resampled the data with replacement 10,000 times. To

each resampled dataset, I fit the Clifford model. This yielded a distribution of peak-to-peak values from which I selected the boundaries of the 95% confidence interval.
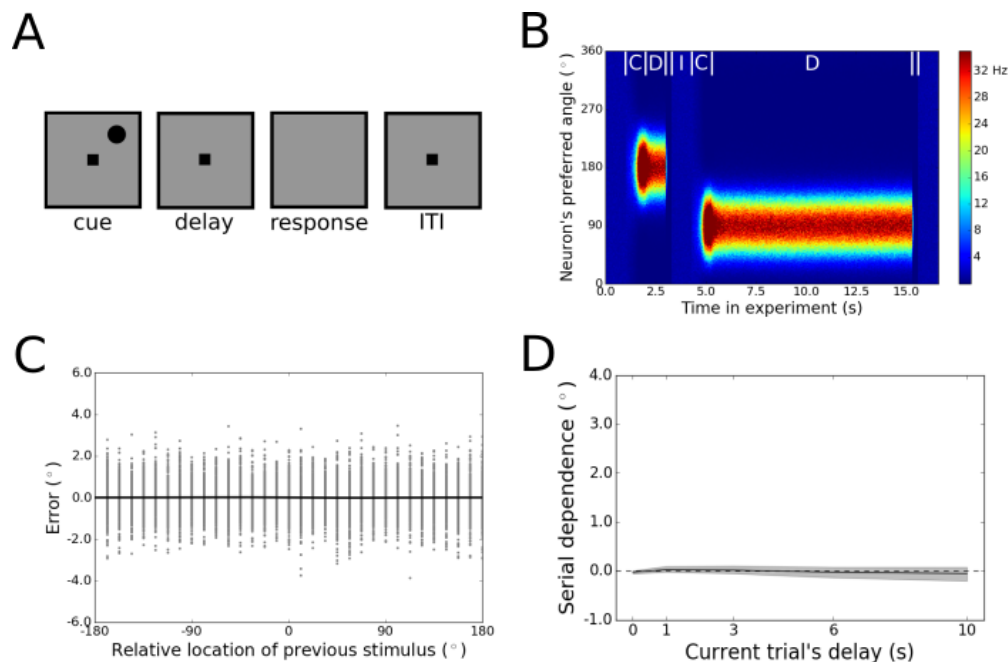
Figure 2.1: Results for Model 1. (A) The events in each trial of the working memory task used to test Models 1-3. Each trial started with the presentation of a cue whose angle from fixation was varied. The first cue in each pair of trials was presented at 180°. One of 32 angles evenly spaced around 360° was chosen for the second. The cue was maintained in the network across a delay period (without continued feedforward stimulus-driven activity). Finally, the removal of the fixation square from the screen signaled the start of the response period, when the network's response was decoded and activity was reset via the passage of unstructured inhibitory current into the network. An intertrial interval separated each response period from the subsequent cue period. (B) Activity in the network, with neurons lined up along the y-axis according to the stimulus angle to which they respond most strongly. Firing rates are color-coded. Two trials are shown with different delay lengths, the first with the stimulus at 180° and the second with the stimulus at 90°. At the start of each trial, during the cue period (labeled C in the figure), a bump of activity forms in the network, centered at the stimulus location. This bump weakens slightly during the transition to the memory delay period (labeled D), but is maintained throughout the delay. The bump is reset at the end of each delay, when the response is made. After the response, the intertrial interval (labeled I) begins. (C) Tuning of serial dependence across all angular differences between the current and previous stimulus. Raw errors for every simulation are in gray, and the best fit of the Clifford function used to model serial dependence is in black. No serial dependence is apparent in the responses – error is distributed evenly around a horizontal line at 0°. (D) Peak-to-peak of the Clifford model across the working memory delay. Shading depicts bootstrapped 95% confidence intervals. There is no evidence of serial dependence regardless of the duration for which the network maintains the current trial's stimulus feature.
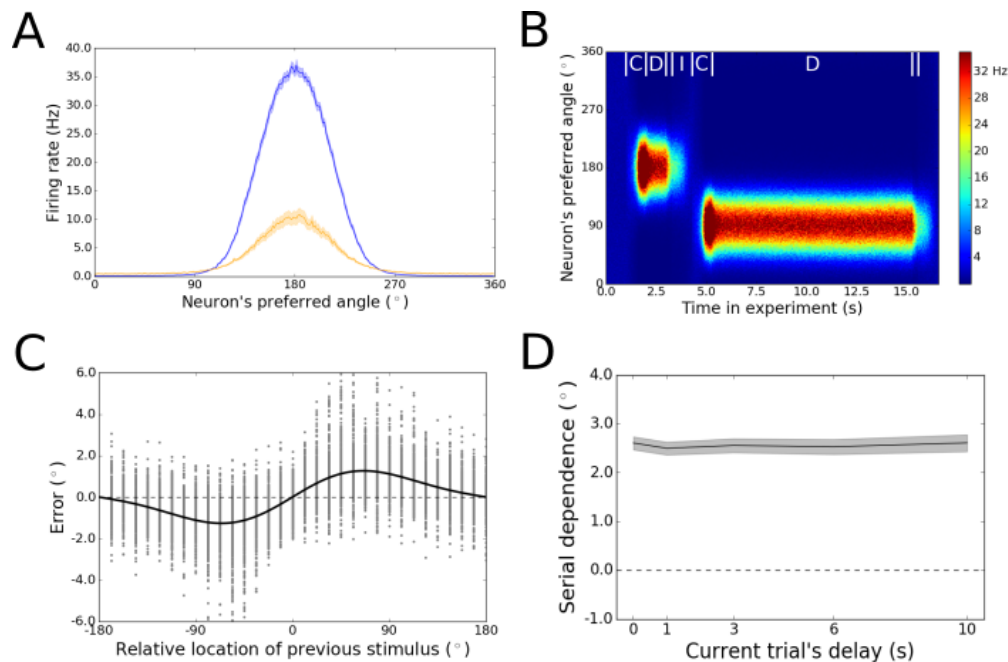
Figure 2.2: Results for Model 2. (A) Consistent with physiological data from monkeys[17], the amplitude of the activity bump during the ITI (yellow) is approximately one-quarter the size of the bump in the middle of the working memory delay (blue) in the activity leak model. Shading indicates the standard error of the mean for a single representative trial. (B) Spatiotemporal activity pattern for the activity leak model for two sample trials. The residual activity that continues beyond the response period into the ITI (labeled I in the figure) is visible. It is quashed as soon as a new stimulus is passed into the network due to lateral inhibition. On some trials, such as the one depicted, the residual activity begins to fade spontaneously before the next cue period (labeled C) has begun, due to insufficient levels of neural firing to keep the bump stable. (C) Tuning of serial dependence across angular differences between consecutive stimuli. Raw errors are in gray, and the best fit of the Clifford function is in black. Serial dependence with an amplitude consistent with what is observed in humans is apparent. (D) Peak-to-peak of the Clifford model across the working memory delay period. Shading represents bootstrapped 95% confidence intervals. Serial dependence is non-zero at all delays. The amplitude is large at the moment of stimulus offset and does not change over time.
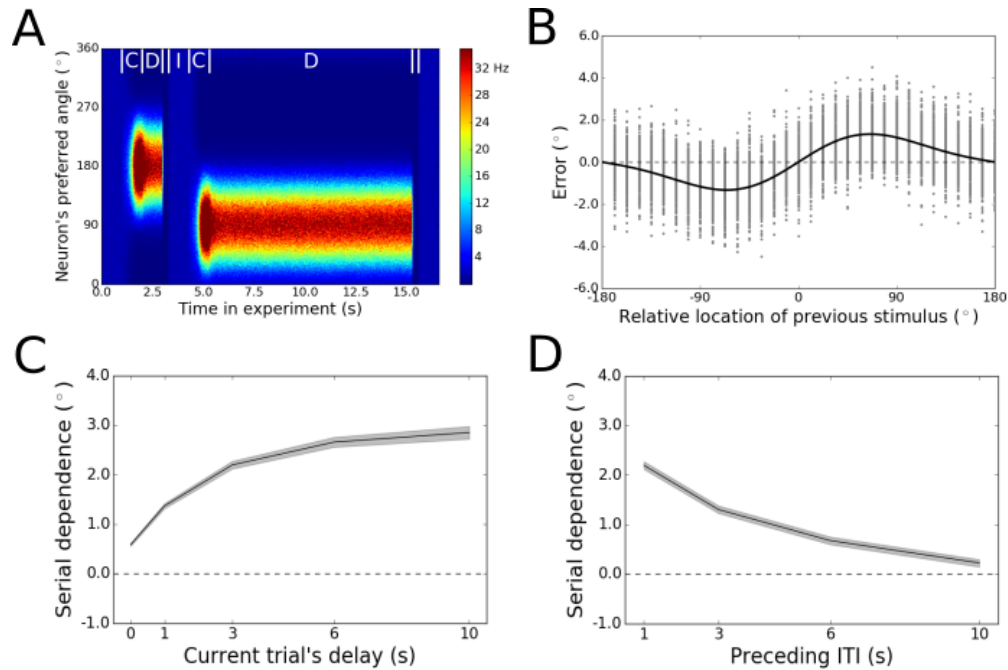
Figure 2.3: Results for Model 3. (A) Spatiotemporal activity pattern for the augmentation model for two sample trials. The effects of augmentation are scarcely visible as a subtle drift of the bump on the second trial in the direction of the center of the previous trial's bump. (B) Tuning of serial dependence across angular differences between consecutive stimuli. Raw errors are in gray, and the best fit of the Clifford function is in black. Serial dependence with an amplitude consistent with what is observed in humans is apparent. (C) Peak-to-peak of the Clifford fit across the working memory delay. Shading represents bootstrapped 95% confidence intervals. Serial dependence increases gradually over time before reaching an asymptote within 10 s. (D) Peak-to-peak of the Clifford fit across different ITIs. When the ITI is increased from 1 to 10 s, serial dependence becomes monotonically weaker.

Chapter 3

Prefrontal cortex promotes serial dependence in visual working memory

Introduction

In visual working memory, human observers tend to blend features of stimuli from previous trials into their representation of the stimulus currently being maintained[1–14]. This leads to a systematic bias in behavioral reports – termed "serial dependence" – that is spatiotemporally tuned[1–7, 12, 14, 15] and sensitive to the featural similarities between successive memoranda[1–3, 7–10, 12, 14, 16, 17]. These properties of serial dependence were replicated in the dataset collected for Chapter 1. Whether this bias is adaptive (as has been assumed in the perceptual literature) or maladaptive (as has been assumed in the memory literature, where serial dependence is termed "proactive interference") is still a matter of debate[32]. Claims of the adaptiveness of this phenomenon have rested on the notion that temporal smoothing would serve to stabilize representations in the face of fluctuations due to saccades, variations in lighting, and occlusion[1, 8]. Given the high temporal autocorrelation present in natural scenes[18], a weighted averaging of successive inputs would remove noise with minimal blurring of signal. In support of this rationale, recent psychophysical evidence indicates that the decay rate of serial dependence for different visual features is sensitive to the average rates of change of these features in typical environments[14].

The role that serial dependence plays in cognition would be clarified by an understanding of its neural mechanisms. While behavioral trial-history biases have been characterized extensively using the techniques of psychophysics, the study of the neural basis of serial dependence is in its infancy. Empirically, functional magnetic resonance imaging (fMRI) has identified signals in primary visual cortex that correlate with the bias in behavior[107]. At the circuit level, models of visual cortex have been designed to show that gain modulation centered at a recently encoded feature value could cause shifts in tuning curves and the population response consistent with serial dependence[1, 108]. However, this work leaves unstated the source of the gain modulation. Given that recent psychophysical experiments

(including Experiment 1 in Chapter 1) have demonstrated that serial dependence is more strongly linked to visual working memory than perception[9, 16], it seems plausible that top-down signals to sensory cortex from areas traditionally associated with attention and working memory – like the prefrontal cortex (PFC) – would contribute to this effect[32]. Viewing serial dependence as a phenomenon of working memory also encourages consideration of its time course over the memory delay period between stimulus and response, something static models of visual cortex have not done[1, 108]. Long-standing biophysical theory holds that synaptic reverberation in PFC sustains working memory representations over this delay period[48, 79, 80, 93]. In a recent modeling effort (Chapter 2), the addition of synaptic augmentation to this reverberatory neural-network model of working memory in the PFC caused the model to reproduce quantitatively serial dependence effects previously measured in human behavior. This adjustment of the model equations to include short-term plasticity was motivated by the empirical finding that synaptic augmentation is enriched specifically in the PFC as compared to visual cortex[86, 87].

To date, investigations into the neural basis of serial dependence (including that in Chapter 2) have been either theoretical or correlational[1, 16, 17, 107]. No cortical regions have been demonstrated, via a causal manipulation, to be necessary for this temporal smoothing operation. Regarding the PFC, a debate has persisted for decades about whether this area is required for working memory at all, let alone for serial dependence in particular[53, 55, 74, 75, 77, 109–112]. To some extent, the controversy is alleviated (or at least made clearer) when different sub-areas of the PFC are distinguished. The first study to show evidence for the maintenance of a continuous visual feature over the working memory maintenance period in PFC focused on an area commonly referred to as the dorsolateral PFC (DLPFC)[78]. DLPFC encompasses the cytoarchitectonic areas around the principal sulcus in monkeys and along the middle frontal gyrus in humans (areas 46 and 9/46 in the scheme of Petrides and Pandya[113])[77, 114]. Caudal to this region are the frontal eye fields (FEF), where neurons similarly show tuned, elevated activity during the working memory delay period[78]. FEF is defined as the zone in the PFC from which electrical stimulation elicits eye movements, near the arcuate sulcus in monkeys and around the precentral sulcus in humans[115]. And rostral to the DLPFC, still on the lateral convexity, is a third zone that can be delineated anatomically and functionally, which I refer to as anterior PFC (aPFC)[116] and which corresponds to area 10 in both monkeys and humans[113]. A number of neuroimaging and lesion studies support (approximately) this tripartite division of lateral PFC[116–124].

Of these sub-areas, the DLPFC is the one that has classically been viewed as necessary for mnemonic performance on delayed response tasks, based primarily on data from non-human primates[74, 75, 77]. These data inspired the conclusion that DLPFC lesions cause "mnemonic scotomas"[125] – impairments in working memory localized to patches of the visual field – as measured using the oculomotor delayed response (ODR) task, a test of spatial working memory (Fig. 3.1). Subsequent work, however, has cast doubt on this interpretation. On the one hand, a reanalysis of data from monkeys has underscored the importance of distinguishing stages of the behavioral response in the ODR task, which earlier studies did not do[109]. At the end of each trial, there is the initial memory-guided

saccade (MGS), which reflects the prepared motor plan, and then subsequent corrections
to this primary response that determine the final eye position (FEP) before feedback is
given[126]. Disruption of activity in monkey DLPFC impairs the MGS[125, 127–129], but
appears to have little effect on FEP[109]. A recent experiment using human participants, on
the other hand, suggests that even this more limited involvement of the DLPFC in working
memory may be specific to monkeys, as humans with DLPFC lesions have intact MGS and
FEP components of the memory-guided response[130]. Lesions to putative human FEF, in
contrast, cause an impairment in the MGS similar to that observed in monkeys with DLPFC
lesions[130]. This pattern of effects in humans has since been replicated using transcranial
magnetic stimulation (TMS) to induce temporary virtual lesions in healthy subjects[126].
The aPFC has largely been left out of this debate, in part, perhaps, because it is notoriously
difficult to access surgically for experiments in monkeys[131].

In the present study, I investigate the causal involvement of all three of these lateral
PFC areas – FEF, DLPFC, and aPFC – in serial dependence, the temporal smoothing
operation applied by the brain to working memory representations. I analyze datasets from
two separate experiments. These experiments employed distinct TMS protocols to disrupt
activity in PFC, with one using a brief train of pulses applied at the center of the delay period
of each trial and the other using the continuous theta-burst stimulation (cTBS) procedure to
induce a longer-lasting reduction in cortical excitability[132] prior to testing. Furthermore,
participants in the two experiments completed different tasks in which not only the stimulus
types were distinct, but the perceptual decision required to make a response at the end of
the trial was also variant. One dataset was acquired while subjects completed the ODR
task, which requires a continuous response – an eye movement away from fixation that is
measured in polar angle (0-360°). For these data I distinguished errors in the MGS from
those in the FEP. The other dataset came from a two-alternative forced choice (2AFC) task
in which subjects indicated via a binary button press whether a previously seen orientation
was tilted clockwise or counterclockwise from vertical. Since this task involves a single,
final decision at the end of each trial, I equate the response on this task to the FEP in
the ODR task. A no-TMS baseline of serial dependence was measured for responses from
both tasks, and, in addition to being applied to the prefrontal nodes, TMS was applied in
separate sessions to areas in parietal cortex as a control. I hypothesized that TMS to PFC
would reduce serial dependence in behavior relative to the control conditions – reasoning
(per Chapter 2) that signals in PFC may be responsible for this trial-history effect[32]. The
only previously published data that speak to this possibility are from an anecdotal report
that GABA blockade in DLPFC (which drives up network activity[133]) causes an extreme
form of serial dependence (perseveration) in the MGS component of responses on the ODR
task[109].

Results

Part 1: Cortical support for serial dependence in the continuous ODR task

Nine neurologically healthy adult participants (two female) completed the version of the ODR task depicted in Figure 3.1. Results from this dataset (unrelated to serial dependence) have been published previously[126]. Each trial began with the presentation of a yellow dot at a random angle from fixation. Participants were instructed to remember the location of this circle across a delay period (3-5 s). At the end of the delay period, a sound coupled with the disappearance of the fixation square signaled the start of the response period. Participants made an initial MGS to report the location of the cue. They were then allowed to adjust their gaze within the 800-ms response window before assuming their FEP. Error in both the MGS and the FEP was measured in degrees of polar angle. After the response period, the target reappeared on the screen (but colored green instead of yellow) as feedback. Subjects were trained to shift their gaze as necessary to fixate on this target. For each TMS condition, each participant completed 300 trials divided into 10 runs over the course of one experimental session.

Seven of the subjects completed four total sessions: one without TMS, and then three with TMS to either FEF, DLPFC, or the control site in posterior parietal cortex (PPC). (By referring to PPC as a control site, I do not mean to imply that it is not important for working memory. Indeed, lesions to PPC have been shown to cause impairments in both the MGS and FEP on the ODR task[134]. In fact, in the current TMS dataset, these effects have been replicated[126]. My goal here is not to expose general impairments in working memory, however, but to measure changes in the magnitude of serial dependence, a particular kind of systematic error that I hypothesize may be a consequence of synaptic dynamics specific to prefrontal networks, where augmentation is prominent[86, 87]. Since the PPC is likely crucial to general working memory maintenance, using it as a control in this analysis represents an especially strong test of my hypothesis. A null result for PPC in the context of the systematic bias of serial dependence would imply that the increase in errors under PPC stimulation found in other analyses[126] is likely due to a non-systematic increase in random noise.) TMS was not applied to aPFC in this experiment. Of the remaining subjects, one completed just the FEF TMS session, and the other completed just the DLPFC and PPC sessions. This resulted in a sample of eight participants for each of the three TMS conditions and nine for the no-TMS condition. In all TMS sessions, a brief train of three pulses was applied at the center of the delay period of every trial during behavioral testing – under the assumption that this would disrupt ongoing neural firing in the targeted cortical area.

First, I examined the pattern of subjects' errors in the MGS. I confirmed that serial dependence was present in this dataset in the no-TMS condition at levels significantly greater than zero (group permutation test, $p < 10^{-4}$). For this analysis, I measured the magnitude of serial dependence as the peak-to-peak of the Clifford model[20] (see Methods) fit to the

behavioral responses, as I did in Chapter 1. No such bias was present in the data in the
direction of the stimulus on the upcoming trial ($n.s.$, group permutation test), which upholds
my conclusion that the dependence of behavior on the previous trial is not due to spurious
correlations in the particular randomized sequences of stimuli generated for the subjects[10,
14]. The observed tuning of serial dependence in the MGS across the range of possible
differences between past and current visual input is displayed in Figure 3.2A.

Next, I examined whether the amplitude of the serial dependence effect was altered in each
of the TMS conditions, again focusing on the initial MGS (Fig. 3.2B). In the PPC condition,
serial dependence was significantly greater than zero ($p < 10^{-4}$, group permutation test) but
did not differ in magnitude from the no-TMS condition ($n.s.$), in support of my predictions.
Serial dependence was also present in the two prefrontal TMS conditions (both $p < 10^{-4}$).
In the FEF condition, the magnitude of the effect was no different from either the no-TMS
baseline or the PPC TMS condition (both comparisons $n.s.$), though it was numerically
reduced. The depression of serial dependence was greater in the DLPFC condition than
in the FEF condition, such that the contrast with the no-TMS baseline was significant
($p = 0.002$) as was that with the PPC control ($p = 10^{-4}$). Thus, TMS to both prefrontal
nodes produced an effect on the MGS in the predicted direction, but that was significant
only for the DLPFC site. The magnitude of serial dependence did not differ between the
two prefrontal conditions ($n.s.$).

I repeated these analyses using the FEP as the dependent variable. First, I checked
whether serial dependence in the FEP differed from that obtained for the MGS in the
absence of TMS, and found that the extent of temporal smoothing in the two response
components was equivalent ($n.s.$, group permutation test). Although serial dependence has
previously been demonstrated in the ODR task[16, 17], the MGS and FEP have not been
compared with respect to this trial-history bias, so this represents a novel finding (though
the result is null). As for the MGS, serial dependence was present in the FEP dataset for the
no-TMS condition at levels significantly greater than zero ($p < 10^{-4}$). Once again, no bias
was present in the data in the direction of the stimulus on the upcoming trial ($n.s.$), further
dismissing the possibility of spurious correlations in the trial sequence[10, 14]. Figure 3.3A
depicts the tuning of the serial dependence effect for the FEP across all stimulus differences.

Although some data indicate that the targeting of the MGS and the FEP is set via
dissociable neural systems[126, 130, 134], my neural hypothesis for serial dependence does
not distinguish between these two elements of the response. Hence, I predicted that TMS to
the prefrontal nodes would reduce the amplitude of serial dependence in the FEP, as in the
MGS, relative to the no-TMS and PPC control conditions. The results for the different TMS
conditions are displayed in Figure 3.3B. When TMS was applied to PPC, serial dependence
remained significantly greater than zero in the FEP ($p < 10^{-4}$, group permutation test).
There was no difference between this condition and the no-TMS baseline ($n.s.$). This result
mirrors what I observed for the MGS. When TMS was applied to the two prefrontal sites,
serial dependence remained present in the FEP component of the response (both $p < 10^{-4}$),
but the magnitude of the effect was numerically weakened relative to both the no-TMS and
PPC controls. This time, the reduction for FEF was greater than that for DLPFC, such that

the magnitude of serial dependence under FEF stimulation was significantly smaller than in
the PPC control condition ($p = 0.0005$) and marginally depressed relative to the no-TMS
condition ($p = 0.01$, Bonferroni-corrected $\alpha = 0.008$). Results for DLPFC differed neither
from the control conditions nor from the results for FEF (all comparisons $n.s.$). Overall, my
findings for the MGS and the FEP were in the same predicted direction, but I observed that
DLPFC appeared to contribute more to serial dependence in the initial memorized motor
plan[109], and FEF more to serial dependence in the ultimate fine-tuned perceptual decision.

Part 2: Cortical support for serial dependence in the binary 2AFC task

Having obtained preliminary support for my hypothesis that serial dependence originates
in PFC, I sought to replicate these results in an independent dataset. The protocol used
to collect this second dataset contrasted sharply with that from Part 1 – in terms of TMS
methodology, stimulus presentation, and requirements for the behavioral response – which
implies that any consistency observed in the effects of TMS cannot be due to superficial
features of the experimental design. For Part 2, 17 adult participants (11 female) completed
a 2AFC perceptual decision making task, depicted in Figure 3.4. Results from this dataset
(unrelated to my hypotheses) have been published previously[116]. On each trial of the task,
two oriented gratings were presented to the left and right of fixation. The gratings were tilted
independently of each other at either 45 or 135°, and this tilt was updated randomly from
trial to trial. Upon the offset of the stimuli, a cue was presented that instructed participants
to report the orientation of one of the gratings by pressing one of two buttons (for clockwise or
counterclockwise). Hence, the recorded error was binary for each trial (correct or incorrect).
No feedback was given. In each testing session, every participant completed four runs, with
four blocks per run, for a total of 480 trials.

Participants first completed a no-TMS baseline behavioral testing session, which occurred
in the fMRI scanner. In subsequent days, they underwent four TMS sessions – in which TMS
was applied to the three prefrontal nodes of interest (FEF, DLPFC, and aPFC) and to a
control site, again in parietal cortex. This time, primary somatosensory cortex (S1) was
used as the control node. (Unlike PPC, S1 is not thought to be crucial for the maintenance
of representations in visual working memory.) TMS was applied using the cTBS procedure:
Participants received five bursts of three 50-Hz pulses every second for 40 s at the start of
testing. This type of TMS stimulation likely reduces cortical excitability[132].

Because only two discrete orientations were tested, rather than the full range of possible
angles, I was unable to extract the tuning of serial dependence as in Figures 3.2A and 3.3A.
Instead, I quantified serial dependence using logistic regression, as has been done previously
with binary responses[107]. My analyses consider only stimuli that were cued for response
(only one of the orientations on each trial), as the angle of the uncued stimulus was not
recorded in the raw dataset. First, I conducted a preliminary analysis to determine the trial
types in which serial dependence occurred in the absence of TMS (during the preliminary

fMRI session) and with S1 stimulation. Consistent with published findings[107], I found
that in both of these control conditions, binary representations of orientation were subject
to serial dependence only when consecutive stimuli occurred at the same location relative
to fixation (both $p = 0.03$, one-sample t-tests). When consecutive stimuli were on opposite
sides of fixation, the effect was null (both $n.s.$). No differences in the magnitude of this
effect between the no-TMS and S1 condition were noted, supporting my interpretation of S1
stimulation as a control condition ($n.s.$, paired-sample t-tests). These results are depicted
in Figure 3.5A.

A previous study has also observed that even at the same location, serial dependence
only occurs when the response to the first stimulus is correct (i.e., when the stimulus is suc-
cessfully encoded)[107]. I separately modeled correct and incorrect trials in a second logistic
regression as an attempt to replicate this effect in my control conditions. The replication
was successful (Fig. 3.5B). Serial dependence was strongly positive in both the no-TMS and
S1 conditions when the previous response was correct (both $p < 10^{-5}$, one-sample t-tests).
Following incorrect trials, repulsion was observed – an effect in the direction opposite se-
rial dependence (both $p < 10^{-5}$). No differences between the two control conditions were
noted (all comparisons $n.s.$, paired-sample t-tests). The repulsion after incorrect trials could
be taken as evidence that subjects tended to form an erroneous percept on these trials –
encoding clockwise when they should have encoded counterclockwise, for example – and
that responses were drawn toward this percept rather than toward the actual presented
stimulus[107]. In effect, this would be the same psychological process as attractive serial
dependence after correct trials. However, the repulsive effect could also imply that when
subjects failed to attend to the previous trial's stimulus and had no representation in mind
for it at all, repulsive sensory adaptation dominated – a phenomenon distinct from serial
dependence that appears to be independent of attention[1, 9].

Hence, the one type of trial pair for which serial dependence unambiguously occurs in
this task is that in which both cued stimuli were presented at the same location, and in
which the response to the first of these stimuli was correct. I used this trial type to examine
the effects of TMS. Again, my hypothesis is that TMS to frontal nodes should decrease serial
dependence relative to the control TMS condition. Because S1 stimulation is closer to a null
control than PPC stimulation (given the latter's importance to visual working memory and
attention), and because I found no differences between S1 TMS and the no-TMS baseline
in my preliminary analyses, I discarded the no-TMS condition from this second stage of the
analysis. This allowed us to boost statistical power in light of my having dropped a large
number of trials on which serial dependence does not occur. I performed just the planned
comparisons between the different prefrontal TMS conditions and the S1 control. For these
comparisons, my general prediction was confirmed (Fig. 3.6). Numerically, TMS to all
three frontal nodes caused a reduction in serial dependence relative to S1 condition – with
serial dependence measured as the regression weight for the influence of the previous trial's
stimulus on the current trial's response. This reduction was significant for TMS to FEF
($p = 0.009$, paired-sample t-test) and to aPFC ($p = 0.01$), but not to DLPFC ($n.s.$). For
all three prefrontal TMS conditions, serial dependence was significantly greater than zero

(all $p < 0.0005$, one-sample t-tests). Importantly, the effects of TMS were specific to the serial dependence phenomenon – the regression weight for the influence of the current trial's stimulus on the current trial's response did not differ among the conditions (all comparisons $n.s.$, paired-sample t-tests).

Discussion

Serial dependence has been identified as a fundamental feature of the visual working memory system, in that it operates across disparate feature dimensions and smooths experience in such a way as to promote its continuity and stability[32]. Recent work has made clear that the serial dependence phenomenon is not specific to any particular experimental paradigm, but arises regardless of whether human observers are asked to reproduce the contents of memory[1, 2, 9, 14], make a binary forced choice[1, 2], or even make a subjective judgment that is only abstractly related to perceptual features[3]. Yet, in spite of this progress made in the psychophysics of serial dependence, the neural basis of this temporal smoothing operation has until now remained elusive.

Inspired by behavioral studies that have linked serial dependence to attention and working memory[1, 9, 16], as well as theoretical modeling that has shown the effect could be a consequence of synaptic augmentation in prefrontal networks[86, 87] – that is, Chapters 1 and 2 of this dissertation – I attempted to demonstrate, using TMS, that the PFC makes a causal contribution to this trial-history effect. To this end, I analyzed datasets in which serial dependence could be measured reliably, but which otherwise came from studies with as many methodological discrepancies as could be achieved: The experiments in Parts 1 and 2 tested different types of stimuli, required different perceptual decisions, taxed different motor output systems, and applied TMS using different protocols. Nevertheless, the results I obtained were remarkably consistent. For Part 1, subjects completed the ODR task, which requires a behavioral response split into two parts: an initial MGS (which represents the motor plan maintained during the working memory delay period) and rapid subsequent corrections that result in the FEP assumed just before feedback. I found that TMS to DLPFC significantly disrupted serial dependence in the MGS, whereas TMS to FEF affected the FEP. In the task used for Part 2, participants executed a unitary final response at the end of each trial that was categorical in nature rather than continuous. I view the finality of this decision as analogous to the FEP in the ODR task. In line with this interpretation, it was TMS to FEF that caused a reduction in the trial-history effect. Furthermore, TMS to aPFC, an area that was not tested in Part 1, also resulted in reduced temporal smoothing relative to control stimulation. I note that in every prefrontal TMS condition analyzed, serial dependence was numerically smaller in magnitude than in the corresponding control conditions (in which TMS was delivered to parietal cortex or not delivered at all).

While the general pattern of results matched my predictions, I did not enter this analysis with the expectation that FEF and DLPFC would differ with regard to their relationship to the initial and final stages of response execution. More research is needed to determine

whether this aspect of my results is due to mere statistical fluctuations in the finite datasets I analyzed or to a genuine difference between these sub-areas of PFC. It does seem that the MGS and FEP stages of the memory-guided oculomotor response arise from distinct mechanisms. For example, they develop at different rates across the lifespan[135], are differentially affected in schizophrenia[136], and have been found to undergo different rates of adaptation over longer timescales than measured in the current experiment[137]. Much is unknown, however, about the neural pathways that support these distinctions. Future work is needed as well to clarify the role of the aPFC in serial dependence. This area has previously been linked to "confidence leak" – the extent to which confidence in a perceptual report made on one trial biases the confidence felt on future trials[116, 138]. My results suggest that aPFC influences not just temporal leak in metacognition, but in the maintained perceptual representations themselves. One possibility is that this influence on the contents of working memory is what drives the aPFC's impact on the confidence the subject feels about whether that content matches reality.

My findings complement an anecdotal report that a manipulation of PFC activity in the opposite direction – amplifying network activity via GABA blockade[133] – causes monkeys to direct their MGS towards the stimulus location from the previous trial (an exaggerated form of serial dependence)[109]. Several circuit mechanisms could be responsible for this effect, and it is as yet unclear whether this kind of perseveration is equivalent to the serial dependence studied in my experiment. Neural firing in PFC causes increases in synaptic weights due to augmentation as well as other short-term synaptic dynamics[86, 87], and, as argued in Chapter 2, these weight changes are one candidate for the signal that persists across trials to make serial dependence possible[32]. Other authors have noted that endocannabinoid signaling in PFC could exert similar influences on behavior[139]. Intrinsic timescales in PFC are longer than in other cortical areas[140], which may help instantiate dependencies between activity patterns on consecutive trials. These dependencies could serve to facilitate the monitoring and manipulation of temporal contingencies between events, an ability associated with primate intelligence that theorists have argued derives from the enlargement of the PFC across evolutionary time[141]. The serial dependence in memory-guided behavior that I and others have measured may be a PFC-dependent signature of all of these interacting factors.

Recent years have seen a number of challenges to the notion that the maintenance of information in visual working memory depends on the PFC[53, 109–111]. While my results do not resolve this debate, they do point to the potential importance of distinguishing qualitatively different types of errors in memory-guided behavior when examining the effects of PFC manipulation. For example, in the dataset I analyzed for Part 1 of this study, TMS to DLPFC caused no overall increase in errors for the MGS, and TMS to FEF did not alter raw errors in the FEP[126]. However, when the systematic bias associated with serial dependence was isolated, deficits were found. This implies that the null result for undifferentiated error rates hides opposite effects of TMS on noise (which increased) and serial dependence (which decreased). Similarly, none of the TMS conditions in Part 2 differed from each other with regard to the response's dependence on the current stimulus (accuracy). It was only the bias

in the direction of the preceding stimulus that distinguished them. Unless these nuances in behavior are modeled explicitly, null results after PFC disruption must be interpreted with caution.

Finally, I propose that my findings highlight the potential for serial dependence to act as a fulcrum for differentiating separate contributions to working memory from areas throughout association cortex. For years, it has been known that several cortical areas display persistent activity during the working memory delay period[112]. In particular, activity in PPC is often indistinguishable from that in PFC[142]. However, just as careful analysis of behavior after TMS to these areas can reveal previously unappreciated differences, I believe that a re-analysis of spiking activity with serial dependence in mind[17] may uncover subtle distinctions in the representations that these areas carry. Investigation of how serial dependence propagates among perceptual, association, and motor cortices may yield insights not just into this particular temporal smoothing phenomenon, but into much broader and long-standing questions about cortical organization and function.

Methods

The methods used for the acquisition and preprocessing of the two datasets included in the current analysis – from the ODR task[126] and the 2AFC task[116] – have already been reported in detail. Here, I briefly summarize these procedures and report in full the novel aspects of my approach. All of the Python, MATLAB, and bash code written for this study is available upon request.

Participants

Nine individuals (two female) participated in Part 1 (the ODR task). Data collection was conducted in accordance with the guidelines of the Institutional Review Board at NYU. All nine subjects completed a baseline session of the task without TMS. Seven of them completed additional sessions for the three TMS conditions (DLPFC, FEF, and PPC). One of the remaining subjects completed only the FEF TMS session, and the last completed only the DLPFC and PPC TMS sessions. This yielded eight datasets for each of the three TMS conditions.

For Part 2 (the 2AFC task), data from 17 subjects were available for analysis (11 female). All procedures were approved by the UC Berkeley Committee for the Protection of Human Subjects. The subjects completed four TMS sessions (to S1, FEF, DLPFC, and aPFC) after completing a baseline testing session without TMS in the fMRI scanner.

ODR Task (Part 1)

Participants completed the task in a darkened room and used a chin rest to eliminate head movement. Stimulus presentation was programmed in MATLAB using the MGL toolbox.

The stages of the task are depicted in Figure 3.1. In the description that follows, all angle measurements are reported in degrees of visual angle. All stimuli were displayed against a gray background. While the participants fixated a black circle at the center of the screen, a yellow target (0.5° in diameter) appeared for 200 ms at a random location approximately 10° in eccentricity from fixation. The target never appeared close to the cardinal axes, but otherwise its angle around fixation was unconstrained. Participants remembered the location of the dot for a delay period that varied randomly from trial to trial (3, 3.5, 4, 4.5, or 5 s). At the end of this delay period, a sound coupled with the disappearance of the fixation point signaled that participants should shift their gaze to the location at which the target had been presented. Subjects were allotted 800 ms to respond, after which the target reappeared on the screen for 700 ms (colored green rather than yellow) as feedback. The subjects were instructed to fixate this feedback. Finally, a blue circle appeared at the center of the screen for 1.5 s (the intertrial interval).

Each participant completed 10 runs per session, and each run comprised 30 trials. During every run, monocular eye-movement data were collected at 1000 Hz using an SR Research EyeLink 1000 eye-tracker. Nine-point calibrations were performed at the beginning of each session, as well as between runs as necessary. The eye-movement data were transformed into degrees of visual angle using a third-order polynomial algorithm that fit eye positions to known spatial locations and were then scored offline using the iEye MATLAB toolbox[126]. Trials where the saccadic response time exceeded 900 ms or occurred within 100 ms were discarded. In addition, trials in which the subject broke fixation before the response screen were discarded, as were trials in which the response failed to exceed 5° eccentricity from fixation.

2AFC Task (Part 2)

During the baseline fMRI session without TMS, the participants viewed the task stimuli on a screen mounted to the scanner's rf coil, onto which light from an LCD projector was back-projected. The overhead lights in the scanner room were turned off. For the TMS sessions, the protocol was completed in a darkened room while participants were seated at a computer display. The task was written in MATLAB using the Psychophysics Toolbox[69].

Figure 3.4 depicts the stages of the task. Again, angles are reported here in degrees of visual angle. Participants were trained to fixate a small central square throughout the experiment. Each trial began with the presentation of a precue that indicated with 66.67% accuracy which of the two upcoming stimuli would be probed at the end of the trial. Simultaneous with the precue was an instruction to emphasize either speed or accuracy in the upcoming response. To maximize power in my statistical analyses, I collapsed over these conditions, which were included to address hypotheses unrelated to mine. Other authors investigating serial dependence have similarly collapsed over precue conditions[107]. While the attentional and speed/accuracy instructions remained on the screen, two stimuli were presented for 200 ms. The stimuli were gray-scale gratings (3° in diameter, 0.5 cycles per degree) displayed 9° to the left and right of fixation, superimposed on a noisy background

of uniformly distributed intensity values (8% contrast). Each grating was tilted either 45 or 135° from vertical. When the presentation period ended, the stimuli were replaced with a single white circle (4° in diameter) positioned at the location of one of the gratings. The circle was sized so as not to induce backward masking. Participants indicated the orientation of the probed stimulus by pressing one of two buttons. They were then asked to indicate their confidence in the correctness of their response using a four-point scale, from 1 (low confidence) to 4 (high confidence). The confidence reports were ignored for the present analysis. No feedback was provided during the experimental sessions.

On each day of testing, the subjects completed four runs, each of which comprised four blocks of 30 trials (480 trials total per session). All experimental variables were fully counterbalanced, including the orientation of the cued stimulus on each trial. The orientation of the uncued stimulus was chosen randomly and independently of all other trial variables. The orientations of the uncued stimuli were not recorded during data collection and were not recoverable for the present analysis.

Cortical Area Definitions for TMS

For Part 1, TMS was applied to DLPFC, FEF, and PPC in separate sessions. The DLPFC was defined using anatomical landmarks, with the understanding that cytoarchitectonic areas 46 and 9/46 of the mid-DLPFC fall roughly along the middle frontal gyrus[114]. FEF was defined using nonlinear population receptive field (pRF) mapping[143], conducted based on an fMRI session that occurred before the TMS sessions. This technique identifies three distinct spatial maps in the vicinity of the precentral sulcus, which have been labeled superior PCS1 (sPCS1), sPCS2, and inferior PCS (iPCS)[143]. TMS was directed to the sPCS maps for the FEF condition, as this general area is considered to be homologous to monkey FEF – i.e., electrical stimulation to this location in humans has been shown to elicit eye movements[144]. Nonlinear pRF mapping reveals four separate maps of visual space in PPC[143]. TMS was applied to the third of these maps – termed intraparietal sulcus area 2 (IPS2). This area is regarded as a potential homolog of the monkey lateral intraparietal area (LIP)[126]. TMS was applied to right DLPFC, left FEF, and left PPC.

Four areas were targeted with TMS in Part 2 of this study: S1, FEF, DLPFC, and aPFC. S1 was localized anatomically along the postcentral gyrus. FEF and DLPFC were identified based on fMRI activation during performance of the 2AFC task during the no-TMS baseline session. The statistical contrast used to define these areas was task > background, and $\alpha = 0.001$ was taken as the threshold for activation. "Task" in this contrast was defined as the average of four regressors: the two possible attentional cues crossed with the two possible speed/accuracy instructions. FEF was defined as the site of maximum signal near the junction of the superior frontal sulcus and the ascending limb of the precentral sulcus[115]. DLPFC was defined as the peak in the midlateral PFC[145]. When multiple sites of activation were observed near putative DLPFC, the one on the middle frontal gyrus was chosen. I note that the criteria used for area definitions in Part 2 are not identical to those used in Part 1, though the two schemes are effectively very similar. Both sets of authors

were aiming for the same areas – in terms of their putative functional roles and anatomical organization – so I refer to them using identical terminology (i.e., FEF and DLPFC) in this report. aPFC was defined using the coordinates of a previously published region of interest rostral to DLPFC associated with metacognitive judgments[146]. In all conditions, TMS was applied to the right hemisphere.

TMS Procedures

Parts 1 and 2 employed different TMS protocols. For Part 1, TMS was administered with a Magstim Rapid 2 Magnetic Stimulator with a figure-eight coil (70-mm diameter). Stimulation was applied online, at the center of the delay period of each trial, as a train of three pulses at 50 Hz. The intensity was set to 53% of the maximum stimulator output for each subject.

TMS for Part 2 was delivered using a Magstim Super Rapid Stimulator connected to two booster modules, again with a figure-eight coil (70-mm diameter). The stimulation occurred prior to testing, using the cTBS procedure, which is theorized to reduce cortical excitability[132]. Five bursts of three 50-Hz pulses were delivered every second for a total of 600 pulses over 40 s. Stimulation was delivered at 80% of the individual participant's motor threshold.

Analysis of Serial Dependence for Continuous Data (Part 1)

In order to test the responses from Part 1 for serial dependence, I first removed systematic directional errors from the data, as was done in Chapter 1[16, 17]. I computed this systematic error as the mean response for each stimulus location. This mean was then subtracted from the response on each individual trial (ignoring the location of the previous trial) to obtain the residual error that was then used to characterize serial dependence. The mean response for each location was computed by spatially low-pass filtering the responses as a function of stimulus location using the MATLAB function `loess`[16, 17].

I used the Clifford model[20] to characterize the tuning of serial dependence across all possible differences between past and current visual input, as was done in Chapters 1 and 2. The model is stated as follows:

$$\sin(\theta_A) = \frac{\sin(\theta_0)}{\sqrt{(s\cos(\theta_0) - c)^2 + \sin^2(\theta_0)}}, \tag{3.1}$$

where θ_A is the perceived location (in polar angle) of the current stimulus, θ_0 is its true location, s is a scaling parameter, and c is a centering parameter. Both θ_A and θ_0 are expressed relative to the previous trial's true location. I used the scipy[70] function `least_squares` (in the `optimize` module) to find the values of c and s that minimized the difference, for each θ_0, between the estimated θ_A and the subjects' errors. Across all values of θ_0, I took the magnitude of serial dependence to be the peak-to-peak of $\theta_A - \theta_0$. This peak-to-peak depends on both c and s.

To determine whether the magnitude of serial dependence was significantly greater than zero, or greater in one condition than in another, I submitted the data to permutation testing at the group level, as in Chapter 1[2, 9]. Specifically, I shuffled the values of θ_0 (current trial's location relative to the previous trial's) while leaving in place the corresponding errors. I then fit the Clifford model to the shuffled dataset. This process was repeated 10,000 times. As p-values I report the proportion of permutations that led to equal or higher values for the Clifford peak-to-peak than the one estimated from the unshuffled data. When comparing TMS conditions, I subtracted the null (shuffled) peak-to-peaks for one condition from those for the other, and report the proportion of these differences that had equal or higher values than the empirical difference. The criterion for significance was Bonferroni-corrected for each family of tests.

Analysis of Serial Dependence for Binary Data (Part 2)

For Part 2, I assessed serial dependence using logistic regression, as has been done previously for this kind of data[107]. First, I performed logistic regression with three regressors: one for the influence of the current stimulus, another for the influence of the previous trial's stimulus at the same location, and the third for the influence of the previous trial's stimulus at the other location. The dependent variable was the response on the current trial. A follow-up regression was performed when it was found (replicating previous work[107]) that serial dependence is positive only when the cued stimulus from the previous trial is at the same location as the cued stimulus on the next trial. This second analysis included just trial pairs in which the cued stimulus location stayed the same. The regressors were the current stimulus, the previous trial's stimulus when the previous trial's response was correct, and the previous trial's stimulus when the previous trial's response was incorrect. For the output of both regressions, I used simple t-tests at the group level (paired-sample and one-sample) to determine the robustness of each regressor's contribution.
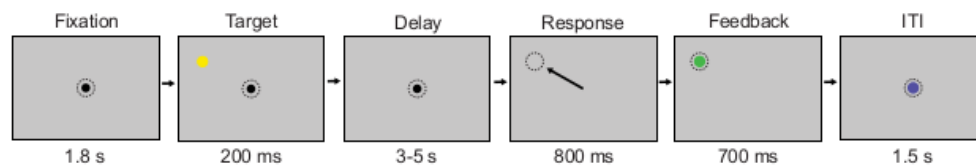
Figure 3.1: The events in each trial of the ODR task used for Part 1[126]. While fixating, the participants maintained the position of a briefly displayed visual target over a working memory delay period, and then made a saccade to the remembered location. The target was presented again after the response (in a different color) for feedback. Dotted circles depict gaze in a hypothetical trial and were not visible to the subjects.
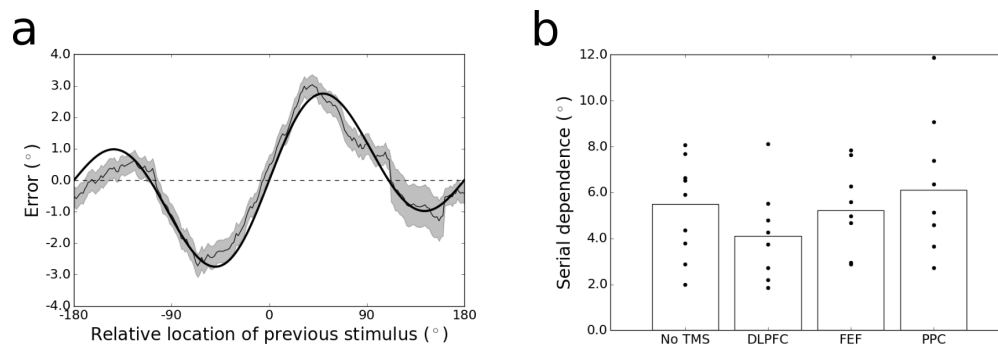
Figure 3.2: MGS results for the ODR task from Part 1. (A) Tuning of serial dependence in the MGS across all possible angular differences between the current and previous stimulus, for the no-TMS condition. The thin black line represents the group moving average of saccade errors (with the standard error in gray shading) and the thick black line is the best-fitting Clifford model curve. Serial dependence is present in the data, detected as an amplitude of the Clifford fit greater than expected by chance (per a group permutation test). (B) Peak-to-peak of the Clifford model fit to the group MGS data for each TMS condition in Part 1. Dots display the peak-to-peak values for the data from individual subjects. Serial dependence is significantly depressed in the DLPFC condition relative to the two control conditions (no-TMS and PPC).
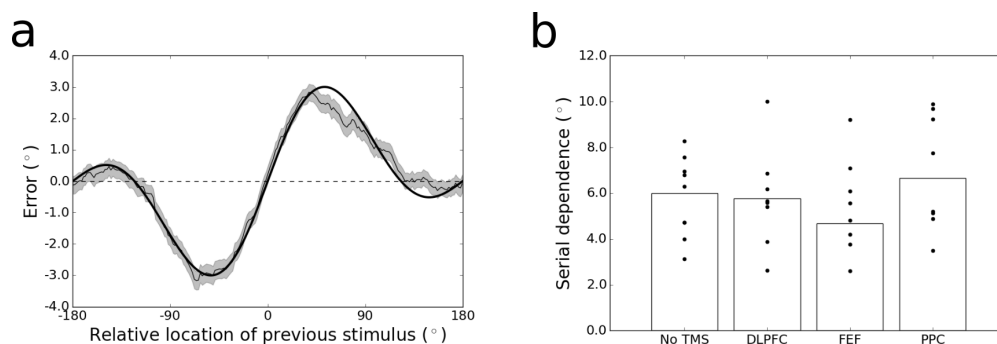
Figure 3.3: FEP results for the ODR task from Part 1. (A) Tuning of serial dependence in the FEP across angular differences between the current and previous stimulus, for the no-TMS condition. The thin black line represents the group moving average of FEP errors (with the standard error in gray shading) and the thick black line is the best-fitting Clifford model curve. As for the MGS, serial dependence is present in the data at levels greater than chance. At baseline (without TMS), the magnitude of serial dependence did not differ between the MGS and the FEP. (B) Peak-to-peak of the Clifford model fit to the group FEP errors for each TMS condition in Part 1. Dots display the peak-to-peak values for the fits to individual subjects' data. Serial dependence is significantly reduced in the FEF condition relative to the two control conditions (no-TMS and PPC).
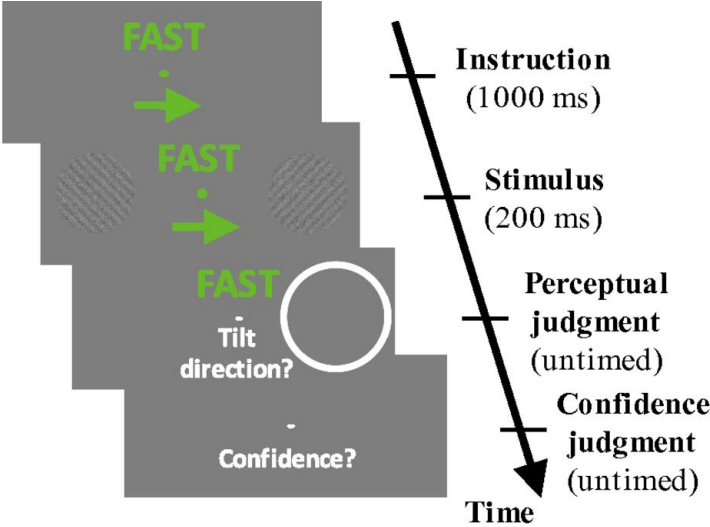
Figure 3.4: Events in a sample trial of the 2AFC task used for Part 2[116]. Each trial began with a pair of instructions, displayed for 1 s. One instruction was to attend to either the left or right stimulus, and the other was to emphasize either speed or accuracy in the response. These instructions were ignored for the present analyses, and were included in the task design to address hypotheses unrelated to serial dependence. Two oriented gratings were then presented for 200 ms, after which a postcue indicated the stimulus that subjects should report. The response was made as a binary button press (clockwise or counterclockwise). Subjects were given as much time as needed to respond. After making their perceptual report, participants indicated their confidence in it, on a scale from 1 to 4. This rating was ignored for the present analysis. The next trial began after a 1-s intertrial interval.
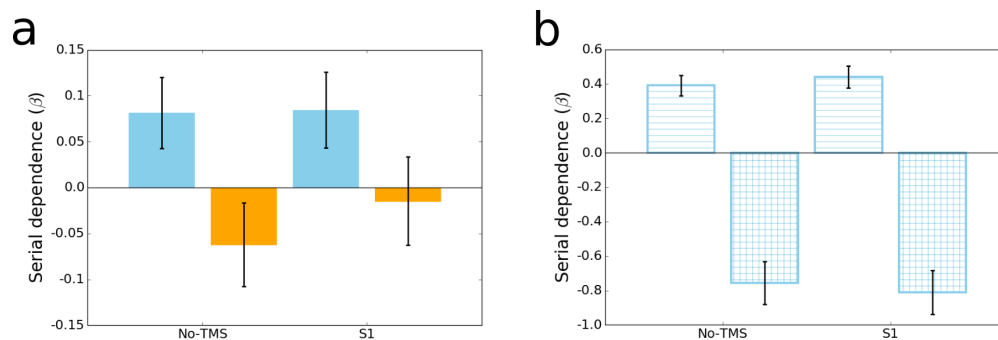
Figure 3.5: Logistic regression results for the control conditions of the 2AFC task used for Part 2. (A) Parameter estimates for the influence of the previous stimulus at the same location (blue) and at the other location (orange) on the current trial's response for the no-TMS and S1 TMS conditions. Equivalently in both conditions, serial dependence occurs only for the previous stimulus at the same location. (B) Parameter estimates for the follow-up regression testing the influence of the previous stimulus at the same location when the response to that stimulus was correct (horizontal lines) or incorrect (cross-hatching). Again equivalently for both conditions, positive serial dependence occurs only for the previous stimulus at the same location when the response to that stimulus was correct. Error bars in (A) and (B) indicate the standard error across subjects.
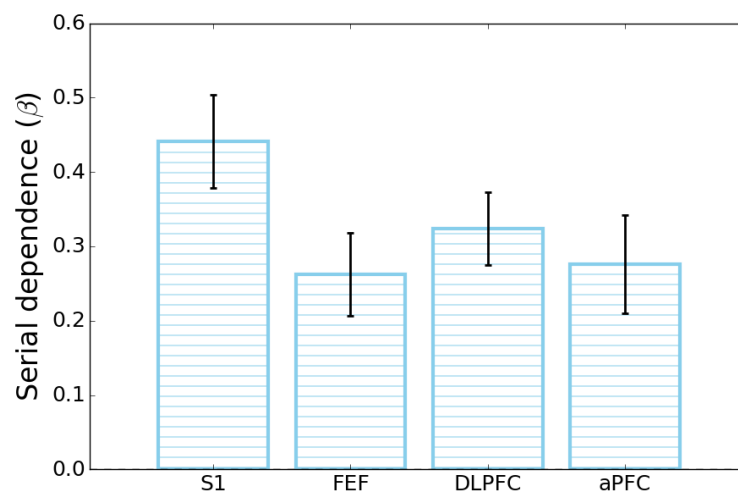
Figure 3.6: Logistic regression results for the TMS conditions of the 2AFC task used for Part 2. Parameter estimates for the regression testing the influence of the previous stimulus at the same location when the response to that stimulus was correct. This is the only trial type in this task that reliably induces serial dependence[107]. (I use the same color/pattern scheme for the bars here as in Figure 3.5, and the bar for S1 in this plot is the same as the corresponding bar in that figure.) Serial dependence is significantly reduced in the FEF and aPFC conditions relative to the S1 control condition. Error bars indicate the standard error across subjects.

References

1. Fischer, J. & Whitney, D. Serial dependence in visual perception. *Nature Neuroscience* **17,** 738–743 (2014).

2. Liberman, A., Fischer, J. & Whitney, D. Serial dependence in the perception of faces. *Current Biology* **24,** 2569–2574 (2014).

3. Xia, Y., Leib, A. Y. & Whitney, D. Serial dependence in the perception of attractiveness. *Journal of vision* **16,** 28–28 (2016).

4. Huang, J. & Sekuler, R. Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision* **10,** 24–24 (2010).

5. Kondo, A., Takahashi, K. & Watanabe, K. Sequential effects in face-attractiveness judgment. *Perception* **41,** 43–49 (2012).

6. Ward, L. M. & Lockhead, G. Response system processes in absolute judgment. *Attention, Perception, & Psychophysics* **9,** 73–78 (1971).

7. Petzold, P. & Haubensak, G. Higher order sequential effects in psychophysical judgments. *Attention, Perception, & Psychophysics* **63,** 969–978 (2001).

8. Liberman, A., Zhang, K. & Whitney, D. Serial dependence promotes object stability during occlusion. *Journal of Vision* **16,** 16–16 (2016).

9. Fritsche, M., Mostert, P. & de Lange, F. P. Opposite Effects of Recent History on Perception and Decision. *Current Biology* (2017).

10. Cicchini, G. M., Anobile, G. & Burr, D. C. Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences* **111,** 7867–7872 (2014).

11. Kondo, A., Takahashi, K. & Watanabe, K. Influence of gender membership on sequential decisions of face attractiveness. *Attention, Perception, & Psychophysics* **75,** 1347–1352 (2013).

12. Ward, L. M. Mixed-modality psychophysical scaling: Inter-and intramodality sequential dependencies as a function of lag. *Attention, Perception, & Psychophysics* **38,** 512–522 (1985).

13. Petzold, P. & Haubensak, G. The influence of category membership of stimuli on sequential effects in magnitude judgment. *Perception & Psychophysics* **66,** 665–678 (2004).

14. Taubert, J., Alais, D. & Burr, D. Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports* **6** (2016).

15. Corbett, J. E., Fischer, J. & Whitney, D. Facilitating stable representations: Serial dependence in vision. *PloS One* **6,** e16701 (2011).

16. Papadimitriou, C., Ferdoash, A. & Snyder, L. H. Ghosts in the machine: memory interference from the previous trial. *Journal of neurophysiology* **113,** 567–577 (2015).

17. Papadimitriou, C., White, R. L. & Snyder, L. H. Ghosts in the Machine II: Neural Correlates of Memory Interference from the Previous Trial. *Cerebral Cortex,* bhw106 (2016).

18. Dong, D. W. & Atick, J. J. Statistics of natural time-varying images. *Network: Computation in Neural Systems* **6,** 345–358 (1995).

19. Gibson, J. J. & Radner, M. Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology* **20,** 453 (1937).

20. Clifford, C. W., Wenderoth, P. & Spehar, B. A functional angle on some after-effects in cortical vision. *Proceedings of the Royal Society of London B: Biological Sciences* **267,** 1705–1710 (2000).

21. Webster, M. A. Visual adaptation. *Annual review of vision science* **1,** 547–567 (2015).

22. Webster, M. A. & MacLeod, D. I. Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366,** 1702–1725 (2011).

23. Webster, M. A. Adaptation and visual coding. *Journal of vision* **11,** 3–3 (2011).

24. Clifford, C. W. *et al.* Visual adaptation: Neural, psychological and computational aspects. *Vision research* **47,** 3125–3131 (2007).

25. Kiyonaga, A. & Egner, T. Working memory as internal attention: toward an integrative account of internal and external selection processes. *Psychonomic bulletin & review* **20,** 228–242 (2013).

26. Chun, M. M. Visual working memory as visual attention sustained internally over time. *Neuropsychologia* **49,** 1407–1409 (2011).

27. Chun, M. M. & Johnson, M. K. Memory: enduring traces of perceptual and reflective attention. *Neuron* **72,** 520–535 (2011).

28. Gazzaley, A. & Nobre, A. C. Top-down modulation: bridging selective attention and working memory. *Trends in cognitive sciences* **16,** 129–135 (2012).

29. Awh, E. & Jonides, J. Overlapping mechanisms of attention and spatial working memory. *Trends in cognitive sciences* **5,** 119–126 (2001).

30. Awh, E., Vogel, E. & Oh, S.-H. Interactions between attention and working memory. *Neuroscience* **139,** 201–208 (2006).

31. Myers, N. E., Stokes, M. G. & Nobre, A. C. Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences* (2017).

32. Kiyonaga, A., Scimeca, J. M., Bliss, D. P. & Whitney, D. Serial Dependence across Perception, Attention, and Memory. *Trends in Cognitive Sciences* (2017).

33. White, J. M., Sparks, D. L. & Stanford, T. R. Saccades to remembered target locations: an analysis of systematic and variable errors. *Vision research* **34,** 79–92 (1994).

34. Wilken, P. & Ma, W. J. A detection theory account of change detection. *Journal of vision* **4,** 11–11 (2004).

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

36. Bays, P. M., Catalao, R. F. & Husain, M. The precision of visual working memory is set by allocation of a shared resource. *Journal of vision* **9,** 7–7 (2009).

37. Van den Berg, R., Shin, H., Chou, W.-C., George, R. & Ma, W. J. Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences* **109,** 8780–8785 (2012).

38. Fougnie, D., Suchow, J. W. & Alvarez, G. A. Variability in the quality of visual working memory. *Nature communications* **3,** 1229 (2012).

39. Van den Berg, R., Awh, E. & Ma, W. J. Factorial comparison of working memory models. *Psychological Review* **121,** 124 (2014).

40. Bays, P. M. Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience* **34,** 3632–3645 (2014).

41. Bays, P. M. A signature of neural coding at human perceptual limits. *Journal of Vision* **16,** 4–4 (2016).

42. Zhang, W. & Luck, S. J. Sudden death and gradual decay in visual working memory. *Psychological science* **20,** 423–428 (2009).

43. Bays, P. M. Evaluating and excluding swap errors in analogue tests of working memory. *Scientific reports* **6,** 19203 (2016).

44. Sligte, I. G., Scholte, H. S. & Lamme, V. A. Are there multiple visual short-term memory stores? *PLOS one* **3,** e1699 (2008).

45. Sligte, I. G., Vandenbroucke, A. R., Scholte, H. S. & Lamme, V. Detailed sensory memory, sloppy working memory. *Frontiers in psychology* **1,** 175 (2010).

46. Almeida, R., Barbosa, J. & Compte, A. Neural circuit basis of visuo-spatial working memory precision: a computational and behavioral study. *Journal of neurophysiology* **114,** 1806–1818 (2015).

47. Ploner, C. J., Gaymard, B., Rivaud, S., Agid, Y. & Pierrot-Deseilligny, C. Temporal limits of spatial working memory in humans. *European Journal of Neuroscience* **10,** 794–797 (1998).

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

49. Harrison, S. A. & Tong, F. Decoding reveals the contents of visual working memory in early visual areas. *Nature* **458,** 632–635 (2009).

50. Serences, J. T., Ester, E. F., Vogel, E. K. & Awh, E. Stimulus-specific delay activity in human primary visual cortex. *Psychological science* **20,** 207–214 (2009).

51. Postle, B. R. Working memory as an emergent property of the mind and brain. *Neuroscience* **139,** 23–38 (2006).

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

53. D'esposito, M. & Postle, B. R. The cognitive neuroscience of working memory. *Annual review of psychology* **66,** 115–142 (2015).

54. Pasternak, T. & Greenlee, M. W. Working memory in primate sensory systems. *Nature Reviews Neuroscience* **6,** 97–107 (2005).

55. Sreenivasan, K. K., Curtis, C. E. & DEsposito, M. Revisiting the role of persistent neural activity during working memory. *Trends in cognitive sciences* **18,** 82–89 (2014).

56. Gabrieli, J. D. Cognitive neuroscience of human memory. *Annual review of psychology* **49,** 87–115 (1998).

57. Fuster, J. M. Network memory. *Trends in neurosciences* **20,** 451–459 (1997).

58. Cappiello, M. & Zhang, W. A Dual-Trace Model for Visual Sensory Memory. *Journal of Experimental Psychology: Human Perception and Performance* (2016).

59. Sreenivasan, K. K., Vytlacil, J. & D'Esposito, M. Distributed and dynamic storage of working memory stimulus information in extrastriate cortex. *Journal of cognitive neuroscience* **26,** 1141–1153 (2014).

60. Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K. & Poggio, T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of neurophysiology* **100,** 1407–1419 (2008).

61. Magnussen, S. Low-level memory processes in vision. *Trends in neurosciences* **23,** 247–251 (2000).

62. Sperling, G. The information available in brief visual presentations. *Psychological monographs: General and applied* **74,** 1 (1960).

63. Sligte, I. G., Wokke, M. E., Tesselaar, J. P., Scholte, H. S. & Lamme, V. A. Magnetic stimulation of the dorsolateral prefrontal cortex dissociates fragile visual short-term memory from visual working memory. *Neuropsychologia* **49,** 1578–1588 (2011).

64. Vandenbroucke, A. R., Sligte, I. G. & Lamme, V. A. Manipulations of attention dissociate fragile visual short-term memory from visual working memory. *Neuropsychologia* **49,** 1559–1568 (2011).

65. Pinto, Y., Sligte, I. G., Shapiro, K. L. & Lamme, V. A. Fragile visual short-term memory is an object-based and location-specific store. *Psychonomic bulletin & review* **20,** 732–739 (2013).

66. Vandenbroucke, A. R., Sligte, I. G., de Vries, J. G., Cohen, M. X. & Lamme, V. A. Neural correlates of visual short-term memory dissociate between fragile and working memory representations. *Journal of cognitive neuroscience* (2015).

67. Ma, W. J., Husain, M. & Bays, P. M. Changing concepts of working memory. *Nature neuroscience* **17,** 347–356 (2014).

68. Luck, S. J. & Vogel, E. K. Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in cognitive sciences* **17,** 391–400 (2013).

69. Brainard, D. H. The psychophysics toolbox. *Spatial vision* **10,** 433–436 (1997).

70. Jones, E., Oliphant, T. & Peterson, P. SciPy: Open source scientific tools for Python. *http://www. scipy. org/* (2001).

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

72. Foulsham, T., Walker, E. & Kingstone, A. The where, what and when of gaze allocation in the lab and the natural environment. *Vision research* **51,** 1920–1931 (2011).

73. Barak, O. & Tsodyks, M. Working models of working memory. *Current opinion in neurobiology* **25,** 20–24 (2014).

74. Funahashi, S. Functions of delay-period activity in the prefrontal cortex and mnemonic scotomas revisited. *Frontiers in systems neuroscience* **9** (2015).

75. Riley, M. R. & Constantinidis, C. Role of prefrontal persistent activity in working memory. *Frontiers in systems neuroscience* **9** (2015).

76. Stokes, M. G. Activity-silentworking memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences* **19,** 394–405 (2015).

77. Curtis, C. E. & D'Esposito, M. Persistent activity in the prefrontal cortex during working memory. *Trends in cognitive sciences* **7,** 415–423 (2003).

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

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

80. Constantinidis, C. & Wang, X.-J. A neural circuit basis for spatial working memory. *The Neuroscientist* **10,** 553–565 (2004).

81. Mongillo, G., Barak, O. & Tsodyks, M. Synaptic theory of working memory. *Science* **319,** 1543–1546 (2008).

82. Barak, O., Tsodyks, M. & Romo, R. Neuronal population coding of parametric working memory. *Journal of Neuroscience* **30,** 9424–9430 (2010).

83. Brody, C. D., Hernández, A., Zainos, A. & Romo, R. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cerebral cortex* **13,** 1196–1207 (2003).

84. Shafi, M *et al.* Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience* **146,** 1082–1108 (2007).

85. Watanabe, K. & Funahashi, S. Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nature neuroscience* **17,** 601–611 (2014).

86. Hempel, C. M., Hartman, K. H., Wang, X.-J., Turrigiano, G. G. & Nelson, S. B. Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *Journal of neurophysiology* **83,** 3031–3041 (2000).

87. Wang, Y. *et al.* Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature neuroscience* **9,** 534–542 (2006).

88. Whitmire, C. J. & Stanley, G. B. Rapid sensory adaptation redux: a circuit perspective. *Neuron* **92,** 298–315 (2016).

89. Felsen, G. *et al.* Dynamic modification of cortical orientation tuning mediated by recurrent connections. *Neuron* **36,** 945–954 (2002).

90. Abbott, L. & Regehr, W. G. Synaptic computation. *Nature* **431,** 796–803 (2004).

91. Zucker, R. S. & Regehr, W. G. Short-term synaptic plasticity. *Annual review of physiology* **64,** 355–405 (2002).

92. Engel, T. A. & Wang, X.-J. Same or different? A neural circuit mechanism of similarity-based pattern match decision making. *Journal of Neuroscience* **31,** 6982–6996 (2011).

93. Wimmer, K., Nykamp, D. Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature neuroscience* **17,** 431–439 (2014).

94. Wang, M. *et al.* NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron* **77,** 736–749 (2013).

95. Constantinidis, C. & Goldman-Rakic, P. S. Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *Journal of neurophysiology* **88,** 3487–3497 (2002).

96. Wei, W. & Wang, X.-J. Downstream effect of ramping neuronal activity through synapses with short-term plasticity. *Neural computation* (2016).

97. Ardid, S., Wang, X.-J. & Compte, A. An integrated microcircuit model of attentional processing in the neocortex. *Journal of Neuroscience* **27,** 8486–8495 (2007).

98. Wong, K.-F. & Wang, X.-J. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience* **26,** 1314–1328 (2006).

99. Georgopoulos, A. P., Schwartz, A. B., Kettner, R. E., *et al.* Neuronal population coding of movement direction. *Science* **233,** 1416–1419 (1986).

100. Engel, T. A., Chaisangmongkon, W., Freedman, D. J. & Wang, X.-J. Choice-correlated activity fluctuations underlie learning of neuronal category representation. *Nature communications* **6** (2015).

101. Barak, O. & Tsodyks, M. Persistent activity in neural networks with dynamic synapses. *PLoS Comput Biol* **3,** e35 (2007).

102. Bays, P. M. & Husain, M. Dynamic shifts of limited working memory resources in human vision. *Science* **321,** 851–854 (2008).

103. Wei, Z., Wang, X.-J. & Wang, D.-H. From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. *Journal of Neuroscience* **32,** 11228–11240 (2012).

104. Bays, P. M. Spikes not slots: noise in neural populations limits working memory. *Trends in cognitive sciences* **19,** 431–438 (2015).

105. Renart, A., Song, P. & Wang, X.-J. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron* **38,** 473–485 (2003).

106. Stimberg, M., Goodman, D. F., Benichoux, V. & Brette, R. Equation-oriented specification of neural models for simulations. *Frontiers in Neuroinformatics* **8,** 6 (2014).

107. John-Saaltink, E. S., Kok, P., Lau, H. C. & de Lange, F. P. Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *Journal of Neuroscience* **36,** 6186–6192 (2016).

108. Compte, A. & Wang, X.-J. Tuning curve shift by attention modulation in cortical neurons: a computational study of its mechanisms. *Cerebral Cortex* **16,** 761–778 (2005).

109. Tsujimoto, S. & Postle, B. R. The prefrontal cortex and oculomotor delayed response: a reconsideration of the mnemonic scotoma. *Journal of cognitive neuroscience* **24,** 627–635 (2012).

110. Postle, B. R. The cognitive neuroscience of visual short-term memory. *Current opinion in behavioral sciences* **1,** 40–46 (2015).

111. Postle, B. R. How does the brain keep information in mind? *Current directions in psychological science* **25,** 151–156 (2016).

112. Leavitt, M. L., Mendoza-Halliday, D. & Martinez-Trujillo, J. C. Sustained Activity Encoding Working Memories: Not Fully Distributed. *Trends in Neurosciences* (2017).

113. Petrides, M. Comparative architectonic analysis of the human and macaque frontal cortex. *Handbook of neuropsychology* **9,** 17–58 (1994).

114. Petrides, M. Lateral prefrontal cortex: architectonic and functional organization. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **360,** 781–795 (2005).

115. Vernet, M., Quentin, R., Chanes, L., Mitsumasu, A. & Valero-Cabré, A. Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. *Frontiers in integrative neuroscience* **8** (2014).

116. Rahnev, D., Nee, D. E., Riddle, J., Larson, A. S. & DEsposito, M. Causal evidence for frontal cortex organization for perceptual decision making. *Proceedings of the National Academy of Sciences* **113,** 6059–6064 (2016).

117. Koechlin, E., Ody, C. & Kouneiher, F. The architecture of cognitive control in the human prefrontal cortex. *Science* **302,** 1181–1185 (2003).

118. Badre, D. & D'Esposito, M. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of cognitive neuroscience* **19,** 2082–2099 (2007).

119. Koechlin, E. & Summerfield, C. An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences* **11,** 229–235 (2007).

120. Badre, D. Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in cognitive sciences* **12,** 193–200 (2008).

121. Badre, D. & DEsposito, M. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature reviews. Neuroscience* **10,** 659 (2009).

122. Badre, D., Hoffman, J., Cooney, J. W. & D'esposito, M. Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nature neuroscience* **12,** 515–522 (2009).

123. Azuar, C *et al.* Testing the model of caudo-rostral organization of cognitive control in the human with frontal lesions. *Neuroimage* **84,** 1053–1060 (2014).

124. Nee, D. E. & D'Esposito, M. The hierarchical organization of the lateral prefrontal cortex. *Elife* **5,** e12112 (2016).

125. Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic" scotomas". *Journal of Neuroscience* **13,** 1479–1479 (1993).

126. Mackey, W. & Curtis, C. E. Distinct contributions by frontal and parietal cortices support working memory. *bioRxiv,* 104430 (2017).

127. Sawaguchi, T. & Iba, M. Prefrontal cortical representation of visuospatial working memory in monkeys examined by local inactivation with muscimol. *Journal of Neurophysiology* **86,** 2041–2053 (2001).

128. Sawaguchi, T. & Goldman-Rakic, P. S. The role of D1-dopamine receptor in working memory: local injections of dopamine antagonists into the prefrontal cortex of rhesus monkeys performing an oculomotor delayed-response task. *Journal of Neurophysiology* **71,** 515–528 (1994).

129. Sawaguchi, T. & Goldman-Rakic, P. S. D1 dopamine receptors in prefrontal cortex: involvement in working memory. *Science* **251,** 947–951 (1991).

130. Mackey, W. E., Devinsky, O., Doyle, W. K., Meager, M. R. & Curtis, C. E. Human dorsolateral prefrontal cortex is not necessary for spatial working memory. *Journal of Neuroscience* **36,** 2847–2856 (2016).

131. Wallis, J. D. Polar exploration. *Nature neuroscience* **13,** 7–8 (2010).

132. Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P. & Rothwell, J. C. Theta burst stimulation of the human motor cortex. *Neuron* **45,** 201–206 (2005).

133. Rao, S. G., Williams, G. V. & Goldman-Rakic, P. S. Destruction and creation of spatial tuning by disinhibition: GABAA blockade of prefrontal cortical neurons engaged by working memory. *Journal of Neuroscience* **20,** 485–494 (2000).

134. Mackey, W. E., Devinsky, O., Doyle, W. K., Golfinos, J. G. & Curtis, C. E. Human parietal cortex lesions impact the precision of spatial working memory. *Journal of neurophysiology* **116,** 1049–1054 (2016).

135. Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A. & Sweeney, J. A. Maturation of cognitive processes from late childhood to adulthood. *Child development* **75,** 1357–1372 (2004).

136. Krappmann, P. & Everling, S. Spatial accuracy of primary and secondary memory-guided saccades in schizophrenic patients. *Schizophrenia research* **30,** 183–185 (1998).

137. Srimal, R. & Curtis, C. E. Secondary adaptation of memory-guided saccades. *Experimental brain research* **206,** 35–46 (2010).

138. Rahnev, D., Koizumi, A., McCurdy, L. Y., DEsposito, M. & Lau, H. Confidence leak in perceptual decision making. *Psychological science* **26,** 1664–1680 (2015).

139. Carter, E. & Wang, X.-J. Cannabinoid-mediated disinhibition and working memory: dynamical interplay of multiple feedback mechanisms in a continuous attractor model of prefrontal cortex. *Cerebral Cortex* **17,** i16–i26 (2007).

140. Murray, J. D. *et al.* A hierarchy of intrinsic timescales across primate cortex. *Nature neuroscience* **17,** 1661–1663 (2014).

141. Genovesio, A., Wise, S. P. & Passingham, R. E. Prefrontal–parietal function: from foraging to foresight. *Trends in cognitive sciences* **18,** 72–81 (2014).

142. Chafee, M. V. & Goldman-Rakic, P. S. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memorytask. *Journal of neurophysiology* **79,** 2919–2940 (1998).

143. Mackey, W. E., Winawer, J. & Curtis, C. E. Visual field map clusters in human frontoparietal cortex. *eLife* **6,** e22974 (2017).

144. Blanke, O *et al.* Visual activity in the human frontal eye field. *Neuroreport* **10,** 925–930 (1999).

145. Van Veen, V., Krug, M. K. & Carter, C. S. The neural and computational basis of controlled speed-accuracy tradeoff during task performance. *Journal of Cognitive Neuroscience* **20,** 1952–1965 (2008).

146. Fleming, S. M., Huijgen, J. & Dolan, R. J. Prefrontal contributions to metacognition in perceptual decision making. *Journal of Neuroscience* **32,** 6117–6125 (2012).