Title
Neural and computational underpinnings of serial dependence

Author
Sheehan, Timothy C.

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Neural and computational underpinnings of serial dependence

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

in

Neurosciences with Specialization in Computational Neurosciences

by

Timothy C. Sheehan

Committee in charge:

Professor John Serences, Chair
Professor Ed Callaway
Professor Eric Halgren
Professor Anastasia Kiyonaga

2023

The Dissertation of Timothy C. Sheehan is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DoG	Derivative of Gaussian
DoVM	Derivative of von Mises
FEF	Frontal Eye Fields
PFC	Prefrontal Cortex
V1	Primary Visual Cortex
fMRI	functional Magnetic Resonance Imaging
2AFC	2-Alternitive Forced Choice Task

ACKNOWLEDGEMENTS

This work would not have been possible without the help and support of my former and current lab mates: Maggie, Vy, Rosanna, Chaipat, Nuttida, Kirsten, Sunyong, Angus, Janna, Holly, Leah, and Stella. A special thanks to Chaipat for getting me to look for serial dependence in the first place, Sunyong for coming with me to Montreal, and Kirsten for helping set the lab up so we could keep things going during the pandemic. I am also grateful for the help of undergraduate research assistance including: Shuangquan Feng, Anika Jallorina, Muru Zhang, Sofia Fransico, Daniella Carey, Ben Carfano, and Dianthe Richmond who all help immensely with collecting data.

I will be forever thankful to my advisor John who has provided nothing but kindness and support for these past 5 years. John simultaneously supports my work by waking up at 4:39 AM during a family vacation in Berlin to edit my abstract 4 hours before the submission deadline and texting me at the end of a backpacking trip to make sure I made it out alright.

I am very thankful to my committee members for proving feedback at many stages of my PhD. Ed Callaway who facilitated a collaboration with Peichao Li which helped ground chapter 1. Eric Halgren who mentored me for INC fellowship and provided valuable feedback on experimental design. Marcelo Mattar who provided critical feedback on the modeling work in chapter 1 and Anastasia Kiyonaga who helped me think through large parts of Chapter 3.

I would not have even started this process without lucking in to a series of supportive mentors over the years through summer research, my post-bac, and rotations at UCSD including Steven Rice, Takashi Buma, Kareem Zaghloul, Sara Inati, Tim Gentner, and Saket Navlaka.

Thanks to my finding sources including a grant from the Institute of Neural Computation.

Chapter 1, in full, is a reprint of the material as it appears in PLOS Biology, 2022. Sheehan, Timothy C.; Serences, John T. The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material under preparation. Sheehan, Timothy C.; Serences, John T. The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material under preparation. Sheehan, Timothy C.; Carfano, Ben; Richmond, Dianthe; Serences, John T. The dissertation/thesis author was the primary investigator and author of this paper.

VITA

2015 Bachelor of Science in Bioengineering, Union College (NY)

2023 Doctor of Philosophy in Neurosciences with Specialization in Computational Neurosciences, University of California San Diego

PUBLICATIONS

Sheehan, T. C., & Serences, J. T. (2023). Distinguishing response from stimulus driven history biases. bioRxiv.

Sheehan, T. C., & Serences, J. T. (2022). Attractive serial dependence overcomes repulsive neuronal adaptation. PLoS biology, 20(9), e3001711.

Steinhardt CR, Sacré P, **Sheehan TC**, Wittig JH, Inati SK, Sarma S, Zaghloul KA. (2020) Characterizing and predicting cortical evoked responses to direct electrical stimulation of the human brain. *Brain Stimulation*, Oct;13(5):1218-1225

El-Kalliny, M. M., Wittig, J. H., **Sheehan, T. C.,** Sreekumar, V., Inati, S. K., & Zaghloul, K. A. (2019). Changing temporal context in human temporal lobe promotes memory of distinct episodes. *Nature communications*, 10(1), 203.

Dasgupta, S., **Sheehan, T. C.,** Stevens, C. F., & Navlakha, S. (2018). A neural data structure for novelty detection. *Proceedings of the National Academy of Sciences*, 115(51), 13093-13098.

Sheehan, T. C., Sreekumar, V, Inati, SK, Zaghoul, K. A. (2018) Signal complexity of human intracranial EEG tracks successful associative memory formation across individuals. *Journal of Neuroscience*. Feb 14;38(7):1744 –1755.

Sreekumar V., Wittig, J. H. Jr., **Sheehan, T. C.**, and Zaghloul K. A. (2017) Principled Approaches to Direct Brain Stimulation for Cognitive Enhancement (Review). *Frontiers in Neuroscience*. Nov 30; 11,650.

Buma T., Wilkinson B. C., **Sheehan T. C.**, (2015) Near-infrared spectroscopic photoacoustic microscopy using a multi-color fiber laser source, *Biomedical Optics Express*. 6(8), 2819–2829.

ABSTRACT OF THE DISSERTATION

Neural and computational underpinnings of serial dependence

by

Timothy C. Sheehan

Neurosciences with Specialization in Computational Neurosciences

University of California San Diego, 2023

Professor John Serences, Chair

Human perception and behavior are shaped by past experiences. Neural representations are constrained to utilize statistical regularities to encode the world efficiently while decision making should utilize heuristics to optimize the use of uncertain information. Recently, there has been heightened interest in serial dependence – a feature-specific attraction towards previously seen stimuli – to better understand these distinct objectives. Serial dependence differs from the more familiar perceptual adaptation effects as it is attractive, can be induced by weak stimuli, and can persist for 10s of seconds. Accounts

explaining serial dependence have varied greatly, with some positing a low-level perceptual phenomenon and others positing a post-perceptual origin operating during decision-making. That said, most existing studies did not separate the influence of previous stimuli, decisions, and motor actions so directly comparing possible mechanisms is challenging. We first examined the neural underpinning of serial dependence by having participants complete a delayed orientation discrimination task while measuring brain activity with functional magnetic resonance imaging (fMRI, Chapter 1). While behavioral responses indicated an attraction towards the previous stimulus, orientation-specific activation patterns in visual cortex exhibited a repulsive bias. We reconciled these apparently divergent findings with an ideal-observer model in which readout from perceptual areas during decision-making accounts for the attractive biases. We next developed a technique to distinguish stimulus from response generated biases using a simulated observer (Chapter 2). Utilizing this approach, we consistently found that reports were attracted towards previously reported – as opposed to previously presented – stimuli in an orientation report task. Finally, we sought to experimentally disentangle the role of sensory, decisional, and motor contributions to serial dependence (Chapter 3). Through a series of experiments, we found that attraction operates on a perceptual level, unrelated to attention or decisions, as well as on a decisional level, unrelated to motor or sensory experiences. We develop a model in which serial dependence is not the result of processing at a single stage. Instead, all levels of processing are influenced by a canonical prior for stability to optimize the efficiency of neural circuits that contribute to different cognitive operations.

GENERAL INTRODUCTION

Serial Dependence General

Serial dependence is the attraction of perceptual reports towards items seen in the recent past. This attraction is feature tuned (e.g. stronger for more similar items) and has been found for a wide range of features including orientation, numerosity, and faces (Cicchini et al., 2014; Fischer and Whitney, 2014; Liberman et al., 2014). Serial dependence is generally found to be stronger for noisier stimuli including internally generated noise during memory maintenance (Papadimitriou et al., 2015; Bliss et al., 2017; Manassi et al., 2018) and can persist across long inter-trial intervals and across several intervening trials (Fischer and Whitney, 2014). This is all in stark contrast to repulsive effects which tend to require long, powerful inducers (typically 5-20s) and can occur from stimuli that are not even consciously perceived, pointing to a low level sensory origin (He and MacLeod, 2001).

Compared with other history dependent perceptual biases, feature tuned serial dependence was identified and described very late with the first accounts emerging in 2014 (Cicchini et al., 2014; Fischer and Whitney, 2014). This is in stark contrast to other perceptual biases such as the oblique effect (Jastrow, 1892), the waterfall illusion (Addams, 1834), or the McCollough effect (McCollough, 1965) which were all characterized in one-shot experiments before computational psychology. The failure to identify serial biases for so long is not because they are particularly weak effects, in fact they are often much stronger than repulsive effects and can be on the order of 3 times the just noticeable difference between stimuli (Fischer and Whitney, 2014). Instead, the failure to identify such effects for so long likely derives from the fact that only emerge for stimuli with high levels of uncertainty, requiring many trials and serial task designs to reliably extract bias estimates. Here, I will provide a brief overview of the current

understandings and debates in the serial dependence literature. I will describe existing findings relating to the behavioral and neural origins of serial dependence before examining existing theoretical models for the cause and potential functional advantage of these biases.

Behavior

Serial dependence is typically studied by having participants complete a series of perceptual judgements in a row. The bias is then quantified by sorting trials based on the relative orientation of the previous stimulus ($\Delta\theta$) and examining response errors. This is most commonly achieved by utilizing a continuous report paradigm, but 2 alternative forced choice (2AFC) paradigms have also been utilized (Fischer and Whitney, 2014; Cicchini et al., 2017; Cicchini and Burr, 2018; Fritsche and de Lange, 2019). The inducing stimulus can be a stimulus that has been previously judged or passively viewed; in the same or different location as the target stimulus; and within a single trial or a stimulus from a preceding trial (Fornaciai and Park, 2018; Fritsche and de Lange, 2019). Serial dependence is an extraordinarily general phenomenon, having been observed across a wide range of visual features including (and to just name a few): emotion, variance, color, and spatial location (Bliss et al., 2017; Barbosa and Compte, 2018; Liberman et al., 2018; Suárez-Pinilla et al., 2018); across modality (Neto and Bartels, 2021; Zhang and Luo, 2023); and even species (Papadimitriou et al., 2015, 2016; Akrami et al., 2018). Thus, serial dependence may be viewed as a universal phenomenon shaping perception rather than a curious illusion emerging under rare circumstances.

Attention

A critical early finding related to the role of attention. In a sequential report task with eight stimuli presented on a ring, one was pre-cued on each trial (Fischer and Whitney, 2014). Participants were significantly attracted towards the previously attended stimulus regardless of

its relative location to the current stimulus but showed a weak (non-significant) repulsion from unattended stimuli in the same location. Similar results have been found regardless of spatial position with only attended stimuli inducing attractive biases, and those biases being stronger when stimuli share additional trails such as color or presentation order (Fischer et al., 2020). Interestingly, when an orthogonal feature is attended on the previous item (e.g. size vs. orientation), the resulting bias is reduced for that trial but stronger for the 2-back item (Fritsche and de Lange, 2019). Together, this suggests that serial dependence operates between features in the attentional field rather than simply on all stimuli that are encoded. That said, other groups have found attraction to task irrelevant items, but only over very short time periods (Fornaciai and Park, 2018).

Timescale

The timescale of serial dependence is highly variable with some studies producing an attractive bias extending back several trials to 50+ trials (Fischer and Whitney, 2014; Collins, 2020), while others only observe a bias for a single judgement (Fritsche et al., 2020). In the latter case, it was also observed that stimuli presented further back in time consistently induced a repulsive bias suggesting an interplay between a strong but short lived attractive bias being subsumed by a weaker but longer lasting adaptation (Fritsche et al., 2020). This phenomenon seems to be mostly determined by time-elapsed rather than intervening trials encountered as the same trends hold (attractive biases weakening and becoming repulsive) with longer inter-trial intervals (Papadimitriou et al., 2015; Bliss et al., 2017). On the level of single trials, studies utilizing spatial working memory consistently find stronger serial dependence with increasing delay periods (Papadimitriou et al., 2016; Bliss et al., 2017). In these studies, serial dependence does not emerge at all for very short delays, suggesting a critical role of working memory. An

alternative account, however, is that delay period is only critical in that it weakens perceptual representations. This claim is supported by a study that found serial attraction emerges on no-delay trials only when a backward mask was introduced in a spatial memory task (Manassi et al., 2018).

Uncertainty

One of the most reliable findings related to serial dependence is its correlation with uncertainty. Biases are consistently stronger on trials with greater uncertainty determined by either spatial frequency, orientation (oblique vs. cardinal), internal noise (measured through fMRI decoding), or delay period duration (Bliss et al., 2017; Cicchini and Burr, 2018; van Bergen and Jehee, 2019). Owing to publication selectivity, the lack of studies demonstrating attractive biases for high contrast stimuli without long delays is also telling. This relation to sensory uncertainty offers a strong hint that serial dependence is related to potentially optimal integration of sensory information to stabilize perception and minimize noise (Burr and Cicchini, 2014; Kiyonaga et al., 2017; Cicchini and Burr, 2018; van Bergen and Jehee, 2019).

Origins of serial dependence

Perhaps the most debated question relating to serial dependence is where it emerges. Many studies have shown that serial dependence can emerge towards a previously seen stimulus even if it is not reported and that it can bias the judgement of simultaneous comparisons (Fischer and Whitney, 2014; Liberman et al., 2014; Cicchini et al., 2017; Manassi et al., 2018). These findings suggest that serial dependence is a perceptual phenomenon likely arising from changes in sensory encoding. In contrast, more recent studies have asserted an alternative view, that serial dependence operates on decisions with attraction occurring between successive responses, not

the stimulus per se (Pascucci et al., 2019; Moon and Kwon, 2022). This account is supported by a failure to observe attractive biases (and a finding of weak repulsion) when previously attended stimuli weren't reported. It is unclear why the different sets of experiments came to such distinct conclusions, but it seems to be mediated by stimulus uncertainty with more "decisional" based findings generally having more easily perceivable stimuli. Adding to this confusion, even in papers claiming largely perceptual level effects, re-analyses suggest that the attractive bias could instead be related to the previous report (Sadil et al., 2021) although further work in this area is needed. One recent study utilizing both spatial and temporal biases may shed some light on this debate, suggesting that influences from past trials may be fed from a late decisional stage to an early perceptual one (Cicchini et al., 2021).

Neural underpinning

The origin of serial dependence has also been examined using neuroscientific methods, although progress in this area has been limited until very recently. Neuroimaging studies in humans offered conflicting findings, with one study finding an attractive behavioral bias was reflected in early visual cortex while another finding repulsive adaptation signals in visual cortex and attractive signals in dorsomedial right prefrontal cortex (dmPFC), bilateral intraparietal sulcus (IPS), and other non-sensory areas (Schwiedrzik et al., 2014; St. John-Saaltink et al., 2016). Unfortunately, neither of these studies utilized a paradigm reveling typical behavioral serial dependence with the first study utilizing only two orthogonal stimuli such that the resulting effect was indistinguishable from a motor or decisional response bias, while the second showed an attractive effect towards the previous response without feature tuning.

Later work in primates showed a strong and consistent behavioral effect of serial dependence on a spatial memory guided saccade task while recording units in FEF

(Papadimitriou et al., 2016). Paradoxically, they observed a repulsive bias from the previous stimulus and proposed a model whereby attention shifts receptive fields to optimally encode the current stimulus but changes slowly. The mismatch between the average tuning properties, and the responses on a given trial result in repulsion for a fixed decoder. Other studies have used a less direct approach. Van Bergen & Jehee, 2019 used the uncertainty (rather than the mean) of sensory representations to show that serial dependence was greater when sensory noise was higher on the current than the previous trial, consistent with Bayesian accounts (van Bergen and Jehee, 2019). Work combining human EEG and primate single unit recordings found representations of the previous stimulus were reactivated shortly before the next stimulus (Barbosa et al., 2020). This finding has been found in other EEG experiments and suggests that reactivation of activity silent memory may give rise to these biases (Bae and Luck, 2019; Luo and Collins, 2023). Lastly, a study of a highly related "contraction bias" in rat auditory processing revealed an active store of history information in posterior parietal cortex (PPC) that was causally related to history effects (they disappeared when PPC was optogenetically silenced) (Akrami et al., 2018).

Conceptual models

Perhaps more interesting is not the individual findings but the attempts at contextualizing why these effects might emerge in the first place. On the algorithmic level, several proposals have emerged relating to activity dependent plasticity or shifts in attentional fields. The most simple proposal is a stimulus specific gain in early sensory circuits, in other words anti-adaptation (Fischer and Whitney, 2014). Such a proposal would be supported by prior research into iconic memory and visual persistence which finds visual representations can outlive their physical presence, particularly for briefly presented stimuli (Coltheart, 1980; Benucci et al.,

2009) but those effects seem to be much too short lived and low-level to apply in most serial dependence paradigms. Additionally, while the sensory gain model is parsimonious, it seems to be largely inconsistent with neural studies of the phenomenon (Hajonides et al., 2023). An alternative proposal is tied to activity dependent plasticity. Representing a stimulus at any stage of sensory processing can elicit activity dependent plasticity changes that could strengthen connections and lower activation energies (e.g. through Ca^{++} loading in presynaptic terminals (Zucker and Regehr, 2002)) tuned to the previously encoded stimulus. A simulated network featuring activity dependent plasticity was able to recreate many of the timescale dependent effects of serial dependence (Bliss and D'Esposito, 2017). Related work extended this model to bump attractor circuits for working memory maintenance in PFC supported by activity patterns recorded from non-human primates (Barbosa et al., 2020). Lastly, an account on shifting receptive fields in later areas due to slowly fluctuating attention from the previous trial was found to account for both repulsive FEF activity and attractive behavior towards previous saccade targets (Papadimitriou et al., 2016).

On a computational level, a lot has been made of the critical role of sensory uncertainty on serial dependence effects. This finding fits nicely with more general Bayesian accounts of perception. When one is less sure about incoming sensory information (e.g. due to occlusions or noise) they should rely more on prior information (Kiyonaga et al., 2017). Well the optimal prior is hard to derive, analyses of natural videos has found that features such as orientation are highly stable across time (Dong and Atick, 1995). Thus, multiplying stimulus likelihood by a prior centered on the previous stimulus should lead to optimal performance and predict larger biases when under grater uncertainty (van Bergen and Jehee, 2019). In line with this, one study found response biases were larger when sensory uncertainty, decoded from V1 was larger on the

current trial than the previous (van Bergen and Jehee, 2019). The full Bayesian model leaves open the question of how a prior would be instantiated and stored, and in which scenarios such an undertaking would be metabolically worth it. Others have suggested a simpler but largely equivalent Kalman filter model that averages consecutive stimuli but only when they are similar. This averaging can reduce sensory noise even when stimuli are not correlated across time (Burr and Cicchini, 2014; Cicchini and Burr, 2018). Thus, there are many potential explanations for the serial dependence effect deriving from the biophysical properties of circuits tasked with maintaining memories to a more computational objective of minimizing uncertainty in a noisy world.

Closing

In this dissertation I attempt to shed light on the neural and computational underpinnings of serial dependence. In Chapter 1, I examined biases in the visual cortex of observers who perform an orientation memory task in a scanner. Participants exhibited a robust attractive bias towards past stimuli and were also more precise when consecutive stimuli were similar. Next., I built a decoder to predict what orientation was presented on each trial and examined the errors of that decoder to assess how stimulus history shaped neural representations of the stimuli. Contrary to expectations, I observed a robust repulsive bias from past stimuli that extensive modeling revealed to be unrelated to hemodynamic artifacts. I built an observer model to account for these conflicting results. While the neural repulsion was well explained by a gain-based adaptation model, a post-perceptual readout scheme that itself was shaped by task history led to a behavioral output that was attractive. This model explained changes in response variance it was not trained on, providing a strong external validation. More generally, this work suggests that it may be optimal to encode stimulus changes (maximizing the dimensionality of bandwidth limited

sensory areas) while later areas smooth encoded information across time. This chapter is a reprint of published work and our central findings have since been replicated in a study utilizing MEG based decoding (Sheehan and Serences, 2022; Hajonides et al., 2023).

In chapter 2, I explored the origins of serial dependence and whether they are better explained by the previous stimulus or the previous response. As noted earlier, it is an area of strong debate in the serial dependence literature whether biases arise at a perceptual or decisional level. Here, I focus first on methodological concerns relating to context independent biases and how they can introduce artifacts into certain analyses. I demonstrate that by correcting these artifactual biases, we can reliably capture infer whether the source of the bias is the pervious stimulus or the previous response. Importantly, this study rigorously tests a method for distinguishing correlated sources of serial dependence and suggests that certain noise conditions may obscure serial dependence effects in many cases due to the competing impact of repulsive adaptation and attractive serial dependence. We applied these techniques to an orientation working memory experiment and found that an attractive bias towards the previous stimulus was completely mediated by an attraction towards the previous response. This chapter has been released as a pre-print (Sheehan and Serences, 2023).

Finally, in chapter 3 I explore the role of stimulus and response experimentally. Across a series of 5 experiments, I disentangle the role of visual experience, attention, and motor action to triangulate the origins of serial dependence. I introduced a novel paradigm where spatial working memory is tested for both physical and imagined (compass coordinates) stimuli and manipulate how responses are made by eliminating motor and visual components associated with stimulus representation. We first find that serial dependence is largely indifferent to stimulus encoding format and tracks perceptual reports, not the stimulus. This data is consistent with a decisional

level attraction. However, we also find that attraction is significantly stronger for physical over imagined stimuli suggesting feedforward sensory activity also contributes to the observed serial dependence. Lastly, we find biases for imagined stimuli persist for substantially longer relative to low-level stimuli, pointing to distinct circuits driving the attraction. Together we find data that is neither wholly consistent with a perceptual or a decisional level origin of the bias. To account for this disparity, we propose a "many stages" model of serial dependence whereby it can, and typically does, emerge from biases at several levels of sensory processing. This chapter is under preparation for submission.

Chapter 1 Attractive serial dependence overcomes repulsive neuronal adaptation

Abstract

Sensory responses and behavior are strongly shaped by stimulus history. For instance, perceptual reports are sometimes biased towards previously viewed stimuli (*serial dependence*). While behavioral studies have pointed to both perceptual and post-perceptual origins of this phenomenon, neural data that could elucidate where these biases emerge is limited. We recorded fMRI responses while human participants (male and female) performed a delayed orientation discrimination task. While behavioral reports were *attracted* to the previous stimulus, response patterns in visual cortex were *repelled*. We reconciled these opposing neural and behavioral biases using a model where both sensory encoding and readout are shaped by stimulus history. First, neural adaptation reduces redundancy at encoding and leads to the repulsive biases that we observed in visual cortex. Second, our modeling work suggest that serial dependence is induced by readout mechanisms that account for adaptation in visual cortex. According to this account, the visual system can simultaneously improve efficiency via adaptation while still optimizing behavior based on the temporal structure of natural stimuli.

Introduction

Natural stimuli are known to have strong statistical dependencies across both space and time, such as a prevalence of vertical and horizontal (cardinal) orientations and a higher probability of small orientation changes in given spatial region over short time intervals (Dong and Atick, 1995; Felsen et al., 2005; Girshick et al., 2011; van Bergen and Jehee, 2019a). These regularities can be leveraged to improve the efficiency and accuracy of visual information processing. For example, regularities can yield attenuated neural responses to frequently occurring stimuli in early visual cortex (*adaptation*), reducing metabolic cost and redundancy in neural codes (Dragoi et al., 2001, 2002; Benucci et al., 2013; Patterson et al., 2014; Fritsche et al., 2022). At readout, regularities might support the formation of Bayesian priors that can be used to bias decision-making in favor of higher probability stimuli (Stocker and Simoncelli, 2006; Cicchini et al., 2014; Wei and Stocker, 2015). While the effects of stimulus history on sensory coding and behavior have been studied extensively, it is unclear how changes in sensory coding shape behavior.

Adaptation increases coding efficiency by modulating sensory tuning properties as a function of the recent past. For instance, reducing the gain of neurons tuned to a recently seen adapting stimulus reduces the temporal autocorrelation of activity when similar stimuli are presented sequentially, improving the overall efficiency of sensory codes (Clifford et al., 2000; Dragoi et al., 2000; Durant et al., 2007; Barlow, 2012; Benucci et al., 2013). Importantly, adapted representations early in the processing stream (e.g. in LGN) are inherited by later visual areas meaning the changes in coding properties could in turn shape decision making (Gardner et al., 2005; Dhruv and Carandini, 2014; Patterson et al., 2014). Although adaptation increases coding efficiency, it comes at a cost to perceptual fidelity as adaptation can lead to repulsion

away from the adapting stimulus for features such as orientation and motion direction (He and MacLeod, 2001; Moradi et al., 2005; Dekel and Sagi, 2015). For example, after continuously viewing and adapting to motion in one direction, stationary objects will appear to be moving in the opposite direction (i.e., current perceptual representations are *repelled* away from recent percepts). However, this potentially deleterious aftereffect is accompanied by better discriminability around the adapting stimulus, which may be more important than absolute fidelity from a fitness perspective (Phinney et al., 1997; Clifford et al., 2001; Abbonizio et al., 2002; Durant et al., 2007).

In contrast to the repulsive perceptual biases typically associated with neural adaptation, perceptual reports are sometimes attracted to recently presented items – a phenomenon termed "serial dependence". Studies utilizing low contrast oriented stimuli suggest serial dependence can be perceptual in nature as it operates before a peripheral tilt illusion, impacts the perception of simultaneously presented items, biases perceptual reports even when no probe is presented, and does not require a working memory delay (Fischer and Whitney, 2014; Cicchini et al., 2017, 2021; Manassi et al., 2018; Murai and Whitney, 2021). This perceptual account could arise from activity changes in early visual cortex, consistent with a fMRI study which measured early sensory biases that match 'attractive' behavioral reports (St. John-Saaltink et al., 2016). This neural finding, however, is challenging to interpret as consecutive trials were always the same or orthogonal orientations which, by definition, cannot distinguish attractive from repulsive biases. Related studies decoding past stimuli from EEG activity do not measure how current stimulus representations are biased, precluding a connection to behavioral biases (Fornaciai and Park, 2018; Bae and Luck, 2019; Bae, 2021).

Counter to studies reporting a perceptual locus of serial dependence which utilized brief or low contrast stimuli, other behavioral studies utilizing high contrast spatial stimuli have found that serial dependence does not emerge immediately but instead emerges only, and increases with, a working memory maintenance period (Papadimitriou et al., 2015; Bliss et al., 2017; Stein et al., 2020). This observation suggests that serial dependence could be implemented by a later readout or memory maintenance circuit (Papadimitriou et al., 2015; Bliss and D'Esposito, 2017; Pascucci et al., 2019; Barbosa et al., 2020). There is evidence that such a readout mechanism is Bayesian, as the influence of the "prior" (the previous stimulus) is larger when sensory representations are less precise due to either external or internal noise (Cicchini et al., 2018; van Bergen and Jehee, 2019a). Thus, the existing behavioral evidence suggests that serial dependence can operate both on perceptual and working memory representations (Papadimitriou et al., 2015; Cicchini et al., 2017; Kiyonaga et al., 2017). It is open question how and where past trial information interacts with incoming sensory and memory representations.

To determine what role visual cortex plays in driving serial dependence, we applied multivariate fMRI decoding techniques to data collected while subjects performed a delayed orientation discrimination task (Figure 1-1A). We replicated classic serial dependence findings where behavioral reports were attracted to the orientation of the previous stimulus. However, this attractive behavioral bias was not accompanied by attractive biases in visual cortex, as predicted by early sensory models of serial dependence. Rather, we observed *repulsive biases* in early visual cortex that were consistent with adaptation. We then examined several possible read-out mechanisms and found that only decoding schemes that account for adaptation can reconcile the neural and behavioral biases found in our data. More generally, these results explain a mechanism where the visual system can reduce energy usage without sacrificing precision by

optimizing sensory coding and behavioral readout relative to the temporal structure of natural environments.

Results

Behavior

To probe the behavioral effects of serial dependence, we designed a delayed discrimination task where participants judged whether a bar was tilted clockwise (CW) or counterclockwise (CCW) relative to the orientation of a remembered grating (Figure 1-1A). We first report the results from a behavior-only study (n=47) followed by an analysis of neural activity for a cohort completing the same task in the fMRI scanner (n=6). Task difficulty was adjusted for each participant by changing the magnitude of the probe offset ($\delta\theta$) from the remembered grating and was titrated to achieve a mean accuracy of ~70% (accuracy 69.8±0.82%, $\delta\theta$: 4.61±0.27°; all reported values mean ±1SEM unless otherwise noted). Fixing subjects at this intermediate accuracy level helped to avoid floor/ceiling effects and improved our sensitivity to detect perceptual biases while keeping participants motivated.

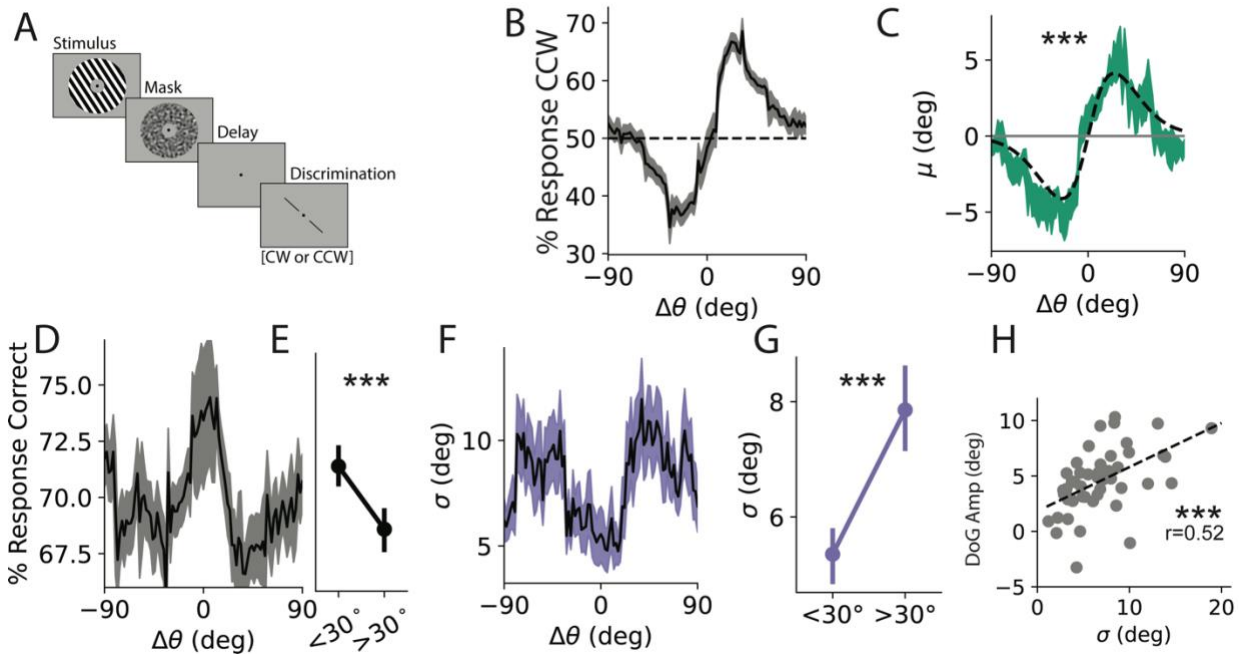


Figure 1-1 Behavior.

A: Task Schematic. An orientated stimulus is followed by a probe bar that is rotated <15° from the stimulus. Participants judged whether the bar was CW or CCW relative to the stimulus in a binary discrimination task. B: Response bias: % of responses that were CCW as a function of $\Delta\theta=\theta n\text{-}1\text{-}\theta n$ (± SEM across participants). C: Behavioral bias, green: average model-estimated bias as a function of $\Delta\theta$ (± SEM across participants); gray: average DoG fit to raw participant responses sorted by $\Delta\theta$ (± 1SEM across participants). D: Response accuracy as a function of $\Delta\theta$. E: Responses are significantly more accurate for $|\Delta\theta|<30°$. F: Behavioral σ as a function of $\Delta\theta$. G: Behavioral variance is significantly less for $|\Delta\theta|<30°$. Note that in computing variance we 'flip' the sign of errors following CCW inducing trials to avoid conflating bias with variance (see Methods) H: Bias is positively correlated with variance across participants. ***, p<.001.

To quantify the pattern of behavioral responses, we modelled the data as the product of a noisy encoding process described by a Gaussian distribution centered on the presented orientation with standard deviation σ and bias μ. Optimal values for σ and μ were found by maximizing the likelihood of responses for probes of varying rotational offsets from the remembered stimulus, thus converting pooled binary responses into variance and bias measured in degrees (see Response Bias, Figure 1-6). This allowed us to measure precision for individual participants and also allowed us to measure how responses were biased as a function of the

orientation difference between the remembered gratings on consecutive trials $\Delta\theta = \theta_{n-1} - \theta_n$, an assay of serial dependence.

Responses were robustly biased towards the previous stimulus (Fig 1C, green curve), which we quantified by fitting a Derivative-of-Gaussian (DoG) function to the raw response data for each participant (gray curve; amplitude: 4.53°±0.42°, $t(46) = 7.8$, $p = 5.9*10^{-10}$, one sample t-test; full width at half max (FWHM): 42.9°±1.8°, see Serial Dependence). The magnitude and shape of serial dependence is consistent with previous reports (Fischer and Whitney, 2014; Fritsche et al., 2017). This bias is not an artifact of our parameterization as the same pattern is observable in the raw proportion of CCW responses (Fig 1B). Note that as participants are reporting the orientation of the probe relative to the grating stimulus, a greater proportion of reports that the probe was CCW corresponds to a CW shift in the perception of the grating.

We next examined how response precision (σ) varied as a function of $\Delta\theta$ and found that responses were more precise around small trial-to-trial orientation changes (Fig 1F), again consistent with previous reports (Cicchini and Burr, 2018). We quantified this difference in precision by splitting trials into 'close' and 'far' bins (greater than or less than 30° separation) and confirmed that responses following 'close' stimuli were more precise ($t(46)=-3.72$, $p=0.0003$, paired 1-tailed t-test, Figure 1-1G , see Response Precision). Note that the choice of 30° was arbitrary, but all threshold values between 20° and 40° yielded significant ($p<.05$) results. As with bias, this variance result was not an artifact of our parameterization as raw accuracy showed a similar pattern such that responses were more accurate following close stimuli ($t(46)=3.66$, $p=.0003$, Figure 1-1D-E). We additionally confirmed that our finding of reduced bias around small changes in orientation is not driven by a higher proportion of 'cardinal' orientations (here defined as being ±22.5° of 0 or 90°) as the proportion of cardinal

orientations did not differ between close and far bins of $\Delta\theta$ (mean % cardinal close: 50.6±0.5%, far: 50.2±0.3%, t(46)=0.9, p=0.39, paired t-test).

Previous work has shown that serial dependence is greater when stimulus contrast is lower (Manassi et al., 2018) and when internal representations of orientation are weaker due to stimulus independent fluctuations in encoding fidelity (van Bergen and Jehee, 2019a). We tested a Bayesian interpretation of these findings by asking whether less precise individuals are more reliant on prior expectations and therefore more biased. Consistent with this account, we found a positive correlation between DoG amplitude and σ (Figure 1-1H, r(45)=0.52, p=.0001, 1-tailed Pearson's correlation). This relationship was not dependent on our response parameterization as we report found similar relationships between DoG amplitude and both accuracy (r=-0.41, Pearson correlation, p<.005) and average task difficulty $\delta\theta$ (r=.44, p<.005).

A subset of participants completed a version of the experiment with inhomogeneities in their stimulus sequences (such that consecutive orientations were more likely to be between ±22.5 and 67.5° from the previous stimulus). We repeated all of the above analyses excluding these participants and found all of our findings were qualitatively unchanged (Figure 1-7).

Stimulus history effects in visual cortex

To examine the influence of stimulus history on orientation-selective response patterns in early visual cortex, six participants completed between 748 and 884 trials (mean 838.7) of the task in the fMRI scanner over the course of four, two-hour sessions (average accuracy of 67.7% ±0.4% with an average probe offset, $\delta\theta$, of 3.65°). As with the behavior-only cohort, behavioral reports in these participants showed strong attractive serial dependence (Figure 1-2A, green) that was significantly greater than 0 when parameterized with a DoG function (amplitude=3.50°±0.27°, $t(5)=11.93$, $p=.00004$; FWHM=35.9°±2.34°, Figure 1-2A black dotted

line). This bias was not significantly modulated by inter-trial interval, delay period, or an interaction between the two factors (all p-values > 0.5, mixed linear model grouping by participant). Similar to the behavioral cohort, we found that variance was generally lower around small values of $\Delta\theta$. We quantified variance in the same manner as the behavioral cohort (flipping responses to match biases and down-sampling the larger group) and found that responses were more precise following close (<30°) relative to far stimuli (>30°, $t(5) = -9.96$, $p=0.00009$, 1-tailed paired t-test, Figure 1-2B). This pattern was significant ($p<0.05$) for thresholds between 20° and 40°. A subset of these participants completed some sessions where consecutive stimuli were not strictly independent as they were more likely to be between ±22.5 and 67.5° from the previous stimulus (see Methods, Behavioral Discrimination Task, 4 out of 6 subjects had between 357-408 trials that were non independent accounting for between 40-50% of their trials and 32% of all trials completed). However, we replicated all of our main analysis excluding these sessions and found that our conclusion remained unchanged with the exception that our finding of reduced variance trended in the same direction but no longer reached significance (Figure 1-8).

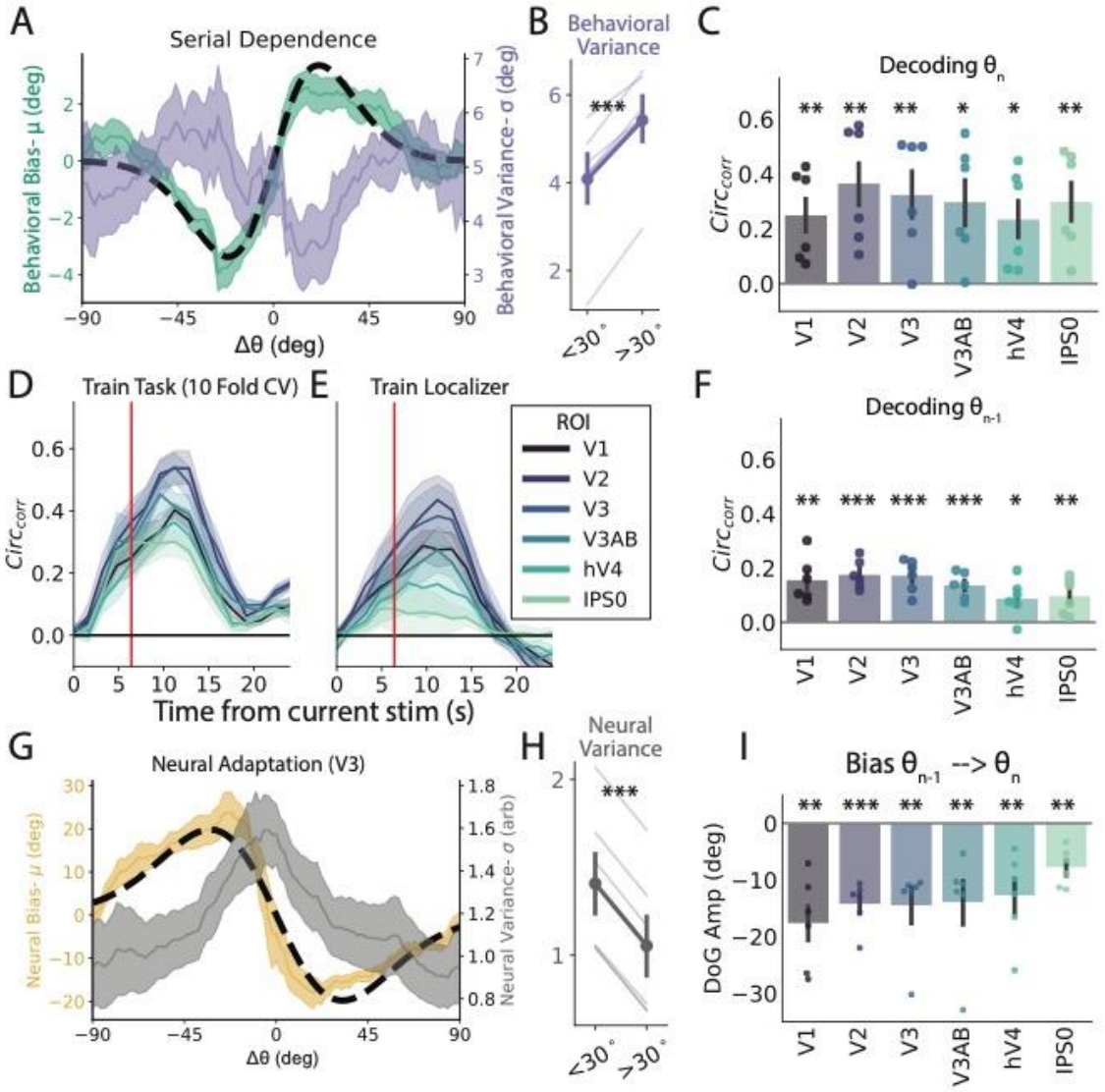


Figure 1-2 Behavioral and Neural Bias.

A: Left-axis, Behavioral serial dependence. Shaded green: average model-estimated bias as a function of Δθ (± SEM across participants); dotted black line: average DoG fit to raw participant responses sorted by Δθ. Right-axis, variance. Purple shaded line: model-estimated variance as a function of Δθ (± SEM across participants). B: Behavioral σ is significantly less for |Δθ|<30°. C: Decoded orientation was significantly greater than chance when indexed with circular correlation for all ROIs examined. Error bars indicate ±SEM across participants. Dots show data from individual participants. D: Decoding performance across time for a subset of ROIs. Vertical red line indicates time point used in most analysis. E: Decoding performance across time for a decoder trained on a separate sensory localization task. F: Performance of task decoder trained and tested on identity of previous stimulus across all ROIs. G: Left-axis, decoding bias. Shaded yellow line: decoded bias (μ_{circ} of decoding errors) sorted by Δθ (± SEM across participants); dotted black line: average DoG fit to raw decoding errors sorted by Δθ. Right-axis, decoded σ_{circ}. Shaded gray line: average decoding variance (σ_{circ}) as a function of Δθ (± SEM across participants). Note that σ_{circ} can range from [0, inf] and has no units. H: Decoded variance is significantly greater for |Δθ|<30° I: Decoded errors are significantly repulsive when parameterized with a DoG in all ROIs. *, p<.05; **, p<.01; ***, p<.001.

To characterize activity in early visual areas, independent retinotopic mapping runs were completed by each subject to identify regions of interest (ROIs) consisting of: V1, V2, V3, V3AB, hV4, and intraparietal sulcus area IPS0. In addition, a separate localizer task was used to sub-select the voxels that were most selective for the spatial position and orientation of the stimuli used in our task (see Voxel Selection).

To examine how visual representations are affected by stimulus history, we trained a decoder on the orientation of the sample stimulus on each trial based on BOLD activation patterns in each ROI. We used the vector mean of the output of an inverted encoding model (IEM) as a single trial measure of orientation using a leave-one-run-out cross-validation across sets of 68 consecutive trials (4 blocks of 17 trials) that had orientations pseudo randomly distributed across all 180° of orientation space (see Orientation Decoding for details). We first quantified single-trial decoding performance using circular correlation (r_{circ}) between the decoder-estimated orientations and the actual presented orientations and found that all ROIs had significant orientation information (Figure 1-2C). Our ability to decode extended for the duration of the trial, peaking around 12s after stimulus presentation (Figure 1-2D). This memory signal seems to be largely in a 'sensory code' as a decoder trained on a separate localizer task where participants viewed stimuli without holding them in memory achieved similar performance over a similar timescale (see fMRI Localizer Task, Figure 1-2E). Thus, visual ROIs showed robust orientation information that could be decoded across the duration of the trial. For all analyses not shown across time, we used the average of four TRs (spanning 4.8-8.0s) following stimulus presentation to minimize the influence of the probe stimulus (which came up ≥6s into the trial and thus should have a negligible influence on activity in the 4.8-8.0s window after accounting for hemodynamic delay, see Figure 1-5A).

We are interested in the how the identity of the previous stimulus influences representations of the current stimulus, akin to previous EEG studies that have demonstrated the ability to decode the previous stimulus during the current trial (Bae and Luck, 2019). We performed a similar analysis by training and testing our task decoder on the identity of the previous stimulus using the same time-points as the current trial decoder. This decoder was able to achieve above chance decoding in all ROIs examined indicating trial history information is present in the activity patterns (Figure 1-2F). As a control analysis, we attempted but were unable to decode the identity of the next stimulus using the same procedure (Figure 1-8). The performance of the memory decoder for the previous stimulus peaked around 6s after stimulus presentation but remained above chance throughout the delay period (Figure 1-9A). Notably, we were generally unable to decode the identity of the previous stimulus using our decoder trained on a localizer task suggesting representations of past trial stimuli are not in a 'sensory code' (Figure 1-9B).

The high SNR of the BOLD decoder additionally allowed us to examine residual errors on individual trials. When measuring the bias (circular mean, μ_{circ}, *see* Neural Bias) of these decoding errors as a function of stimulus history ($\Delta\theta$), we observed a strong *repulsive* bias reflecting neural adaptation (V3, Figure 1-2G yellow). This bias was significant when quantified with a DoG (amplitude=-14.5°±2.9°, $t(5)$=-3.56, p=.0029; FWHM=52.2°±2.94°, Fig 2G black dotted-line), and all ROIs had a significantly negative amplitude ($p<.01$, Figure 1-2I). Critically, this bias was present across all TRs for both the task and localizer decoders and was visible in the bias curve computed for each individual participant (Figure 1-9). In addition to the model-based analysis of responses in visual cortex, we also performed a model-free assessment of the dimensionality of activation patterns conditioned on the prior stimulus. Consistent with our main

analysis, responses following close stimuli have a higher dimensionality than responses following far stimuli. This suggests that changes due to neural adaptation should assist pattern separation regardless of stimulus identity (see Dimensionality Analysis, Figure 1-10).

We also examined how the precision of neural representations changed as a function of stimulus history. In sharp contrast to behavior, σ_{circ} exhibited a monotonic trend such that neural decoding was *least* precise when the previous stimulus was similar (Figure 1-2G, gray curve, see Neural Variance). We quantified this difference in sensory uncertainty in a similar manner to the behavioral data and found that variance in the sensory representations was significantly greater following a similar stimulus (<30°, $t(5)=72.4$, $p=4.8*10\text{^-}9$, paired 1-tailed t-test, V3, Fig 2H). This pattern was significant ($p<.05$) in all ROIs except IPS0 (Figure 1-11A). The results did not change qualitatively when we utilized vector length as a proxy for decoding precision derived directly from our channel estimates (Figure 1-11C-D), or when we used other thresholds between 20° and 40°. The repulsion of sensory representations and the corresponding reduction in decoding precision around the previous orientation is consistent with neural adaptation where recently active units are attenuated, thus leading to lower SNR responses in visual cortex.

Accounting for the Timecourse of the Hemodynamic Response Function

We considered whether the repulsive adaptation we observed in visual cortex could be explained by residual undershoot of the hemodynamic response function (HRF) from the previous stimulus. To address this concern, we directly modeled the evoked response in each voxel to the stimulus and probe using a deconvolution approach and used a parameterization of the resulting filter (double gamma function) to model the stimulus evoked response on each trial (see kernel based decoding). Notably the stimulus response has an undershoot that extends up to

25s following stimulus presentation (see Figure 1-3A for an example voxel and parameterization). Estimating responses using this filter on individual trials and using the resulting weights to train a decoder removes the linear contribution of previous stimulus/ probe presentations (Dale, 1999; Glover, 1999). Any bias in the resulting decoder should thus be due to changes in BOLD activity driven by neuronal activity rather than a hemodynamic artifact. We repeated all analyses after correcting for the shape of the HRF, and while the resulting decoder was less precise than one trained on the time course data (eg. V3 r_{circ}=0.19±.07 versus 0.32±0.08 with time course decoder), it was still significantly predictive across all visual ROIs (ps<.05) except IPS0. Despite the noisier decoding, we still observed a significant repulsive bias in all visual ROIs that matched the pattern found when decoding the raw BOLD timecourse (Figure 1-3B).

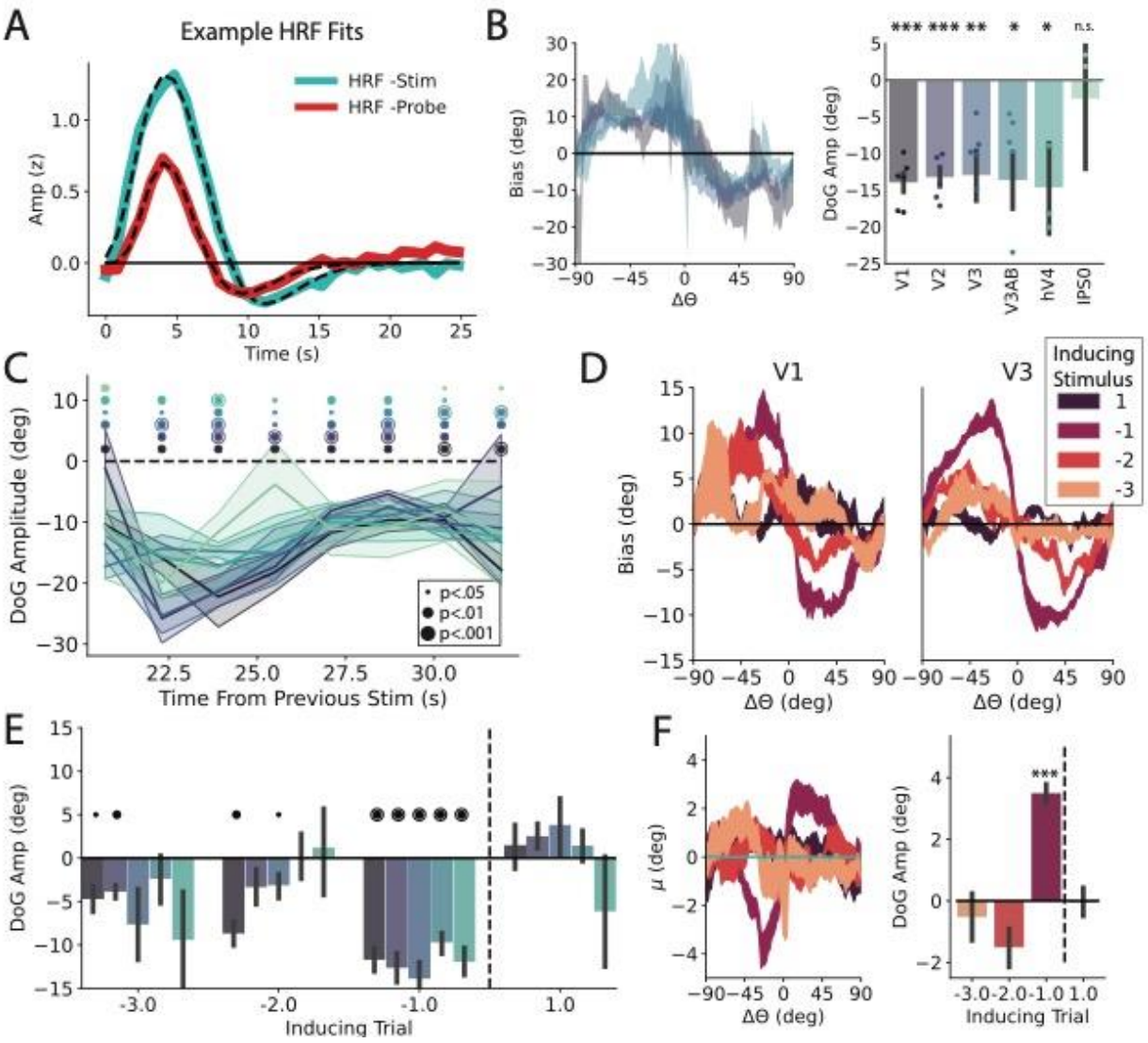


Figure 1-3 Influence of BOLD specific biases on repulsive bias.

A: average V1 HRF through deconvolution for stimulus and probe. Average best fit double gamma function overlaid in dotted lines. B: (left) Bias curves from decoder trained on response patterns from deconvolved double-gamma functions (± SEM across participants). Here excluding hV4 and IPS0 for clarity. (right) bias quantified with a DoG function across ROIs. C: Bias across time including only trials with an ISI of at least 17.5s. X-axis reflects minimum time from previous stimulus. Repulsion significant in all ROIs at 32s. D: Bias as a function of various relative orientations for V1 and V3 (± SEM across participants). E: Bias across early visual ROIs for N-1, -2, -3. Color scheme same as C. N+1 control analysis to ensure effects not driven by some unknown structure in stimulus sequence. F: Behavioral bias for various relative orientations. N-1 data same as data presented in Fig 2. *, p<.05, **, p<.01, ***, p<.001.

To further understand whether the time course of our task could lead to artifacts, we also simulated responses to our task using tuned voxels that were modeled after the task sequence and

estimated HRFs observed in our experiment (see supplementary modeling section, Figure 1-12). These simulations show that repulsive biases like the ones we observed with both our time course and deconvolution-based decoders are only possible when the underlying tuning of voxels is adapted by past stimuli/responses.

We additionally examined the time-course of the bias. Significant repulsive biases were observable through the duration of the trial, in all early visual ROIs (Figure 1-9). As the undershoot portion of the HRF extended to ~25s, we examined the bias relative to the time of the presentation of the previous stimulus. We included only trials with an inter-stimulus interval (ISI) greater than the median of 17.5s and plotted bias as a function of the minimum time from the previous stimulus (Figure 1-3C). Notably bias was still significantly repulsive for 30s following the previous stimulus presentation in all early visual ROIs, further shrinking the possibility that our biases are driven by the slow timecourse of the HRF (Figure 1-3C, last time point). Finally, we examined how far back previous stimuli shape early visual representations. We examined the influence of not just the N-1 stimulus, but N-2 and N-3 stimuli as well, corresponding to median ISIs of 35.1 and 52.5s respectively (Figure 1-3D-E). As any influence of these more distant stimuli should be diminished relative to N-1, we maximized our sensitivity by taking the average decoded representation from 4-12s. While the control N+1 stimulus showed no impact on decoded orientation as expected, we continued to see biases that are significantly repulsive through the N-3 stimulus in V1 and V2 (Figure 1-3E). These neural biases were surprisingly persistent and are in line with recent studies which have found adaptation signatures extending 22s in mouse visual cortex spiking activity (Fritsche et al., 2022). It is not clear why our effects persist even longer, but it is likely driven in part by the long ISIs, resulting in fewer intervening stimuli compared to the paradigm utilized in (Fritsche et al., 2022). We

separately extended our analysis of behavioral biases and found no significant effect of trials except for N-1, although biases were trending towards being repulsive for N-2 and N-3 reflecting the pattern reported in (Fritsche et al., 2020) (Figure 1-3F). Together these analyses suggest that our observed biases are driven by adaptation in the underlying neural population and provide additional evidence that behavior is not directly linked to early visual representations.

Encoder-Decoder Model

We observed an attractive bias and low variability around the current stimulus feature in behavior, and a repulsive bias and high variability around the current feature in the fMRI decoding data. Thus, the patterns of bias and variability observed in the behavioral data are opposite to the patterns of bias and variability observed in visual cortex. To better understand these opposing effects, we reasoned that representations in early visual cortex do not directly drive behavior but instead are read out by later cortical regions that determine the correct response given the task (Crick and Koch, 1995; Grunewald et al., 2002; Gold and Shadlen, 2007; Siegle et al., 2021). In this construction, the decoded orientations from visual cortex represent only the beginning of a complex information processing stream that, in our task, culminates with the participant making a speeded button press response. Thus, we devised a two-stage encoder-decoder model to describe observations in both early visual cortex and in behavior (see modeling).

The encoding stage consists of cells with uniformly spaced von Mises tuning curves whose amplitude is adapted by the identity of the previous stimulus (θ_{n-1}, Figure 1-4A). The decoding stage reads out this activity using one of three strategies (Fig 4B). The *unaware* decoder assumes no adaptation has taken place and results in stimulus likelihoods $p(m|\theta)$ that are repelled from the previous stimulus (Figure 1-4C, yellow, where m is the population activity

at the encoding stage). This adaptation-naïve decoder is a previously hypothesized mechanism for behavioral adaptation (Seriès et al., 2009) and likely captures the process that gives rise to the repulsive bias we observe in visual cortex using a fMRI decoder that is agnostic to stimulus history (Figure 1-2G). Alternatively, the *aware* decoder (Figure 1-4C, green) has perfect knowledge of the current state of adaptation and can thus account for and 'un-do' biases introduced during encoding. Finally, the *over-aware* decoder knows the identity of the previous stimulus but over-estimates the amount of gain modulation that takes place, resulting in a net attraction to the previous stimulus (Figure 1-4C, red). We additionally built off of previous work showing stimuli are generally stable across time by implementing a prior of temporal contiguity (van Bergen and Jehee, 2019a). In our implementation, a Bayesian prior centered on the previous stimulus (Figure 1-4C, black) is multiplied by the decoded likelihood to get a Bayesian posterior (Figure 1-4C, bottom). We applied this prior of temporal contiguity to both the *aware* decoder as well as the *unaware* decoder to test the importance of awareness at decoding. We did not apply a prior to the *over-aware* model to balance the number of free parameters between the various decoders and to see if the *over-aware* model could achieve attractive serial dependence without a Bayesian prior (Figure 1-4, S1 Table).

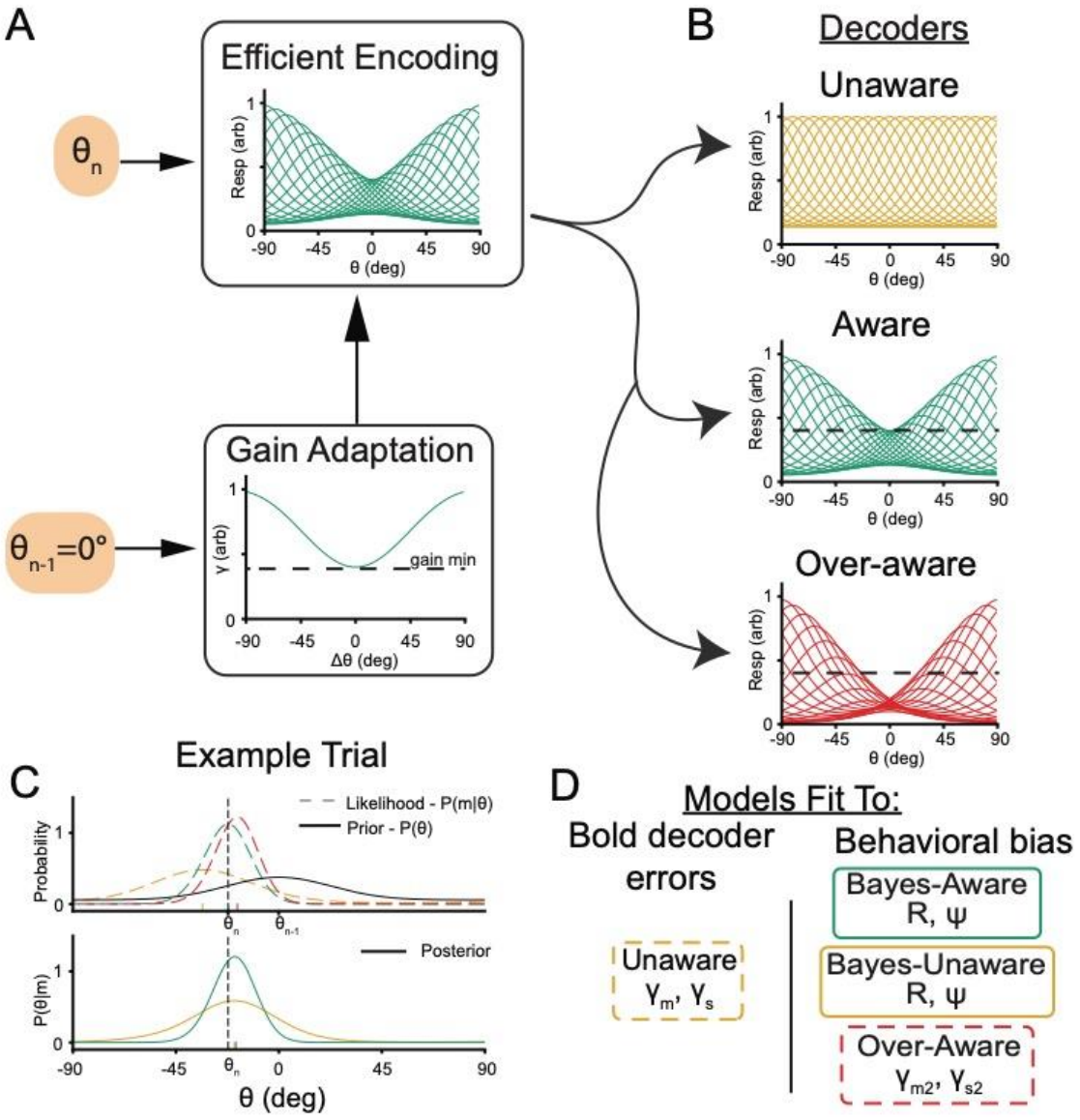


Figure 1-4 Encoder-Decoder model schematic.

A: Encoding. Units with von Mises tuning curves encodes incoming stimuli. The gain of individual units undergoes adaptation such that their activity is reduced as a function of their distance from the previous stimulus. B: Decoding. This activity is then read out using a scheme that assumes one of three adaptation profiles. The unaware decoder assumes no adaptation has taken place, the aware decoder assumes the true amount of adaptation while the over-aware decoder over-estimates the amount of adaptation (note center tuning curves dip lower than the minimum gain line from encoding). C: Example stimulus decoding. Top: The resulting likelihood function for the unaware readout (dotted yellow line) has its representation for the current trial (θn=-30°) biased away from the previous stimulus (θn-1=0°). The aware readout (dotted green line) is not biased, while the over-aware readout is biased towards the previous stimulus. These likelihood functions can be multiplied by a prior of stimulus contiguity (solid black line) to get a Bayesian posterior (bottom) where Bayes-unaware and Bayes-aware representations are shifted towards the previous stimulus. Tick marks indicate maximum likelihood or decoded orientation. D: Summary of models and free parameters being fit to both BOLD decoder errors and behavioral bias

For each participant, we fit the encoder-decoder model in two steps (Figure 1-4D). All model fitting was performed using the same cross-validation groups as our BOLD decoder and each stage had two free parameters that were fit using grid-search and gradient descent techniques. We first report results from the encoding stage of the model. The gain applied at encoding was adjusted to minimize the residual sum of squared errors (RSS) between the output of the *unaware* decoder and the residual errors of our BOLD decoder. The *unaware* readout of the adapted encoding process (Figure 1-5A, yellow) provided a good fit to the average decoding errors obtained with the BOLD decoder (Figure 1-5A, black outline, ρ=0.99) and across individual participants (Figure 1-13A, ranges: ρ= [0.84,0.98]). The *unaware* readout provided a better fit to the outputs of our neural decoder than the null alternative of the presented orientation ($t(5)$=3.41, p=.01) because it captured a significant proportion of the variance in decoding errors as a function of $\Delta\theta$ ($t(5)$=7.5, p=.0007). This analysis demonstrates that our adaptation model does a reasonable job of recovering our empirical decoding data (both of which use a decoder unaware of sensory history).

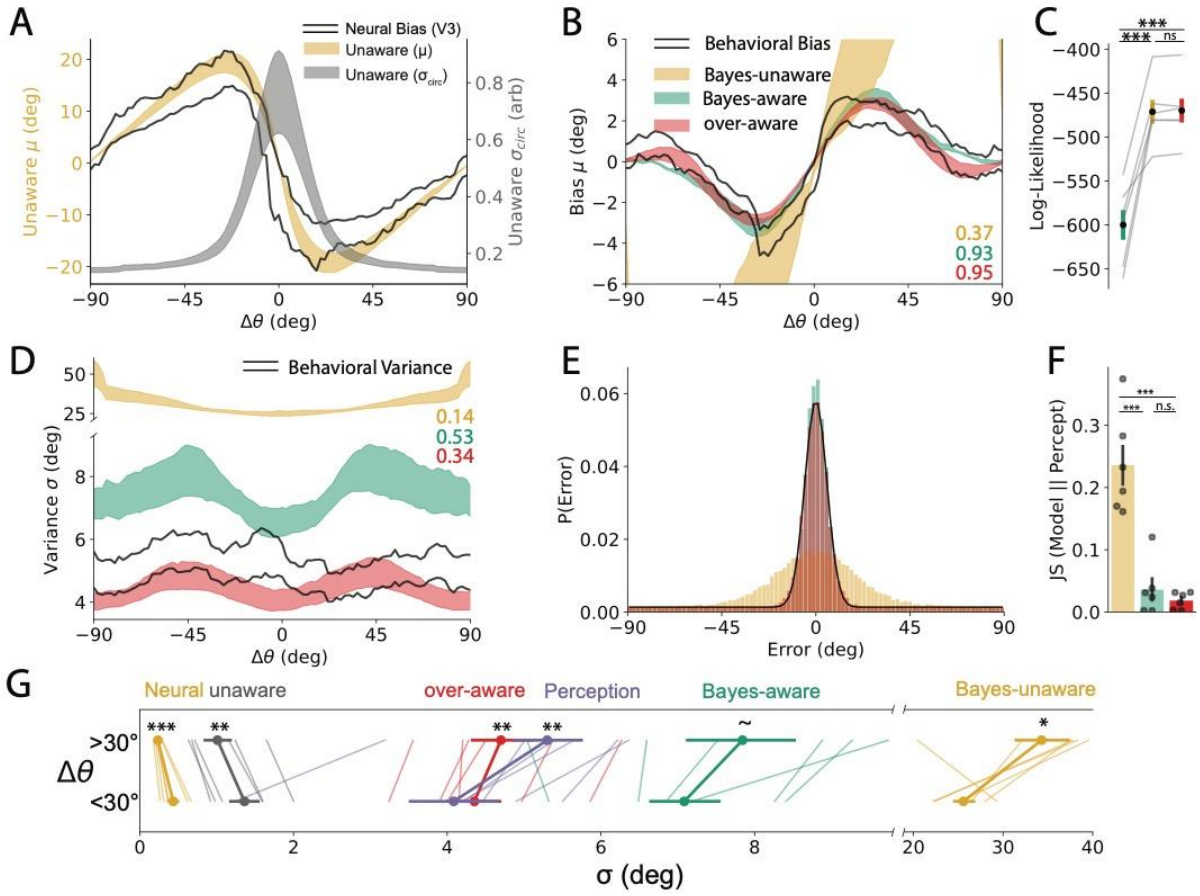


Figure 1-5 Model performance bias.

A-C Neural/Behavioral Bias, D-G Neural/Behavioral Variance. A: Unaware decoder (yellow) provides a good fit to neural bias (black outline). Decoded variance decreases monotonically with distance from previous stimulus. (± SEM across participants). B: Perceptual bias (black outline) was well fit by the Bayes-aware and over-aware models but not the Bayes-Unaware model (± SEM across participants). C: Participant responses were significantly more likely under aware models. D: Behavioral variance had a similar shape and magnitude to Bayes-aware and over-aware model fits. Bayes-unaware model output was much less precise and had a different form. E: Distribution of empirically predicted response errors (black line) and simulated model fits for an example participant. F: The unaware model's error distribution had significantly higher Jenson-Shannon Divergence from BOLD decoder than either aware model. G: Visualization of all uncertainties split as a function of close and far stimuli. Note that the Bayes-unaware model had an average uncertainty that was on average 6x that of perception. *, p<.05; **, p<.01; ***, p<.001.

We next considered three readout schemes of this adapted population to maximize the likelihood of our behavioral responses (Figure 1-5B). The *Bayes-aware* decoder is consistent with previous Bayesian accounts of serial dependence (van Bergen and Jehee, 2019a), but additionally asserts that Bayesian inference occurs after encoding and that readout must account

for adaptation. Alternatively, the *Bayes-unaware* decoder tests whether this awareness is necessary to achieve attractive serial dependence. Both aware models achieved biases that were significantly more likely than the unaware model ($t(5)=6.53$, p=.001, Bayes-aware; t(5)=6.6, p=.001, over-aware, t-test on log-likelihood, Figure 1-5C) but were indistinguishable from each other (p=.36). Thus, both aware models were able to explain the response biases while the unaware model did a relatively poor job, suggesting some awareness of the adapted state is necessary.

Finally, we examined the variance of our decoders to see if this mapped onto our empirically observed variance. As model coefficients were fit independent of observed variance, correspondence between model performance and BOLD/behavioral data would provide convergent support for the best model. While the models were trained using noiseless activity at encoding, we simulated responses using Poisson rates to induce response variability. We simulated 1000 trials from each cross-validated fit and pooled the model outputs. We first confirmed that the variance of the *unaware* decoder was highest following small changes of $\Delta\theta$ (Figure 1-5A, gray; Figure 1-5G $t(5)=3.93$, p=.005, paired 1-tailed t-test <30° vs >30°) matching the output of our neural decoder (Figure 1-2G) and providing additional support for gain adaptation causing the observed repulsion in the fMRI data. Next, we compared the different decoders and found that, matching real behavioral responses, all three decoders were more precise following small values of $\Delta\theta$ (Figure 1-5G, Bayes-unaware, $t(5)=2.25$, p=.037; Bayes-aware $t(5)=1.90$, p=.058; and over-aware $t(5)=5.43$, p=.001). While the pattern of the Bayes-unaware variance matched behavior, it's overall variance was much higher than our behavioral data such that it diverged from the behavioral data significantly more than either of the aware models (Figure 1-5E-G; ps<.005, paired t-test comparing Jenson-Shannon divergence of error

distributions). Together, the variance data provides additional evidence in favor of adaptation driving the repulsive biases that were observed in the BOLD data and awareness of the current state of adaptation being a requisite condition for the observed attractive serial dependence. More generally, this model has notable advantages that can lead to enhanced discrimination, reduced energy usage, and improved discrimination in naturalistic conditions over a static labeled-line representation.

Discussion

In this study, we sought to understand the neural underpinning of attractive serial dependence, and how changes in tuning properties at encoding shape behavior. Based on previous behavioral and neural studies, we expected to observe attractive biases in line with observed behavior and decoding from early visual areas (St. John-Saaltink et al., 2016). Instead, we found that representations were significantly repelled from the previous stimulus starting in primary visual cortex and continuing through IPS0 (Figure 1-2I). This repulsion is consistent with bottom up adaptation beginning either at or before V1 and cascading up the visual hierarchy (Dhruv and Carandini, 2014; Patterson et al., 2014; Fritsche et al., 2022). As repulsive biases are in the opposite direction as behavioral biases, we built a model to link these conflicting patterns. The critical new insight revealed by the model is that only readout schemes that account for adaptation can explain the attractive behavioral bias observed in our paradigm. More generally, our BOLD data argue against an early sensory origin of serial dependence for orientation and instead suggest that serial dependence is driven by post-perceptual or mnemonic circuits (Pascucci et al., 2019; Barbosa et al., 2020). However, because we used a paradigm that required working memory, our results may not generalize to other situations in which serial dependence is observed even in the absence of a memory delay (Fischer and Whitney, 2014; Cicchini et al.,

2017; Collins, 2020; Murai and Whitney, 2021). Thus, future work is needed to better understand the role of sensory representations in paradigms with low contrast stimuli, that do not require a memory delay period, and that utilize other features besides orientation.

There have been many prior studies arguing for either a perceptual or post-perceptual origin of serial dependence. Some behavioral studies have found that serial dependence emerges almost immediately after the offset of a stimulus, pointing to an early perceptual origin of the effect (Fischer and Whitney, 2014; Cicchini et al., 2017, 2018, 2021). One study additionally demonstrated that attraction to the previous stimulus seems to occur before the 'tilt-illusion' driven by concurrently presented flanking stimuli (Cicchini et al., 2021). If history biases indeed operate before spatial context, this could point to a distinct assimilative mechanism for serial dependence in early visual processing which may only emerge under low stimulus drive. As our experiment always utilizes a working memory delay, it is unclear if the bias towards past stimuli is driven by a change in their perception of the stimulus itself or instead somehow biases their comparison with the probe stimulus only after the working memory maintenance period.

Others have found that serial dependence is repulsive at very short delays and only becomes attractive when items are held for an extended time in working memory (Papadimitriou et al., 2015; Bliss et al., 2017). This apparent discrepancy was reconciled by (Manassi et al., 2018), who showed that attractive biases disappear without a working memory delay, unless the stimuli are rendered at a very low contrast. This observation suggests that serial dependence may emerge immediately when high sensory uncertainty is induced by low contrast stimuli, and it may emerge later if high sensory uncertainty is induced by extended working memory delay periods. It is curious that unlike some spatial working memory studies (Papadimitriou et al., 2015; Bliss et al., 2017; Stein et al., 2020), we did not find that behavioral biases increased with

delay time. One possible explanation is that this phenomenon is actually unique to spatial working memory due to either a more consistent increase in sensory uncertainty of spatial location due to eye movements or a separate mechanism of memory maintenance that becomes more susceptible to proactive interference relative to orientation memories. Separately, as our stimuli were presented at the fovea (unlike spatial paradigms) they are encoded by a larger population and thus may be less susceptible to degradation across time.

Evidence for an early sensory origin of serial dependence comes from an fMRI study with low contrast stimuli and a short (500 ms) delay period which reported that both behavioral responses and V1 representations were more precise following a matching stimulus (St. John-Saaltink et al., 2016). This departure from our own finding could be driven by the stimuli that were rendered to have a very high uncertainty. Past work studying adaptation in non-human primates found repulsive patterns following long (4s and 40s) but attractive patterns following short (0.4s) stimulus presentations suggesting stimulus duration may have a large influence on how past stimuli shape future sensory processing (Patterson et al., 2013). That said, the stimuli used in the fMRI study of (St. John-Saaltink et al., 2016) were always one of two orthogonal orientations, which, given a circular feature space like orientation, precludes an assessment of attraction or repulsion. Furthermore, correct motor responses were directly yoked to the stimulus so any behavioral tendency to report seeing the same stimulus on successive trials could be due to motor priming rather than stimulus based serial dependence (e.g. a “stay” bias). Related work has shown the ability to decode the previous stimulus from EEG activity patterns (Fornaciai and Park, 2018; Bae and Luck, 2019; Bae, 2021) but it is important to note that our study also showed robust decoding of the previous stimulus that did not also correspond with an attractive bias in the neural representation of the current stimulus (Figures 1-2F-G and 1-9). This is

because the representations of current and past stimuli are not necessarily stored using the same code. Thus, while previous neural studies have argued that serial dependence emerges in visual cortex, no study has demonstrated an attraction towards the previous stimulus dependent on feature similarity consistent with behavioral biases. Further work examining neural biases using low contrast stimuli will shed further light on a potential role of coding changes in sensory cortex driving serial dependence.

In contrast to studies favoring an early sensory account – and more in line with the paradigm and findings reported in this manuscript – a single unit recording study in non-human primates used high contrast stimuli and a longer working memory delay (1.4-5.6s) (Papadimitriou et al., 2016). Under these conditions, neural responses in the Frontal Eye Fields (FEF) were *repelled* from the previously remembered location even though saccades were attracted to the previously remembered location. Given the tight link between the FEF and attentional control (Moore and Fallah, 2001; Schall, 2004; Moore and Zirnsak, 2017), the authors speculated that the observed neural repulsion was due to residual attentional shifts carrying over from the previous trial. However, our observation of repulsive biases starting in V1 and persisting across later visual areas suggests that bottom-up adaptation may be a viable alternative explanation (which the authors also acknowledged). Further support for this account comes from a recent MEG study showing that representations were repelled from past stimuli both within the current trial and from the previous trial (Hajonides et al., 2021). As in our study, this neural repulsion contrasts with attractive behavioral biases to the previous stimulus, suggesting sensory representations do not directly shape behavior even in simple sensory paradigms (Siegle et al., 2021). Behavioral studies using similar high contrast orientation stimuli to our own have also shown that responses are attracted to past decisions and repelled from past stimuli, further

suggesting these attractive biases do not emerge in early sensory areas (Pascucci et al., 2019; Sadil et al., 2021; Moon and Kwon, 2022). Several modeling studies additionally suggest that serial biases are mediated by later readout circuits due to synaptic changes arising from persistent bump attractor dynamics as opposed to early sensory processing (Bliss and D'Esposito, 2017; Barbosa et al., 2020). Thus, in line with our findings: behavioral, neuronal, and modeling studies utilizing high contrast stimuli in working memory paradigms consistently point to attractive effects emerging in either memory or decision-making circuits and not early sensory areas.

In line with classic accounts, adaptation in visual cortex should lead to a reduction in energy usage during encoding (Clifford et al., 2000). However the main advantage of adaptation may be to decorrelate inputs, thus enhancing the discriminability of incoming stimuli (Clifford et al., 2000, 2007) and even acting as a form of short term memory (Hu et al., 2021). An optimal processing stream may emphasize differences at encoding and only favor stability once a stimulus has been selected by attention for more extensive post-perceptual processing (Pascucci et al., 2019). This motif of pattern separation followed by pattern completion would not be unique to adaptive visual processing. Similar mechanisms have been proposed as a critical component of long term memory processing in the hippocampus and associative memory formation in the fly mushroom body (Cayco-Gajic and Silver, 2019). Thus, the biases introduced by adaptation may be beneficial in part because they expand the dimensionality of the representational space as we found in our recordings (Figure 1-10).

We did not explicitly define how awareness of adaptation is implemented, but it is clear that both attention to and conscious awareness of the previous stimulus are necessary for serial dependence to occur (Fischer and Whitney, 2014; Kim et al., 2020). This is consistent with our model, and it suggests that some representation of information about stimulus history should be a

minimum requirement for an aware decoding scheme. The identity of the previous stimulus for spatial position and angle has been shown to be decodable from the spiking activity of single units in the frontal eye field (FEF) and posterior parietal cortex (PPC) as well as large-scale activity patterns in human EEG and MEG (Papadimitriou et al., 2016; Akrami et al., 2018; Fornaciai and Park, 2018; Bae and Luck, 2019; Bae, 2021; Hajonides et al., 2021). We additionally demonstrate that information about the previous trial is encoded in patterns of fMRI activity in human visual cortex (Figure 1-2F), but not in a sensory-like code (Figure 1-9A-B). These signals could potentially be represented concurrently with representations of the current stimulus in the same populations of sensory neurons but in orthogonal codes analogous to what has been found for sequentially encoded items in primate prefrontal cortex and human EEG (Wan et al., 2020; Xie et al., 2022). An alternate account holds that representations of stimulus history are maintained outside of early visual areas, consistent with findings from mouse parietal and primate prefrontal cortex (Akrami et al., 2018; Barbosa et al., 2020). This anatomical segregation could disambiguate incoming sensory drive from representations of stimulus history. Critically, optogenetically suppressing non-sensory representations of stimulus history eliminated history effects, thus providing strong support for some form of an aware readout mechanism (Akrami et al., 2018).

For the decoding stage of our model, we established that only readout schemes that are aware of adaptation can explain attractive serial dependence. The *Bayes-aware* model is an extension of previously proposed models that employ an explicit prior but that did not consider effects of adaptation at encoding (van Bergen and Jehee, 2019b). In contrast, the *over-aware* model is a novel account that can achieve similar performance without needing an explicit prior based on stimulus history. While model fit metrics did not readily distinguish one of these two

models as superior, the *over-aware* model may prove to be more flexible. For instance, one of our fMRI participants showed significant repulsion from far stimuli, an observation also reported by others (Bliss et al., 2017; Fritsche et al., 2017). While the *over-aware* model can fit this repulsive regime, the *Bayes-aware* model is incapable of generating repulsive patterns (compare models fits for subj #3, Figure 1-13). This limitation of a purely Bayesian account of serial dependence is also observable in prior work (Figure 1-6B in (van Bergen and Jehee, 2019b)).

The *over-aware* model proposed in our study may instead be a special condition of a decoder with "fixed awareness" that is based on temporal transition probabilities in natural scenes that are steeply peaked around 0 (no change) over short time scales (Dong and Atick, 1995; Felsen et al., 2005; van Bergen and Jehee, 2019b). Such a readout would correct for the most encountered levels of adaptation by accounting for the transition probabilities of stimuli while being 'fixed', or inflexible, when stimuli violate these expectations. This decoder could account for additional phenomena not directly assessed in the present study such as the tilt after-effect (TAE). The TAE and other forms of (repulsive) behavioral adaptation are often ascribed to an unaware decoder (Seriès et al., 2009; Benucci et al., 2013) but might instead reflect levels of adaptation that exceed the fixed level of adaptation expected by a "fixed-aware" decoder due to long presentations or high contrast stimuli. This is supported by an apparent disconnect in the magnitude of repulsive biases between behavior and neural representations (Dragoi et al., 2001; He and MacLeod, 2001). In contrast, the fixed awareness decoder would lead to attractive biases (serial dependence) when stimuli create less bottom-up drive than expected (e.g., through brief presentations or low contrast items). This 'fixed-aware' decoder is consistent with previous findings of attractive biases disappearing or switching to repulsion when stimulus contrast or duration is increased (Fischer and Whitney, 2014; Manassi et al., 2018). This scheme could

extend to spatial adaptation such as the tilt-illusion where the joint probability of center and surround orientations being perfectly distinct would be vanishingly rare in natural scenes (Howe and Purves, 2005; Schwartz et al., 2007, 2009).

In this study, we extended previous descriptions of serial dependence by quantifying how both bias and variance are shaped by stimulus history. We report a robust pattern of perception being most precise following small changes in successive stimulus features (Figure 1-1F-G, 1-2A-B). This relationship violates a proposed perceptual 'law' that bias is inversely proportional to the derivative of discrimination thresholds (Wei and Stocker, 2017). This account would assert that our attractive bias should come with a less precise representation following small changes (or a repulsive bias to account for our enhanced precision). We argue that serial dependence is not violating this law, but rather believe this is further evidence for delay dependent serial dependence being a post-sensory phenomenon. Neural representations exhibit repulsive biases, expanding the perceptual space and allowing greater discriminability (Figure 1-10). When these representations are read out by an aware decoder, the bias is undone but the enhanced discriminability remains (Figure 1-5D,G).

Methods

Participants

Behavioral study: 56 participants (male and female) were drawn from a subject pool of primarily undergraduate students at UC San Diego. All participants gave written consent to participate in the study in accordance with the UC San Diego IRB (approval number 180067) and were compensated either monetarily or with class credit. Of these 56 participants, 9 were removed from further analysis for completing less than 200 trials (2) or getting less than 60% of

trials correct (7). We included the remaining 47 participants who completed on average 421 trials, range: [204, 988], in our lab over the course of 1 to 3 sessions.

fMRI study: 6 participants (3 female, mean age 24.6 ±0.92) participated in four, 2-hour scanning sessions. Each subject completed between 748 and 884 trials (mean 838.7). For two participants, one session had to be repeated due to technical difficulties that arose during scanning.

Behavioral Discrimination Task

Participants in the behavior-only study completed the task on a desktop computer in a sound attenuated room. Subjects were seated with a chin rest to stabilize viewing 50 cm from a 39 by 29 cm CRT monitor (1600x1200 px) with a visual angle of 42.6° (screen width). Each trial consisted of a full-field oriented grating (1000 ms) which had to be remembered across a delay period (3,500 ms) before a test. At test, the participant judged whether a line was slightly clockwise (CW) or counter-clockwise (CCW) relative to the remembered orientation (max response time window: 3,000ms, Figure 1-1A). The oriented grating consisted of a sine wave grating (spatial frequency 1.73 cycles/°, 0.8 Michelson contrast) multiplied by a 'donut' mask (outer diameter Ø=24.3°, inner Ø=1.73°). The stimulus was then convolved with a 2D Gaussian filter (1.16° kernel, SD = 0.58°) to minimize edge artifacts (Roth et al., 2018). Phase and orientation were randomized across trials, and the stimulus was phase-reversed every 250ms. After the offset of the oriented grating, a mask of filtered noise was presented for 500ms. The mask was generated by band passing white noise [low 0.22, high 0.87 cycles/°], multiplying by the same donut mask, and convolving with a 2D Gaussian filter (0.27° kernel, SD = 0.11°). The mask was phase reversed once after 250 ms. A black fixation point (diameter .578°) was displayed throughout the extent of the block and turned white for 500 ms prior to stimulus onset

on each trial. The probe was a white line (width 0.03°, length 24.3°) masked by the same donut. Subjects indicated whether the probe line was CW or CCW from the remembered orientation by pressing one of two buttons ('Q', 'P') with their left and right pointer fingers. The next trial started after a 1000ms inter trial interval (ITI). For some behavioral participants (n=9) delay and ITI were varied between 0.5-7.5s without notable effects on performance.

First, subjects completed a training block to ensure that they understood the task. Next, they completed a block of trials where difficulty was adjusted by changing the probe offset ($\delta\theta$) between the stimulus and probe to achieve 70% accuracy. This $\delta\theta$ was used in subsequent blocks and was adjusted on a per-block basis to keep performance at approximately 70%. Participants completed an average of 5.76±0.24 blocks [min = 3, max = 9]. Some participants completed the task with slight variations in the distribution and sequence of orientations presented. For completeness we include those details here. Note, however, we additionally report a set of control analyses in which we repeat all of our main analyses excluding blocks with binned stimuli and find no relevant difference in behavior. For most participants, stimuli were pseudo-randomly distributed across the entire 180° space such that they were uniformly distributed across blocks of 64 trials (n=25). However, some participants saw stimuli that were binned (with some jitter) every 22.5° to purposefully avoid cardinal and oblique orientations (11.25°, 33.75°, 56.25°, etc.) and the trial sequence was ordered so that a near oblique orientation was always followed by a near cardinal orientation (n=7). This was implemented to maximize our ability to observe serial dependencies in our binary response data as it is typically strongest around orientation changes of 20° and is more pronounced around oblique orientations (Cicchini and Burr, 2018). The remaining participants completed both blocks with uniform and blocks with binned stimuli (n=14). All participants were interviewed after the study and reported that stimuli

were non-predictable and that all orientations felt equally likely. For our main analysis we include all trials from all participants, irrespective of whether they participated in uniform blocks, binned blocks, or both.

fMRI Discrimination Task

In the scanner, participants completed the behavioral task outlined above with slight modifications. fMRI participants completed the task using a fiber-optic button box while viewing stimuli through a mirror projected onto a screen mounted inside of the bore. The screen was 24 by 18 cm and was viewed at a distance of 47 cm (width: 28.6° visual angle; 1024x768 px native resolution). The stimulus timing was the same except that the sample-to-probe delay period was either 5, 7 or 9 s and the ITIs were uniformly spaced between 5s and 9s and shuffled pseudo-randomly on each run of 17 trials. The oriented gratings had a spatial frequency of 1.27 cycles/°, outer Ø=21.2°, inner Ø=2.37° and were smoothed by a Gaussian filter (0.79° kernel, sd=0.79°). The noise patch (SF low 0.16, high 0.63 cycles/°) was also smoothed by a Gaussian filter (0.29° kernel, sd=0.11°). The probe stimulus was a white line (width = 0.03°).

fMRI participants completed 44-52 blocks of 17 trials spread across 4, two-hour scanning sessions for a total of 748-884 trials. As in the behavior-only task described above, 4 out of 6 fMRI subjects had some blocks of trials where the stimuli were binned in 22.5° increments and ordered in a non-independent manner (21-24 blocks/participant). However, all of the fMRI subjects also participated in blocks with a uniform distribution of orientations across the entire 180° space (24-52 blocks/participant). For our main analysis we include all trials from all participants. However, as with the behavioral analyses, we also report control analyses in which we repeat all of our main analyses excluding blocks with non-random stimuli.

fMRI Localizer Task

Interleaved between the main task blocks, participants completed an independent localizer task used for voxel selection where they were presented with a sequence of grating stimuli at different orientations. Stimuli had a pseudo-randomly determined orientation that either matched the spatial location occupied by the *donut* stimuli used in our main task (outer diameter Ø=21.2°, inner diameter Ø=2.37°) or were a smaller foveal oriented Gabor corresponding to the 'hole' in the *donut* stimuli (diameter Ø=2.37°). Participants were instructed to attend to one of three features orthogonal to orientation depending on the block: detect a contrast change across the entire stimulus, detect a small grey blob appearing over part of the stimulus, or detect a small change in contrast at the fixation point. Each stimulus was presented for 6000 ms and was separated by an ITI ranging from 3-8s.

Response Bias

Each trial consisted of a stimulus and a probe separated by a probe offset ($\delta\theta$) that was either positive (probe is CW of stimulus) or negative. We report degrees in a compass-based coordinate system such that 0° is vertical and orientation values increase moving CW (eg. 30° would point towards 1 o-clock). Participants judged whether the probe was CW or CCW relative to the remembered orientation by making a binary response. To quantify the precision and the response bias, we fit participant responses with a Gaussian cumulative density function with parameters μ and σ corresponding to the *bias (mean)* and *standard deviation* of the distribution. The likelihood of a given distribution was determined by the area under the curve (AUC) of the distribution of CW (CCW) offsets between the stimulus and the probe ($\delta\theta$) on trials where the participant responded CW (CCW). In extreme cases, a very low standard deviation (σ) value with no bias would mean that all $\delta\theta$ would lie outside the distribution and the participant would

get every trial correct. A high negative bias (μ) value would mean that $\delta\theta$ would always lie CW relative to the distribution and the participant would respond CW on every trial. The best fitting parameters were found using a bounded minimization algorithm (limited memory BFGS) on the negative log likelihood of the resulting responses (excluded the small number of trials without a response) given the generated distribution (SciPy 1.0 Contributors et al., 2020). We included a constant 25% guess rate in all model fits to ensure the likelihood of any response could never be 0 (critical for later modelling). While this was critical to fitting our model to raw data, the specific choice had no qualitative effect on our behavioral findings besides making the σ values smaller compared to having a 0% guess rate. By having a constant guess rate rather than varying it as a free parameter we were able to directly compare σ values across participants as a measure of performance.

Serial Dependence

To quantify the dependence of responses on previous stimuli, we analyzed response bias and variance as a function of the difference in orientation between the previous and current orientation ($\Delta\theta = \theta_{n-1} - \theta_n$). We performed this analysis using a sliding window of 32°, such that a bias centered on 16° would include all trials with a $\Delta\theta$ in the range [0°, 32°].

We additionally fit a Derivative of Gaussian (DoG) function to parameterize the bias of participant responses. The DoG function is parameterized with an amplitude A and width w

$$y = xAwce^{-(wx)^2} \quad [1]$$

where $c = \sqrt{2e}$ is a normalization constant. For the purpose of fitting to our participant responses, x is $\Delta\theta$ and y corresponds to μ in our response model. For each participant we adjusted three parameters: A, w, and σ to maximize the likelihood of participant responses. We

report the magnitude of our fits as well as the resulting full width at half max (FWHM) estimated numerically.

Response Precision

In addition to quantifying how responses were biased as a function of stimulus history, we also estimated how precise responses were depending on their unsigned distance from the previous stimulus ($|\Delta\theta|$). When quantifying variance difference between close and far trials, we 'folded' trials with $\Delta\theta < 0$ so that the bias would generally point in the same direction and not artificially inflate our variance measure. Values from the bin with more samples (typically 'far') were resampled (31 repetitions) without replacement with the number of samples in the smaller bin and the median chosen to control for sample number differences.

Scanning

fMRI task images were acquired over the course of four 2-hour sessions for each participant in a General Electric Discovery MR750 3.0T scanner at the UC San Diego Keck Center for Functional Magnetic Resonance Imaging. Functional echo-planar imaging (EPI) data were acquired using a Nova Medical 32-channel head coil (NMSC075-32- 3GE-MR750) and the Stanford Simultaneous Multi-Slice (SMS) EPI sequence (MUX EPI), with a multiband factor of 8 and 9 axial slices per band (total slices 72; 2-mm3 isotropic; 0-mm gap; matrix 104 x 104; field of view 20.8 cm; TR/TE 800/35 ms; flip angle 52°; in-plane acceleration 1). Image reconstruction and un-aliasing was performed on cloud-based servers using reconstruction code from the Center for Neural Imaging at Stanford. The initial 16 repetition times (TRs) collected at sequence onset served as reference images required for the transformation from k-space to the image space. Two 17s runs traversing k-space using forward and reverse phase-encoding directions were collected in the middle of each scanning session and were used to correct for

distortions in EPI sequences using FSL top-up (FMRIB Software Library) for all runs in that session (Andersson et al., 2003; Jenkinson et al., 2012). Reconstructed data was motion corrected and aligned to a common image. Voxel data from each run was de-trended (8TR filter) and z-scored.

We also acquired one additional high-resolution anatomical scan for each subject (1 x 1 x 1-mm3 voxel size; TR 8,136 ms; TE 3,172 ms; flip angle 8°; 172 slices; 1-mm slice gap; 256x192-cm matrix size) during a separate retinotopic mapping session using an Invivo eight-channel head coil. This scan produced higher quality contrast between gray and white matter and was used for segmentation, flattening, and visualizing retinotopic mapping data. The functional retinotopic mapping scanning was collected using the 32-channel coil described above and featured runs where participants viewed checkerboard gratings while responding to an orthogonal feature (transient contrast changes). Separate runs featured alternating vertical and horizontal bowtie stimuli; rotating wedges; and an expanding donut to generate retinotopic maps of the visual meridian, polar angle, and eccentricity respectively (Sprague and Serences, 2013). These images were processed using FreeSurfer and FSL functions and visual regions of interest (ROI) were manually drawn on surface reconstructions (for areas: V1-V3, V3AB, hV4, and IPS0).

Voxel Selection

To include only voxels that showed selectivity for the location of the oriented grating stimulus used in our main experimental task, we used responses evoked during the independent *localizer* task (see fMRI Localizer Task). For all analyses we used TRs 5-11 (4-8.8s) following stimulus onset. First, voxels were selected based on their response to the spatial location of the grating stimulus by performing a t-test on the responses of each voxel evoked by the donut and

the donut-hole stimuli, selecting the 50% of the voxels most selective to the donut for a given ROI. Of the voxels that passed this cutoff, we then performed an ANOVA across 10° orientation bins and selected the 50% of voxels with the largest F-score thus retaining ~25% of the initial voxel pool. These selected voxels were used in all main analyses.

Orientation Decoding

We performed orientation decoding by training an inverted encoding model (IEM) (Brouwer and Heeger, 2009) on BOLD activation patterns using a sliding temporal window of 4 TRs. For most analyses we focused on a 3.2s (4 TR) window centered 6.4 s after stimulus presentation. We first designed an encoding model which assumes voxels are composed of populations of neurons with tuning functions centered on one of 8 orientations evenly tiling the 180° space. The response of population i to stimulus θ is given by:

$$c_i(\theta) = \max\left(0, cos^5(\theta - \omega_i)\right) \quad [2]$$

where ω_i is the center of the tuning function. The response of voxel j is defined as a weighted sum of these hypothetical populations:

$$B_j = \sum_{i}^{8} c_i w_i \quad [3]$$

Or in matrix notation,

$$B = CW \quad [4]$$

Where B *(trial x voxel)* is the resulting BOLD activity, C *(trial x channel)* is the hypothetical population response, and W *(channel x voxel)* is the weight matrix. The weight matrix W is estimated as:

$$\widehat{W} = C^{-1}B \quad [5]$$

where C^{-1} (*channel x trial*) is the pseudo-inverse of C (implemented using the NumPy pinv function). We then estimated channel responses using the inverse of our estimated weight matrix:

$$\hat{C} = B\hat{W}^{-1} \quad [6]$$

This channel response corresponds to a representation of orientation activity. To decode orientation, we took the inner product with a vector of the tuning curve centers in polar coordinates. The angle of the resulting vector was taken as the estimated orientation ($\hat{\theta}$) while the vector length was taken as a proxy for model certainty ($\hat{R}$).

$$\hat{\theta} = angle(\hat{C}e^{i\omega}) \quad [7]$$

$$\hat{R} = \left\|\hat{C}e^{i\omega}\right\| \quad [8]$$

The weight matrix of our model was estimated from a subset of our data and used to estimate orientation representations on a held-out portion of the task data. We used leave-1-block-out cross-validation where each block was a set of 4 consecutive runs (64 trials). These blocks had orientations that were linearly spaced across the entire 180°, with a random phase offset for each block, to ensure a balanced training set. We performed an additional analysis training a model on all data from the localizer task and testing on the memory task. This model had lower SNR than models trained on the task but showed qualitatively similar results as our task trained neural decoder.

Kernel Based Decoding

Estimating average voxel HRFs through deconvolution

Because we are measuring the effects of previous stimuli on responses to the current stimulus, we did an additional analysis to quantify any influence of overlapping Hemodynamic Response Functions (HRF) that last for 20-30s (e.g. the "undershoot" that happens

approximately 8-18s post-stimulus; see Figure 1-3A). To account for overlapping HRFs, we used deconvolution to estimate the average univariate response separately in each voxel in each ROI by modeling the responses to both the stimulus and probe for 30 TRs (24s) post-stimulus (Dale, 1999; Glover, 1999). We created a design matrix (rows x columns = total number of TRs x 30) with the first column containing ones corresponding to the onset TR of each stimulus (and zeros elsewhere). Subsequent columns were the same vector shifted forward in time by one TR. Following the same procedure, another design matrix was defined for the probe onset times. These matrices were stacked with a column of ones added for each run as a constant term, yielding a final design matrix X of dimensions (total number of TRs x (60+number of blocks)). We created a related matrix of voxel activity Y (total number of TRs x number of voxels) by concatenating responses in each voxel across blocks. We then estimated the HRF by performing least squares regression using the normal equation:

$$h = (X^T X)^{-1}(X^T Y). \quad [9]$$

The resulting weights corresponded to the average timecourse of the HRF evoked separately by the stimulus and the probe across all trials. We note that this HRF is estimated independent of the orientation of the presented stimuli as we wanted to use these estimates to then decode orientation dependent changes in activation patterns. For each voxel we then parameterized the HRF using a 6-parameter double gamma function using scipy.optimize.minimize so that we could use the voxel-specific HRF model in a GLM to estimate the response magnitude in each voxel on each trial. We excluded the 11% of voxels which failed to converge on a solution.

Estimating trial-by-trial responses using parameterized voxel HRFs

For each voxel, we then created a design matrix Xv (rows x columns = total number of TRs x (number of trials * 2 +number of blocks)) with each column a delta function centered at

the onset of the stimulus (or probe). We then regressed this matrix onto the (total # of TRs) vector Yv of voxel activity using equation 9. This resulted in a simultaneous estimation of the trial-by-trial magnitude of responses to each stimulus grating and each probe which was repeated for each voxel to allow voxel specific HRFs to be utilized in the creation of Xv. The resulting activity pattern associated with each stimulus was used in the same manner as the raw time course of the BOLD response to train and test an IEM, and the resulting estimates should be largely independent of linear contributions of previous stimuli (Dale, 1999).

Neural Bias

To quantify how BOLD representations were biased by sensory history we computed the circular mean of decoding errors ($\theta_{error} = \text{wrap}(\theta_{decode} - \theta_{stim})$):

$$\mu_{circ} = angle(\vec{R}), \quad [10]$$

$$\vec{R} = \frac{1}{nTrials}\sum_{k=0}^{nTrials} e^{i\theta_{error}^{k}}. \quad [11]$$

We estimated this bias using the same 32° sliding window as a function of $\Delta\theta$ used for visualizing response bias from participant responses. We additionally quantified the magnitude of the bias in decoding errors by fitting a DoG function to the raw decoding errors by minimizing the residual sum of squares (RSS) and reporting the amplitude term.

Neural Variance

To quantify the variance of decoded orientations from visual areas, we computed the circular standard deviation on binned decoding errors:

$$\sigma_{circ} = \sqrt{-2\ln|\vec{R}|}. \quad [12]$$

This was visualized using the same sliding window analysis as well as in reference to whether it was close or far from the previous stimulus.

Dimensionality Analysis

To quantify how stimulus history shaped the structure of neural responses independent of neural tuning we utilized principal component analysis (PCA). For a given set of neural responses R (number of trials x number of voxels) we mean centered and performed eigenvalue decomposition on the (number of voxels x number of voxels) covariance matrix. Eigenvalues were sorted in descending order and our response matrix was projected into PCA space (for visualization purposes) by multiplying by the sorted eigenvectors.

To compare dimensionality across conditions, we sub-set our data into trials following close (<30°) or far (>60°) trials and randomly sub selected trials from the larger group (without replacement) to equate trial numbers. We then performed PCA separately for each group and compared the relative proportion of total variance explained as the magnitude of the sorted eigenvalues. We quantified both the minimum number of components to reach at least 90% of the variance explained and also recorded the mean (AUC) of the variance curve.

Modeling

We sought to develop a model that could explain both neural and behavioral biases as a function of stimulus history. For the fMRI data, we focused on explaining changes in encoding that could lead to the observed biases in the output of the BOLD decoder that was specifically designed to be '*unaware'* of stimulus history. To explain the behavioral data, we assumed that a decoder would receive inputs from the same population of sensory neurons that we measured with fMRI and that the decoder would read out this information in a manner that gives rise to attractive serial dependence. We considered readout models that were either *unaware*, *aware*, or *over-aware* of adaptation and additionally applied a Bayesian inference stage, which integrates prior expectations of temporal stability, to the *unaware* and *aware* decoders (van Bergen and

Jehee, 2019a). We then compared performance between these competing models to see which could best explain our behavioral data.

Our full models consisted of two stages: an encoding stage where the gain of artificial neurons was changed as a function of the previous stimulus (adaptation) and a decoding stage where the readout from this adapted population was modified. The encoding population consisted of 100 neurons with von Mises tuning curves evenly tiling the 180° space. The expected unadapted population response is:

$$Resp_N(\theta_n) = R\ \gamma_N e^{\kappa \cos(\Phi - \theta_n) - 1} \quad [13]$$

Where γ_N is the scalar 1 for constant gain without adaptation, Φ is the vector of tuning curve centers, θ_n is the orientation of the current stimulus, κ=1.0 is a constant controlling tuning width, and R is a general gain factor driving the average firing rate. We implemented sensory adaptation by adjusting the gain of tuning curves relative to the identity of the previous stimulus, θ_{n-1} (Figure 1-4A, *Gain Adaptation*):

$$\gamma_A(\theta_{n-1}) = \gamma_N - rect(\gamma_m \cos^3(\gamma_s(\Phi - \theta_{n-1}))) \quad [14]$$

Where γ_m is the magnitude of adaptation, γ_s scales the width of adaptation, and *rect* is the half-wave rectifying function. The responses of the adapted population thus depend on both the current and previous stimulus (Figrue 1-4A, *Efficient Encoding*):

$$Resp_A(\theta_n, \theta_{n-1}) = R\ \gamma_A e^{\kappa \cos(\Phi - \theta n) - 1} \quad [15]$$

Unaware decoder: We first considered a model in which an adapted orientation-encoding representation is being decoded by an *unaware* readout mechanism (Figure 1-4B). The likelihood of each orientation giving rise to the observed response profile across *N* neurons was estimated assuming activity was governed by a Poisson process:

$$P_{unaware}(\,Resp_A|\theta) = \exp\left(\sum_{i=1}^{N} \log P_{Poisson}\left(Resp_A^i(\theta); Resp_N^i(\theta)\right)\right) \quad [16]$$

$$P_{Poisson}(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad [17]$$

Where $Resp_N^i(\theta)$ is the expected response of the unadapted neuron i to stimulus θ and $P_{Poisson}(k;\lambda)$ is the probability of observing k spikes given an expected firing rate of λ. The decoded orientation is then the θ giving rise to the maximum likelihood (MLE).

Aware decoder: In addition to the unaware decoder, we also evaluated the ability of a decoder that was aware of the current state of adaptation to explain behavior. The *aware* decoder differs from the *unaware* decoder in that its assumed activity level for each unit is modulated as a function of stimulus history:

$$P_{aware}(\,Resp_A|\theta_n;\theta_{n-1}) \quad [18]$$
$$= \exp\left(\sum_{i=1}^{N} \log P_{Poisson}\left(Resp_A^i(\theta_n,\theta_{n-1}), Resp_A^i(\theta_n,\theta_{n-1})\right)\right)$$

Note that here the rate parameter $k \equiv \lambda \equiv Resp_A$ such that the observed and expected values perfectly align with the presented orientation. $P_{aware}(\,Resp_A|\theta_n;\theta_{n-1})$ is dependent on sensory history and is non-biased.

Over-Aware decoder: Our final decoding scheme we call the *over-aware decoder*. This model can test whether serial dependence can be achieved without an explicit stage of Bayesian inference introduced in the next section. The decoder has an assumed adaptation defined by a unique set of free parameters $\gamma_{m2}\ and\ \gamma_{s2}$ which shapes a separate gain adaptation:

$$\gamma_{OA}(\theta_{n-1}) = \gamma_N - rect(\gamma_{m2}\cos^3(\gamma_{s2}(\Phi - \theta_{n-1})))$$ [19]

which in turn shapes the response profile of $Resp_{OA}$ in the same manner as $Resp_A$. The likelihood profile is then defined as:

$$P_{over-aware}(\, Resp_A|\theta) = \exp\left(\sum_{i=1}^{N} \log P_{Poisson}\left(Resp_A^i(\theta); Resp_{OA}^i(\theta, \theta_{n-1})\right)\right)$$ [20]

where our expected (assumed) rate λ is designated by $Resp_{OA}$. By having a larger assumed adaptation than implemented at encoding (through either $\gamma_{m2} > \gamma_m$ or $\gamma_{s2} > \gamma_s$) the net effect of the over-aware decoder should be behavioral attraction.

Bayesian Inference: In addition, we explored the effect of applying an explicit Bayesian prior based on temporal contiguity to the likelihood functions derived from these different readout schemes. This type of prior has been previously used to explain behavioral biases without considering how encoding might also be affected by stimulus history (van Bergen and Jehee, 2019a). Specifically, the prior is defined by the transition probability between consecutive stimuli and is defined as a mixture model of a circular Gaussian and a uniform distribution:

$$P_T(\theta_n|\theta_{n-1}) = \frac{1}{Z} e^{-\frac{angle(\theta,\theta_{n-1})^2}{2\psi^2}}$$ [21]

$$P_{Bayesian}(\theta_n|\theta_{n-1}) = P_{SAME} P_T(\theta|\theta_{n-1}) + \frac{1}{2\pi}(1 - P_{SAME})$$ [22]

with P_{SAME} set to 0.64 (as found empirically in (van Bergen and Jehee, 2019a)), Z as a normalization constant so P_T integrates to 1, and ψ is a free parameter describing the variance of the transition distribution. This prior (Figure 1-4C, black line) is multiplied by the *unaware* likelihood (Figure 1-4C, yellow dashed-line): to get the posterior estimate of our *Bayesian-unaware* decoder (Figure 1-4C, yellow solid-line):

$$P_{Bayesian-unaware}(\, \theta_n|Resp_A;\, \theta_{n-1}) = P_{Bayesian}(\theta|\theta_{n-1}) P_{unaware}(\, Resp_A|\theta_n)$$ [23]

We can additionally examine a *Bayesian-aware* decoder by substituting its respective likelihood function. We did not examine a *Bayesian-over-aware* model so that all decoding models would have the same number of free parameters and so that we could directly evaluate the need for an explicit prior.

Model Fitting: The *encoding* stage of the model has two free parameters and for each subject these parameters were optimized to minimize the residual sum of squares (RSS) between our measured fMRI decoding errors and the decoding errors of our *unaware* decoder. For simplicity we only fit our model to decoding errors from V3 as it had the highest SNR, but other early visual ROIs showed similar results. After fitting the *encoding* stage of the model, we then separately fit the three competing *decoding* models to best account for the behavioral data: *Bayes-unaware*, *Bayes-aware*, and *over-aware* (two free parameters each). The output of this readout stage was treated as the behavioral bias (μ) and the free parameters were optimized to maximize the likelihood of the observed responses (assuming constant standard deviation σ estimated empirically for each participant). For the purposes of fitting the model, the firing rates of the modelled neurons were deterministic (no noise process). Having noiseless activity had no effect on the expected bias (verified with additional simulations) and served to make model fitting more reliable and less computationally intensive. Both stages of the model were fit using the same cross-validation groups as our neural decoder. To ensure all models had a sufficient chance of achieving a good fit to behavioral data, we implemented a grid search sampling 30 values along the range of each variable explored (900 locations total) followed by a local search algorithm (Nelder-Mead) around the most successful grid point. We found dense sampling of the initial parameter space was especially important for our *Bayes-unaware* model.

Model Evaluation: For bias of neural and behavioral responses, we evaluated the performance of the two stages of our model separately. These stages must be evaluated in a qualitatively different manner as the neural data gives us an orientation estimate for each trial while the behavioral data consists of binary responses. For the encoding stage, we quantified how well the output of our *unaware* decoder predicted the raw errors of our BOLD decoder using circular correlation. The performance of this model was contrasted with the true presented orientation which is analogous to the representation of an unadapted population. We additionally computed the variance of the neural decoding errors explained by the model bias (R^2). For the decoding stage of our model, we compared the log-likelihood of observed responses for each model.

We additionally estimated the variance of our models using neurons with rates generated by a Poisson process. The average bias was unaffected by allowing random fluctuations in activity, but the trial-to-trial variance increased. To get a stable estimate, we simulated 1000 trials for each set of parameters estimated for a cross-validation loop for each participant and pooled these outputs. We compared the overall variance of our models to our single parameter estimate of participant precision using Jensen-Shannon divergence. We additionally examined relative precision of our model for close and far trials in the same manner as participant responses and decoding errors (Response Precision).

Data/Code Availability

Processed BOLD and behavioral data as well as code necessary to reproduce all analysis in this study can be found here: https://osf.io/e5xw8/?view_only=e7c1da85aa684cc8830aec8d74afdcb4.

Acknowledgements: Chapter 1

Thanks to Chaipat Chunharas for critical discussions in experimental design and assistance with scanning and to Anika Jollorina and Shuangquan Feng for assistance with behavioral data collection. Thanks to Marcelo Mattar for helpful comments on our model, Gal Mishne for help with setting up dimensionality analyses, and to Margaret Henderson, Sunyoung Park, and Kirsten Adam for thoughtful comments on earlier versions of the manuscript.

Chapter 1, in full, is a reprint of the material as it appears in PLOS Biology, 2022. Sheehan, Timothy C.; Serences, John T. The dissertation/thesis author was the primary investigator and author of this paper.

Works Cited

Abbonizio G, Clifford C, Langley K (2002) Contrast adaptation may enhance contrast discrimination. Spat Vis 16:45–58.

Akrami A, Kopec CD, Diamond ME, Brody CD (2018) Posterior parietal cortex represents sensory history and mediates its effects on behaviour. Nature 554:368–372.

Andersson JLR, Skare S, Ashburner J (2003) How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. NeuroImage 20:870–888.

Bae G-Y (2021) The Time Course of Face Representations during Perception and Working Memory Maintenance. Cereb Cortex Commun 2:tgaa093.

Bae G-Y, Luck SJ (2019) Reactivation of Previous Experiences in a Working Memory Task. Psychol Sci 30:587–595.

Barbosa J, Stein H, Martinez RL, Galan-Gadea A, Li S, Dalmau J, Adam KCS, Valls-Solé J, Constantinidis C, Compte A (2020) Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. Nat Neurosci 23:1016–1024.

Barlow HB (2012) Possible Principles Underlying the Transformations of Sensory Messages. In: Sensory Communication (Rosenblith WA, ed), pp 216–234. The MIT Press. Available at:

http://mitpress.universitypressscholarship.com/view/10.7551/mitpress/9780262518420.001.0001/upso-9780262518420-chapter-13 [Accessed August 29, 2020].

Benucci A, Saleem AB, Carandini M (2013) Adaptation maintains population homeostasis in primary visual cortex. Nat Neurosci 16:724–729.

Bliss DP, D'Esposito M (2017) Synaptic augmentation in a cortical circuit model reproduces serial dependence in visual working memory Bazhenov M, ed. PLOS ONE 12:e0188927.

Bliss DP, Sun JJ, D'Esposito M (2017) Serial dependence is absent at the time of perception but increases in visual working memory. Sci Rep 7:14739.

Brouwer GJ, Heeger DJ (2009) Decoding and Reconstructing Color from Responses in Human Visual Cortex. J Neurosci 29:13992–14003.

Cayco-Gajic NA, Silver RA (2019) Re-evaluating Circuit Mechanisms Underlying Pattern Separation. Neuron 101:584–602.

Cicchini GM, Anobile G, Burr DC (2014) Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. Proc Natl Acad Sci 111:7867–7872.

Cicchini GM, Benedetto A, Burr DC (2021) Perceptual history propagates down to early levels of sensory analysis. Curr Biol 31:1245-1250.e2.

Cicchini GM, Burr DC (2018) Serial effects are optimal. Behav Brain Sci 41:e229.

Cicchini GM, Mikellidou K, Burr D (2017) Serial dependencies act directly on perception. J Vis 17:6.

Cicchini GM, Mikellidou K, Burr DC (2018) The functional role of serial dependence. Proc R Soc B Biol Sci 285:20181722.

Clifford CWG, Webster MA, Stanley GB, Stocker AA, Kohn A, Sharpee TO, Schwartz O (2007) Visual adaptation: Neural, psychological and computational aspects. Vision Res 47:3125–3131.

Clifford CWG, Wenderoth P, Spehar B (2000) A functional angle on some after-effects in cortical vision. Proc R Soc Lond B Biol Sci 267:1705–1710.

Clifford CWG, Wyatt AM, Arnold DH, Smith ST, Wenderoth P (2001) Orthogonal adaptation improves orientation discrimination. Vision Res 41:151–159.

Collins T (2020) Serial dependence alters perceived object appearance. J Vis 20:9.

Crick F, Koch C (1995) Are we aware of neural activity in primary visual cortex? Nature 375:121–123.

Dale AM (1999) Optimal experimental design for event-related fMRI. Hum Brain Mapp 8:109–114.

Dekel R, Sagi D (2015) Tilt aftereffect due to adaptation to natural stimuli. Vision Res 117:91–99.

Dhruv NT, Carandini M (2014) Cascaded Effects of Spatial Adaptation in the Early Visual System. Neuron 81:529–535.

Dong DW, Atick JJ (1995) Statistics of natural time-varying images. Netw Comput Neural Syst 6:345–358.

Dragoi V, Rivadulla C, Sur M (2001) Foci of orientation plasticity in visual cortex. Nature 411:80–86.

Dragoi V, Sharma J, Miller EK, Sur M (2002) Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. Nat Neurosci 5:883–891.

Dragoi V, Sharma J, Sur M (2000) Adaptation-Induced Plasticity of Orientation Tuning in Adult Visual Cortex. Neuron 28:287–298.

Durant S, Clifford CWG, Crowder NA, Price NSC, Ibbotson MR (2007) Characterizing contrast adaptation in a population of cat primary visual cortical neurons using Fisher information. J Opt Soc Am A 24:1529.

Felsen G, Touryan J, Dan Y (2005) Contextual modulation of orientation tuning contributes to efficient processing of natural stimuli. Netw Comput Neural Syst 16:139–149.

Fischer J, Whitney D (2014) Serial dependence in visual perception. Nat Neurosci 17:738–743.

Fornaciai M, Park J (2018) Attractive Serial Dependence in the Absence of an Explicit Task. Psychol Sci 29:437–446.

Fritsche M, Mostert P, de Lange FP (2017) Opposite Effects of Recent History on Perception and Decision. Curr Biol 27:590–595.

Fritsche M, Solomon SG, de Lange FP (2022) Brief stimuli cast a persistent long-term trace in visual cortex. J Neurosci:JN-RM-1350-21.

Fritsche M, Spaak E, de Lange FP (2020) A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. Neuroscience. Available at: http://biorxiv.org/lookup/doi/10.1101/2020.01.22.915553 [Accessed March 2, 2020].

Gardner JL, Sun P, Waggoner RA, Ueno K, Tanaka K, Cheng K (2005) Contrast Adaptation and Representation in Human Early Visual Cortex. :14.

Girshick AR, Landy MS, Simoncelli EP (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nat Neurosci 14:926–932.

Glover GH (1999) Deconvolution of Impulse Response in Event-Related BOLD fMRI1. NeuroImage 9:416–429.

Gold JI, Shadlen MN (2007) The Neural Basis of Decision Making. Annu Rev Neurosci 30:535–574.

Grunewald A, Bradley DC, Andersen RA (2002) Neural Correlates of Structure-from-Motion Perception in Macaque V1 and MT. J Neurosci 22:6195–6207.

Hajonides JE, Ede F van, Stokes MG, Nobre AC, Myers NE (2021) Multiple and Dissociable Effects of Sensory History on Working-Memory Performance. :2021.10.31.466639 Available at: https://www.biorxiv.org/content/10.1101/2021.10.31.466639v1 [Accessed March 2, 2022].

He S, MacLeod DIA (2001) Orientation-selective adaptation and tilt after-effect from invisible patterns. Nature 411:473–476.

Howe CQ, Purves D (2005) The Müller-Lyer illusion explained by the statistics of image-source relationships. PNAS Proc Natl Acad Sci U S Am 102:1234–1239.

Hu B, Garrett ME, Groblewski PA, Ollerenshaw DR, Shang J, Roll K, Manavi S, Koch C, Olsen SR, Mihalas S (2021) Adaptation supports short-term memory in a visual change detection task. PLOS Comput Biol 17:e1009246.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. NeuroImage 62:782–790.

Kim S, Burr D, Cicchini GM, Alais D (2020) Serial dependence in perception requires conscious awareness. Curr Biol 30:R257–R258.

Kiyonaga A, Scimeca JM, Bliss DP, Whitney D (2017) Serial Dependence across Perception, Attention, and Memory. Trends Cogn Sci 21:493–497.

Manassi M, Liberman A, Kosovicheva A, Zhang K, Whitney D (2018) Serial dependence in position occurs at the time of perception. Psychon Bull Rev 25:2245–2253.

Moon J, Kwon O-S (2022) Dissecting the effects of adaptive encoding and predictive inference on a single perceptual estimation. :2022.02.24.481765 Available at: https://www.biorxiv.org/content/10.1101/2022.02.24.481765v1 [Accessed May 21, 2022].

Moore T, Fallah M (2001) Control of eye movements and spatial attention. Proc Natl Acad Sci 98:1273–1276.

Moore T, Zirnsak M (2017) Neural Mechanisms of Selective Visual Attention. Annu Rev Psychol 68:47–72.

Moradi F, Koch C, Shimojo S (2005) Face Adaptation Depends on Seeing the Face. Neuron 45:169–175.

Murai Y, Whitney D (2021) Serial dependence revealed in history-dependent perceptual templates. Curr Biol 31:3185-3191.e3.

Papadimitriou C, Ferdoash A, Snyder LH (2015) Ghosts in the machine: memory interference from the previous trial. J Neurophysiol 113:567–577.

Papadimitriou C, White RL, Snyder LH (2016) Ghosts in the Machine II: Neural Correlates of Memory Interference from the Previous Trial. Cereb Cortex:bhw106.

Pascucci D, Mancuso G, Santandrea E, Della Libera C, Plomp G, Chelazzi L (2019) Laws of concatenated perception: Vision goes for novelty, decisions for perseverance Tong F, ed. PLOS Biol 17:e3000144.

Patterson CA, Duijnhouwer J, Wissig SC, Krekelberg B, Kohn A (2014) Similar adaptation effects in primary visual cortex and area MT of the macaque monkey under matched stimulus conditions. J Neurophysiol 111:1203–1213.

Patterson CA, Wissig SC, Kohn A (2013) Distinct Effects of Brief and Prolonged Adaptation on Orientation Tuning in Primary Visual Cortex. J Neurosci 33:532–543.

Phinney RE, Bowd C, Patterson R (1997) Direction-selective Coding of Stereoscopic (Cyclopean) Motion. Vision Res 37:865–869.

Roth ZN, Heeger DJ, Merriam EP (2018) Stimulus vignetting and orientation selectivity in human visual cortex. eLife 7:e37241.

Sadil P, Cowell R, Huber DE (2021) The Yin-yang of Serial Dependence Effects: Every Response is both an Attraction to the Prior Response and a Repulsion from the Prior Stimulus. Available at: https://psyarxiv.com/f52yz/ [Accessed March 14, 2022].

Schall JD (2004) On the role of frontal eye field in guiding attention and saccades. Vision Res 44:1453–1467.

Schwartz O, Hsu A, Dayan P (2007) Space and time in visual context. Nat Rev Neurosci 8:522–535.

Schwartz O, Sejnowski TJ, Dayan P (2009) Perceptual organization in the tilt illusion. J Vis 9:19.

SciPy 1.0 Contributors et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17:261–272.

Seriès P, Stocker AA, Simoncelli EP (2009) Is the Homunculus "Aware" of Sensory Adaptation? Neural Comput 21:3271–3304.

Siegle JH et al. (2021) Survey of spiking in the mouse visual system reveals functional hierarchy. Nature 592:86–92.

Sprague TC, Serences JT (2013) Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nat Neurosci 16:1879–1887.

St. John-Saaltink E, Kok P, Lau HC, de Lange FP (2016) Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. J Neurosci 36:6186–6192.

Stein H, Barbosa J, Rosa-Justicia M, Prades L, Morató A, Galan-Gadea A, Ariño H, Martinez-Hernandez E, Castro-Fornieles J, Dalmau J, Compte A (2020) Reduced serial dependence suggests deficits in synaptic potentiation in anti-NMDAR encephalitis and schizophrenia. Nat Commun 11:4250.

Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci 9:578–585.

van Bergen RS, Jehee JFM (2019a) Probabilistic Representation in Human Visual Cortex Reflects Uncertainty in Serial Decisions. J Neurosci 39:8164–8176.

van Bergen RS, Jehee JFM (2019b) Probabilistic Representation in Human Visual Cortex Reflects Uncertainty in Serial Decisions. J Neurosci 39:8164–8176.

Wan Q, Cai Y, Samaha J, Postle BR (2020) Tracking stimulus representation across a 2-back visual working memory task. R Soc Open Sci 7:190228.

Wei X-X, Stocker AA (2015) A Bayesian observer model constrained by efficient coding can explain "anti-Bayesian" percepts. Nat Neurosci 18:1509–1517.

Wei X-X, Stocker AA (2017) Lawful relation between perceptual bias and discriminability. Proc Natl Acad Sci 114:10244–10249.

Xie Y, Hu P, Li J, Chen J, Song W, Wang X-J, Yang T, Dehaene S, Tang S, Min B, Wang L (2022) Geometry of sequence working memory in macaque prefrontal cortex. Science 375:632–639.

Supplemental Figures

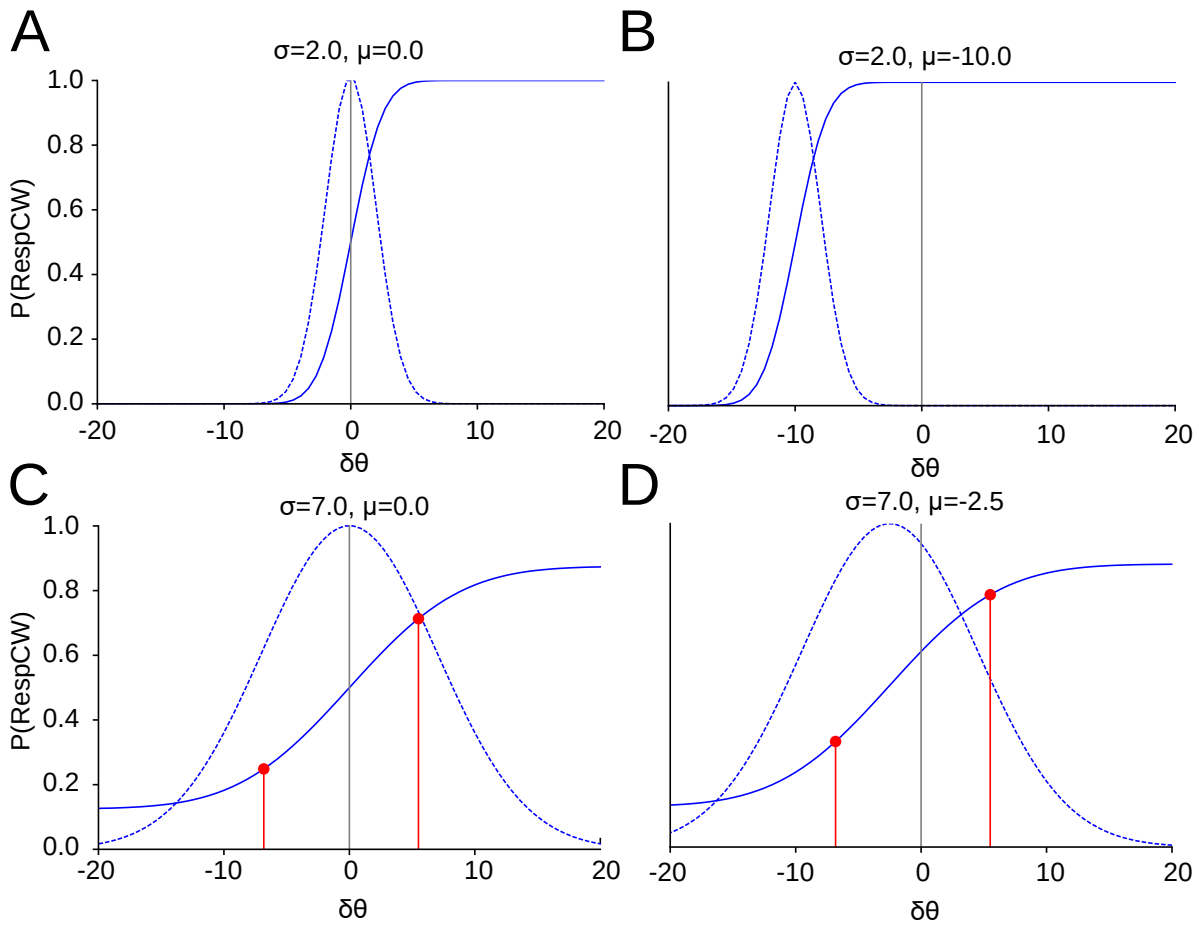


Figure 1-6 Response model.

Encoding of stimulus is assumed to be a noisy process whereby the distribution of encoded orientations is described by a Gaussian pdf with mean μ and standard deviation σ. Dashed line is pdf and solid line is the cdf of encoding distribution. Note that participants are reporting the probes orientation relative to the stimulus so more frequent CCW responses would correspond to a CW perceptual bias. A: Example estimation curve with no bias and a very small σ. If the difficulty was set to $\delta\theta$=6° (3 sd) than this participant would get essentially all (99.7%) trials correct. B: Estimation curve with a μ=-10, this participant would respond CW on almost every trial. C-D: Realistic encoding curves. To aid with fitting and to best describe responses, a constant guess rate of 25% was included in the response model fit to participant responses. C: An unbiased distribution with two theoretical stimuli on which the participant responded CW. The left response $\delta\theta$=-6° is incorrect. D: A CCW biased distribution results in a higher likelihood for all CW responses. Data and code supporting this figure found here: https://osf.io/e5xw8/?view_only=e7c1da85aa684cc8830aec8d74afdcb4

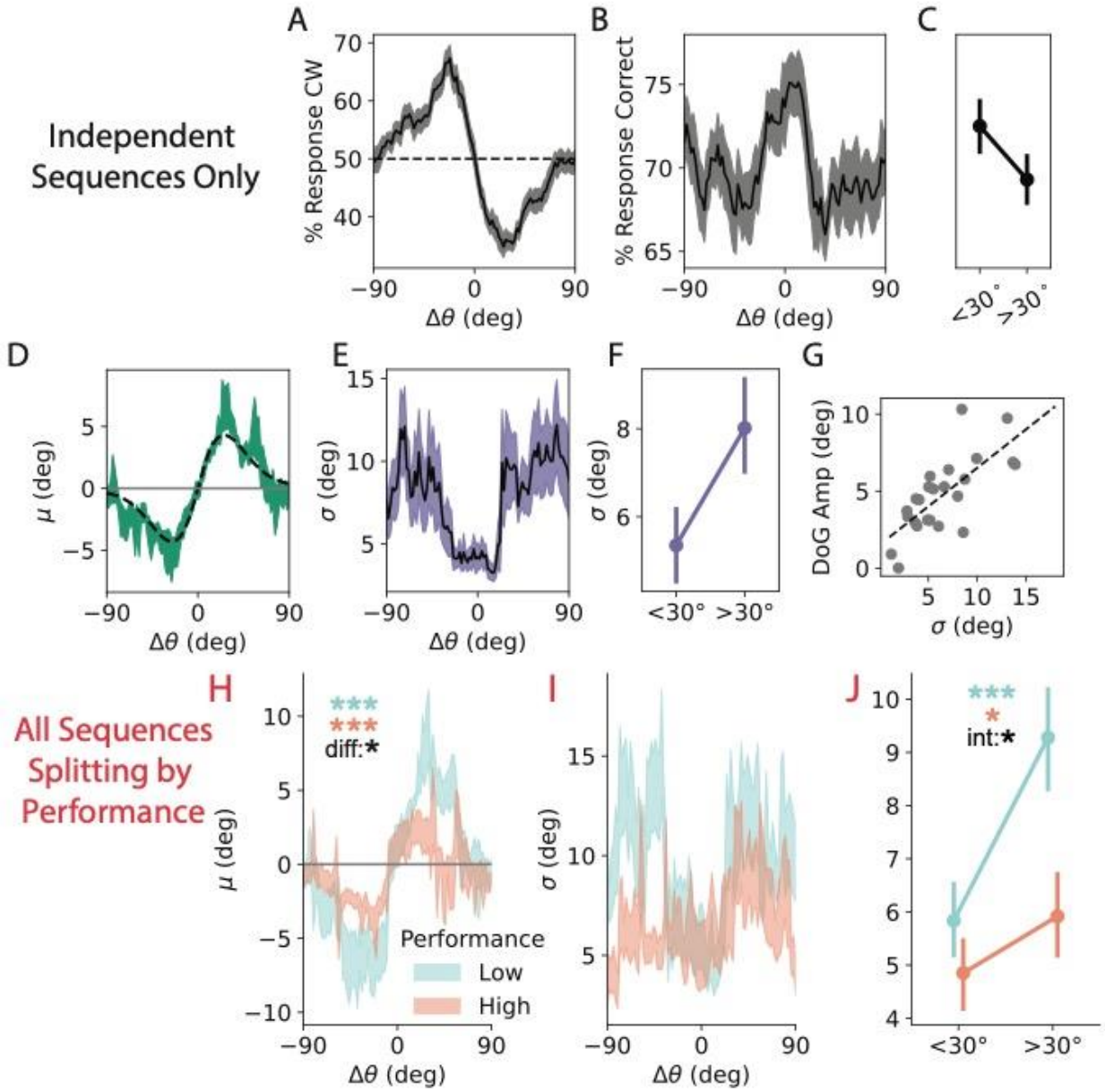


Figure 1-7 A subset of behavior only participants completed a version of the experiment with inhomogeneities in their stimulus sequences (such that consecutive orientations were not independent).

To confirm this manipulation did not drive any of our results, we repeated our behavioral analyses excluding participants with non-independent sequences leaving a cohort of n=25 with an average accuracy of 70.46±1.14° at an average $\delta\theta$ of 4.97±0.35°. A,D: This cohort still showed significant serial dependence (DoG amp =4.71±0.49, t(23) = 9.4, p=2.4*10-9; width 0.027±0.0019, FWHM 43.68±1.86°, B-C: and had responses that were more accurate (t(24)=3.14, p=.0023, E-F: and precise following 'close' stimuli (t(24)=-3.54, p=0.0009, G: Lastly, bias and variance were still positively correlated across this cohort (r(22)=0.72, p=0.00003, H-J: Stimulus history effects are larger for worse performing subjects. H: Serial dependence was significantly greater for less precise participants (t(45)=-2.5, p=.012, unpaired t-test comparing DoG Amplitude). I-J: Variance was modulated significantly by stimulus history (low-performing: t(23)=3.9 p=.0007; high-performing t(22)=2.4, p=.02, one-sample t-tests), with a significant interaction between overall performance and the effect size (p=.017, mixed effects linear model).

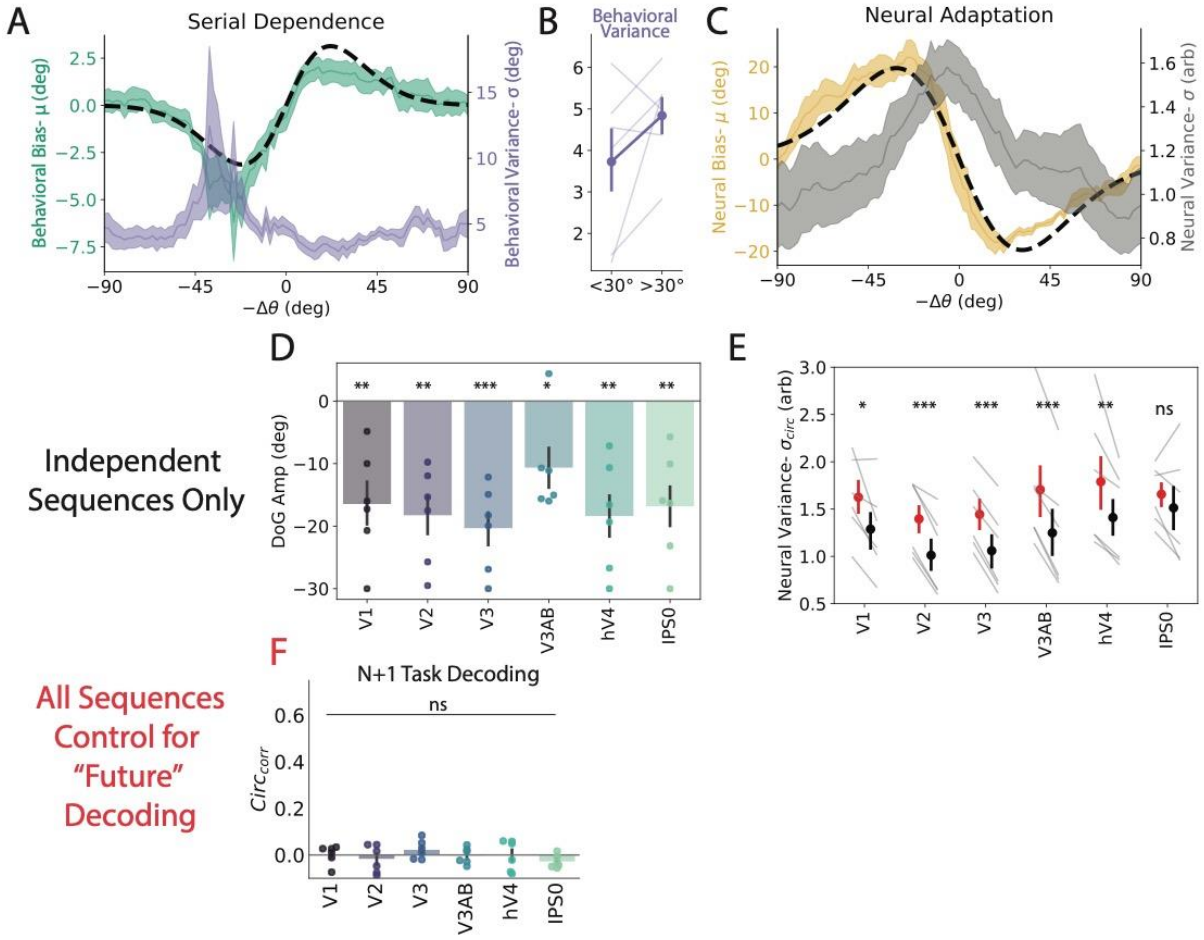


Figure 1-8 A subset of fMRI participants completed some sessions where consecutive stimuli were not strictly independent.

A: To confirm this structure was not driving our results, we repeated our main analyses excluding these sessions and found that responses were still strongly attracted to the previous stimulus (DoG Amp: 3.25± 0.34, t(5)=8.85, p=1.53e-04; DoG FWHM: 36.1±2.9). B: We found that responses were no longer significantly more precise following small changes in orientation but were trending in the same direction as when including all sessions (t(5)=-1.55, p=.09). We additionally confirmed that our finding of reduced bias around small changes in orientation was not driven by the oblique effect in the same manner as the behavioral cohort (mean % cardinal close: 48.6±0.9%, far: 49.8±0.2%, t(5)=-1.0, p=0.36, paired t-test). C-E: We further replicate our finding of neural repulsion and increased uncertainty following 'close' stimuli across all ROIs except IPS0. F: As a control analysis, we attempted but were unable to decode the identity of the next trial in any ROI when including all sequences. ns, not significant; *, p<.05; **, p<.01; ***, p<.001.

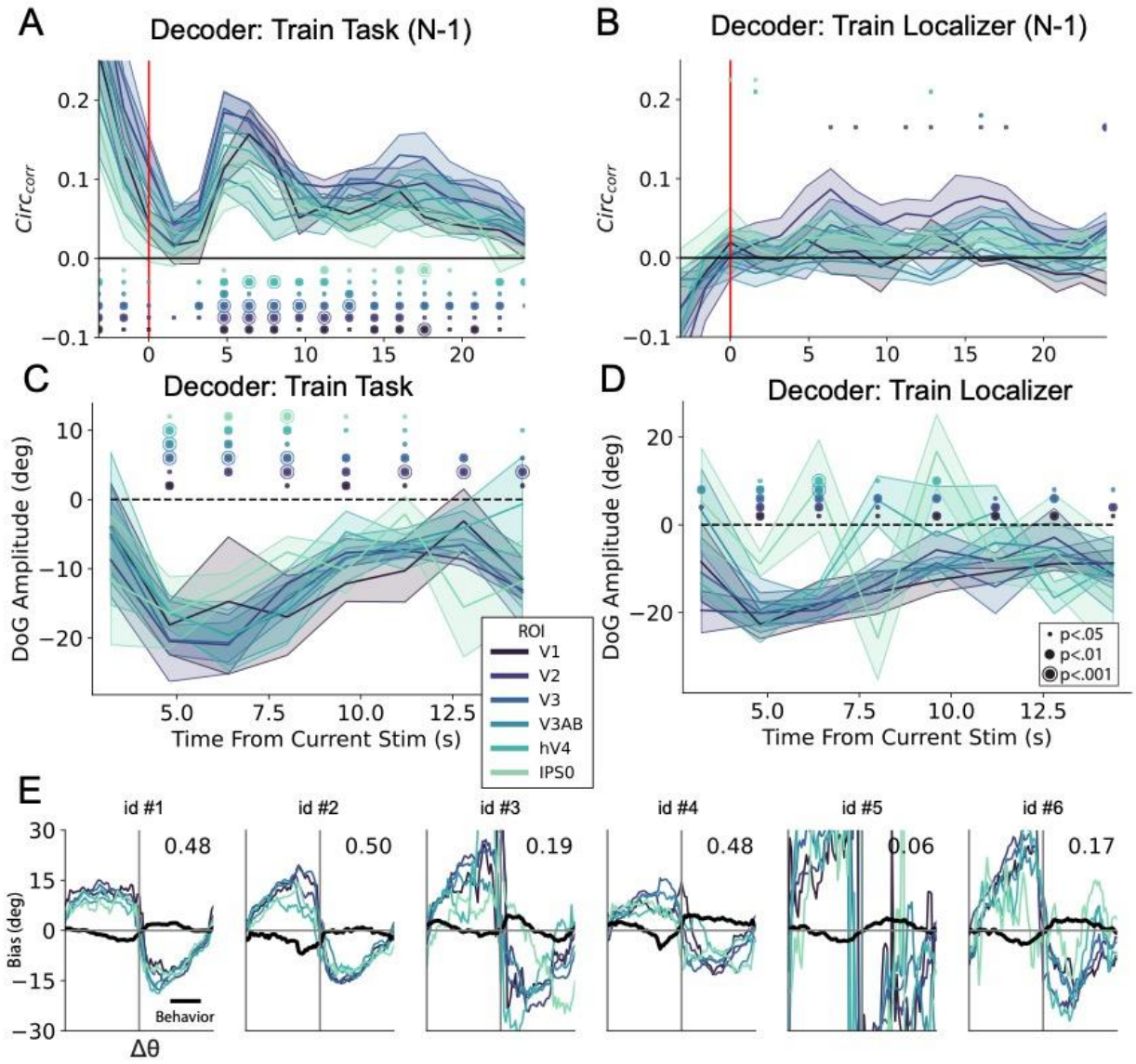


Figure 1-9 Impact of previous trial across time and individuals.

A: Decoding of the previous stimulus dropped to chance around stimulus presentation before rebounding B: decoding using sensory localizer data was consistently at chance during N+1 trial suggesting information relating to past stimulus is not stored in a sensory code. C-D: Decoded biases across time for both decoders are consistently repulsive. E: Bias curves for individual participants using the memory decoder across rois (see legend) overlayed with behavioral biases (black). Neural and behavioral biases are consistently in opposite directions. Note that id#3 exhibits peripheral repulsion.

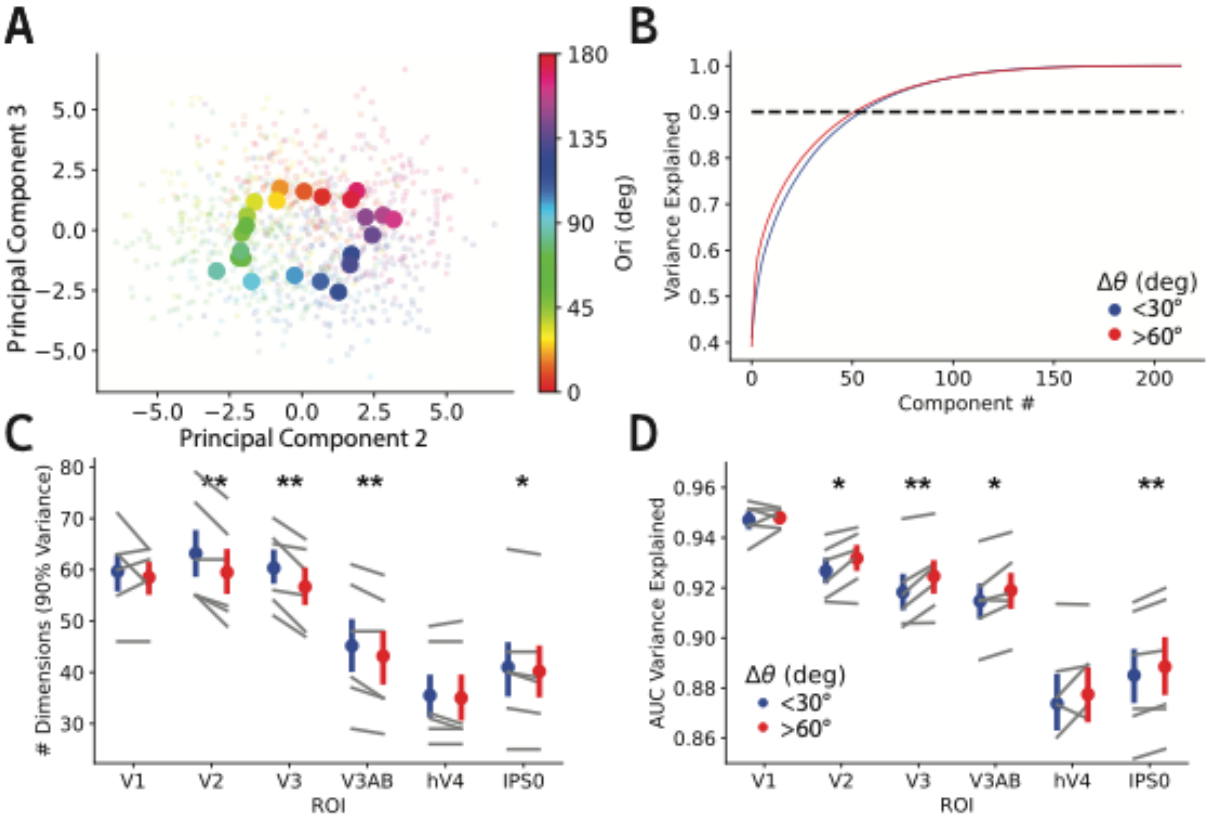


Figure 1-10 Dimensionality Analysis.

To quantify the intrinsic dimensionality of neural representations and whether it changes following a 'close' stimulus, we performed principal component analyses (PCA) on the activity matrix (number of trials x number of voxels) of responses across different ROIs. A: we found that early principal components were correlated with the presented orientation, here presenting both individual trials as well as the average location for different orientation bins (large solid circles) for an example subject and ROI. B: we performed PCA separately for trials following 'close' and 'far' trials, being careful to subsample the number of trials in the larger group. We then sorted the eigenvalues and examined the proportion of variance explained as a function of the number of components included separately for each group. C: we found that it took significantly more components to explain 90% of the variance on the population activity following close versus far stimuli. This suggests that the representations in most visual areas occupy a higher dimensional space following close stimuli, but curiously not V1. Note that the total number of dimensions is shaped by the number of voxels included, so differences between subjects/ROIs should not be interpreted with how these data were processed. D: we additionally looked at the area under the variance curve to avoid any arbitrary effects of choosing 90% and found a similar effect (higher AUC implies lower dimensionality).

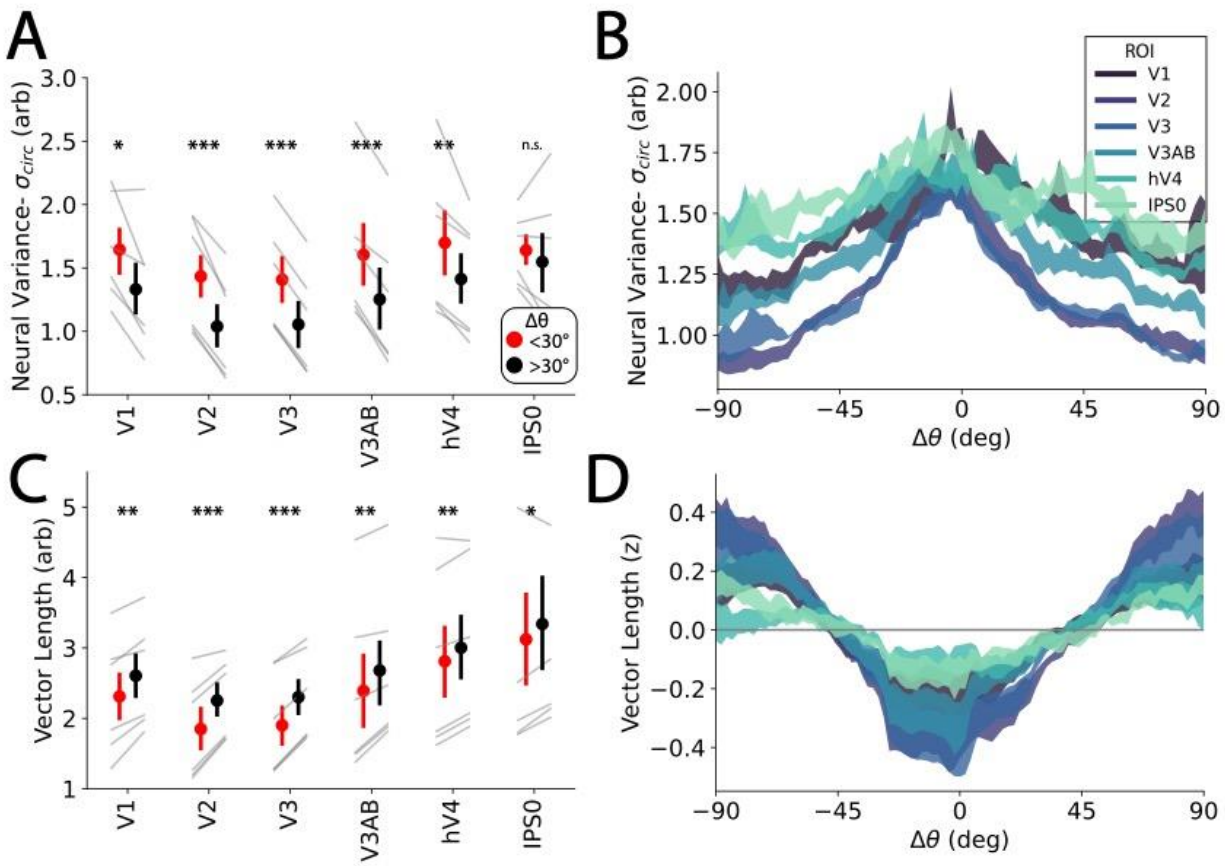


Figure 1-11 Decoded uncertainty as a function of Δθ across ROIs.

A: σ_{circ} of decoding errors is significantly greater for close (<30°) versus far (>30°) stimuli across early visual ROIs (see Neural Variance). Points and error bars are mean ±SEM across participants; gray lines depict individual participants. Error bars depict SEM across participants B: Sliding σ_{circ} for V1-V3 shows a monotonic relationship (± SEM across participants). C-D: Same as A-B but measuring uncertainty directly measured from the single trial posterior (see eq. 8). Results are qualitatively very similar for both techniques. *, p<.05, **, p<.01, ***, p<.001.

Figure 1-12 Trial Simulation.

To better understand how our experiment's trial sequence could impact results, we simulated BOLD signals based on our empirically estimated HRFs and our trial sequences used in the task.

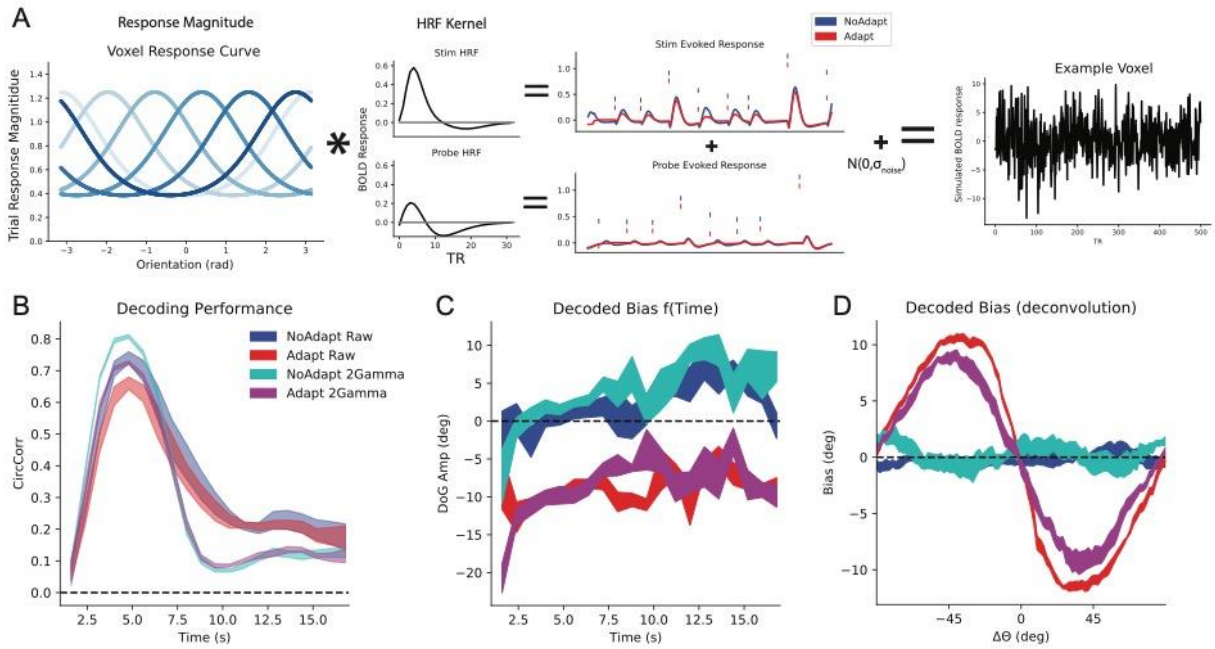
We first created a population of 32 voxels with uniformly distributed von Mises tuning curves. Note that for the purposes of this simulation, we are effectively treating voxels as neurons instead of a summation of the metabolic demands of many neurons. This shortcut comes from experience simulating voxel activity and finding decoding results are unaffected by such a shortcut while making results a bit simpler to understand (and faster to generate). The responses of each voxel were estimated by first generating a design vector based on the stimulus presentation times of both the stimulus and probe for a given subject with the amplitude of the response based on the defined tuning curves. This vector was then convolved with an empirically estimated HRF (both the raw output and when parameterized with a double gamma function) randomly sampled from voxels of the same subject to get the estimated evoked response to both the stimulus and the probe. These two signals were then combined along with gaussian noise to simulate the voxel response (A).

Importantly, the tuning properties of these simulated voxels were unaffected by past stimuli so any biases found by applying our decoding techniques could reflect artifacts of our task design or analysis procedure. We additionally simulated BOLD responses with true adaptation in the underlying neural tuning. For simplicity we simply attenuated the response to the current trial by 40% of the response to the previous trial while keeping all other stages of our analysis the same.

We first applied a decoder across time to the epoched data and found a similar pattern to our empirical data with decoding performance following a parabolic shape before leveling off at some intermediate level, here utilizing HRFs from V3 voxels (B). This was true whether we used parameterized or raw HRFs and whether the simulation included adaptation. We next examined biases in our decoder as a function of stimulus history. With adaptation (red curves), decoded representation were systematically repelled from previous stimuli matching our empirical findings (C). Importantly, without adaptation the resulting bias was never repelled from the previous stimulus (blue curves). This suggests that the timing of our stimuli and the resulting evoked responses should not bias us towards seeing the repulsive results we report.

We finally implemented the regression-based estimation of BOLD responses as we did with our empirical data. As stated before, this technique should remove any linear contributions of past evoked responses to our estimate of the current trial's response. When analyzing the resulting biases, we found that while the unadapted data showed no bias from the previous stimulus (as expected, despite added noise) the adapted response continued to show a repulsive bias (D). Data and code supporting this figure found here:

This analysis demonstrates that 1) while our task design could lead to biases in decoded representations in the absence of any neural history effects, these effects tend to be in the opposite direction of our reported effects and 2) our use of HRF kernels to estimate trial responses is unbiased by across trial contamination and robustly recovers repulsive patterns in the presence of real neuronal adaptation at noise levels similar to our study.


A
Response Magnitude
Voxel Response Curve
Trial Response Magnitidue
Orientation (rad)
HRF Kernel
Stim HRF
Probe HRF
BOLD Response
TR
Stim Evoked Response
Probe Evoked Response
NoAdapt
Adapt
N(0,σnoise)
Example Voxel
Simulated BOLD response
B
Decoding Performance
NoAdapt Raw
Adapt Raw
NoAdapt 2Gamma
Adapt 2Gamma
CircCorr
Time (s)
C
Decoded Bias f(Time)
DoG Amp (deg)
Time (s)
D
Decoded Bias (deconvolution)
Bias (deg)
Δθ (deg)

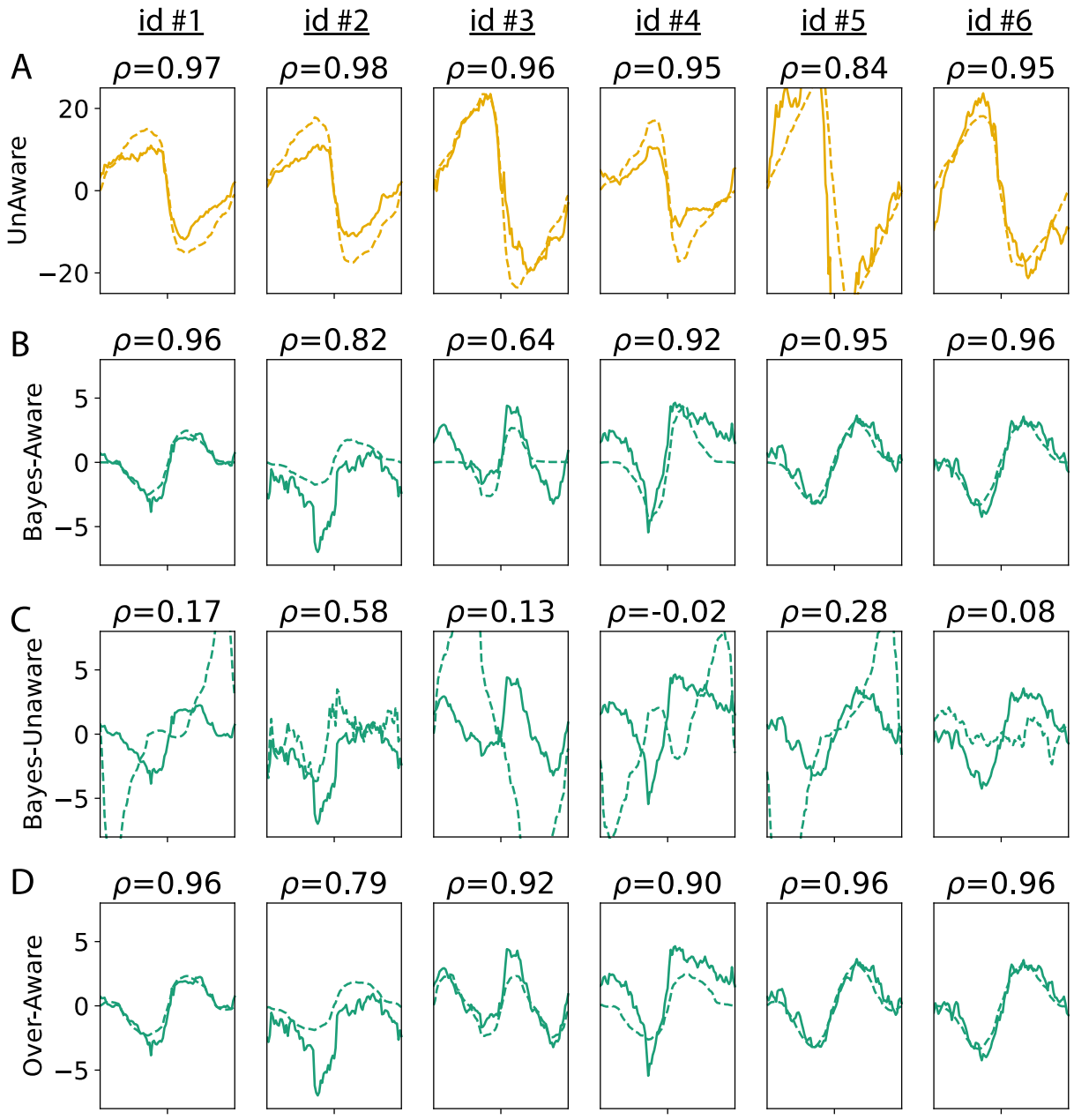


Figure 1-13 Model fits for individual participants (same order as Fig 3).

Solid lines correspond to empirical neural (yellow) or behavioral (green) bias; dashed lines correspond to model fits to BOLD decoding bias (Unaware model, A) or behavior (B-D). Model fits plotted are average of noiseless biases generated by models fit to each CV fold. Note that models are fit to raw data, not binned data presented here. Pearson's correlations are reported above each fit between binned and model estimated bias

Table 1 Fit Parameters

Cells correspond to parameters for proposed decoder. Items with bold values indicate free parameters adjusted to fit empirical data (± SEM across participants). γ_m controls the amplitude and γ_s controls the width of gain adaptation (Fig 4A). These parameters were fit by minimizing the residual sum of squared errors between the unaware decoder and the BOLD decoder output. γ_{m2} and γ_{s2} are the assumed adaptation parameters at decoding. These terms were either set to assume no adaptation (unaware), match the true amount of adaptation (aware) or are free parameters adjusted to maximize the likelihood of responses (over-aware, Fig 4B). Last, R adjusts the average Poisson firing rate and ψ controls the variance of the prior distribution (Fig 4C). These parameters are adjusted for decoders using a Bayesian prior while R is set to the arbitrary value of 5 for non-Bayesian decoders (it has no effect on bias for non-Bayesian decoders). Increasing R increases the precision of the likelihood function and reduces the relative influence of the prior. Increasing ψ increases the range of Δθ over which the prior has an influence. Data and code supporting this table found here: https://osf.io/e5xw8/?view_only=e7c1da85aa684cc8830aec8d74afdcb4

	Fit To:	BOLD Decoder	Behavior		
Stage:		**Unaware**	**unaware Bayes (Prior*unaware)**	**Aware Bayes (Prior*Aware)**	**Over Aware**
Encoding	γ_m	**0.80 ±0.03**	γ_m	γ_m	γ_m
	γ_s	**0.38 ±0.08**	γ_s	γ_s	γ_s
Decoding	γ_{m2}	0	0	γ_m	**0.66 ±0.07**
	γ_{s2}	1	1	γ_s	**0.53 ±0.07**
Bayes	R	5	**2.39 ±0.76**	**0.21 ±0.12**	5
	ψ	N/A	**0.57 ±0.04**	**1.96 ±0.41**	N/A

Chapter 2 Distinguishing response from stimulus driven history biases

Abstract

Perception is shaped by past experience, both cumulative and contextual. Serial dependence reflects a contextual attractive bias to perceive or report the world as more stable than it truly is. As serial dependence has often been examined in continuous report or change detection tasks, it unclear whether attraction is towards the identity of the previous stimulus feature, or rather to the *response* made to indicate the *perceived* stimulus value on the previous trial. The physical and reported identities can be highly correlated depending on properties of the stimulus and task design. However, they are distinct values and dissociating them is important because it can reveal information about the role of sensory and non-sensory contributions to attractive biases. These alternative possibilities can be challenging to disentangle because 1) stimulus values and responses are typically strongly correlated and 2) measuring response biases using standard techniques can be confounded by *context-independent* biases such as *cardinal bias* for orientation (i.e. higher precision, but repelled, responses from vertical and horizontal orientations). Here we explore the issues and confounds related to measuring response biases using simulations. Under a range of conditions, we find that response-induced biases can be reliably distinguished from stimulus-induced biases and from confounds introduced by *context-independent* biases. We then applied these approaches to a delayed report dataset (N=18) and found evidence for response over a stimulus driven history bias. This work demonstrates that stimulus and response driven history biases can be reliably dissociated and provides code to implement these analysis procedures.

Introduction

Perceptual reports can be shaped by past stimuli and actions - the visual system exploits this information to support efficient information processing. To this end, the visual system expends less energy processing expected stimuli and can rely on priors to facilitate processing of new sensory information (Oliver, 1952; Mumford, 1994; Olshausen and Field, 1996). However, even though these adaptive mechanisms support more efficient processing on average, they also lead to a collection of perceptual biases.

For example, over developmental or evolutionary time scales perceptual processing has adapted to represent frequently encountered stimulus features such as vertical and horizontal orientations with greater precision than off-cardinal oblique orientations (the *oblique effect*). While this resource allocation supports more efficient processing in early visual cortex, it also gives rise to a phenomenon of *cardinal bias* where perceptual reports are repelled from vertical and horizontal orientations (Girshick et al., 2011; Wei and Stocker, 2015). Importantly, *cardinal bias*, as well as the *oblique effect*, are thought to be based on long-term exposure to natural image statistics and are highly stable across time (Henderson and Serences, 2021). Hence, we use the term *context-independent biases* to refer to this and related phenomena.

In addition to these *context-independent* biases, dynamic perceptual biases can also emerge based on exposure to recent stimulus features. For instance, viewing a stable image feature for an extended period can lead to a suppressed neural response to that feature (Dragoi et al., 2000; Kohn and Movshon, 2004; Patterson et al., 2013). Given that stimuli are generally stable across time, these *adaptation* effects are also thought to contribute to *efficient coding* as fewer neural resources (i.e., spikes) are dedicated to processing expected stimulus features (Barlow, 1961; Felsen et al., 2005; Benucci et al., 2013). However, attenuated responses in

neurons tuned to the viewed stimulus can bias neural population response profiles away from the adapting stimulus. This neural repulsion is the likely source of perceptual repulsion effects seen in well-known phenomena such as the waterfall illusion or the tilt after-effect (Anstis et al., 1998; He and MacLeod, 2001).

Interestingly, and in contrast to typical adaptation-induced repulsive biases, the repetition of similar stimuli can sometimes lead to an attractive or assimilative bias known as hysteresis or serial dependence (Corbett et al., 2011; Chopin and Mamassian, 2012; Cicchini et al., 2014; Fischer and Whitney, 2014). Typically, attractive serial dependence emerges with briefly presented or near-threshold stimuli that are hard to perceive, as opposed to longer exposure to high contrast stimuli that usually leads to adaptation and perceptual repulsion (Chopin and Mamassian, 2012; Maus et al., 2013; Cicchini et al., 2017; Fritsche et al., 2017). These attractive biases can be explained by invoking a Bayesian prior for stimulus stability over short time scales (Cicchini and Burr, 2018; Pascucci et al., 2019; van Bergen and Jehee, 2019; Fritsche et al., 2020). Given this prior for environmental stability, the precision of near-threshold stimuli can be improved by biasing reports towards recently viewed features (Cicchini and Burr, 2018; Fritsche et al., 2020; Sheehan and Serences, 2022). However, even though attractive biases are observed across a host of stimulus/task domains, their ultimate source is still debated.

Here we address a set of key unanswered questions related to efficient information processing in the human visual system. First, do attractive serial dependence effects depend on the physical identity of recently seen features, or on the responses made to report the identity of recently seen stimuli? Second, how do attractive serial dependence effects interact with adaptation and *context-independent* factors like *cardinal bias*? Parceling out sensory and motor

contributions from these other perceptual biases is critical to better understanding the source of the effect because these factors all jointly contribute to measured perceptual reports.

Disentangling sensory from motor/decisional contributions to attractive serial biases is particularly challenging because most studies of serial dependence have employed delayed recall paradigms where responses are highly correlated with the presented stimulus feature. For example, in a typical task a participant is instructed to report the orientation of a remembered orientation using a mouse pointer. Their response will ultimately be driven by the integration of sensory evidence on that trial, adaptation induced by previous stimuli, *context-independent* biases (e.g., *cardinal bias*), and random errors accumulating from other unmeasured sources. These will cause the response to deviate from the stimulus orientation but only by a few degrees such that even for a low performing participant, stimulus identity and the associated responses will still be highly correlated (r_{circ}=0.63, σ=21.4° for an example continuous report dataset which we analyze in more detail below).

Most studies of serial dependance have focused only on the influence of the previous stimulus and claim that it is the processing or perception of the physical stimulus that induces attractive biases (Fischer and Whitney, 2014; Cicchini et al., 2017; Cicchini and Burr, 2018; Manassi et al., 2018). However, the emerging consensus is not so straightforward. One recent study found evidence that responses are simultaneously repelled (due to adaptation) and attracted (due to the application of Bayesian priors) to past stimuli but at different timescales, leading to both attractive and repulsive effects (Fritsche et al., 2020). In contrast, other work suggests that it is the previous decision, not the stimulus per se, that leads to attractive serial biases (Pascucci et al., 2019).This finding is consistent with subsequent studies that have simultaneously modeled the influence of both the previous response *and* the previous stimulus and found that reports are

simultaneously attracted to previous responses and repelled from previous stimuli, providing an extra layer of distinction between the attractive and repulsive effects described by Fritsche and colleges (2020) (Sadil et al., 2021; Moon and Kwon, 2022).

Trying to ascribe biases to past responses is further complicated by *context-independent* biases (e.g., *cardinal bias*) (Fritsche, 2016). When sorting trials as a function of the previous response ($resp_{N-1}$), the sorting variable ($\Delta R = resp_{N-1} - stim_N$) is dependent on the physical stimulus feature ($stim_N$) in the presence of cardinal bias. This is in contrast to analyzing stimulus biases where (for an independent stimulus sequence) the sorting variable is independent of the physical stimulus identity $\Delta s = (stim_{N-1} - stim_N) \perp stim_N$. As a result, any *context-independent* bias, such as repulsion from the cardinal axes, can lead to a dependence of $resp_N$ on ΔR. This dependance may be why past studies have shown a spurious attraction to future or shuffled trial sequences – an observation that lacks a reasonable causal explanation (Pascucci et al., 2019). Thus, observing a spurious response bias to future or shuffled sequences raises the concern that *any* measured response bias (e.g., even towards the previous trial, ΔR_{N-1}) could also be influenced by the same artifact. In Pascucci et al. (2019) and other studies that followed, this issue was addressed by subtracting the average *context-independent* bias from either participant responses or response errors. This method of correction is reasonable, but may actually be insufficient given other *context-independent* anisotropies (e.g., the *oblique effect*) as noted by others (Fritsche, 2016). Thus, to reconcile these seemingly paradoxical findings, an analytic framework is needed to successfully disentangle the relative contribution of perceptual, motor, adaptation, and *context-independent* factors.

To address these concerns, we created a model observer exhibiting either stimulus or response driven biases from the previous trial. For parsimony, we will only explore orientation

stimuli that feature *cardinal biases* along with the *oblique effect* in this study, but our approach should generalize to other stimulus types (e.g., spatial location, numerosity, pitch). We found that some techniques can reliably distinguish between stimulus and response biases across a range of conditions, but that care needs to be taken to correct for *context-independent* biases. We additionally apply these techniques to an orientation working memory dataset and demonstrate that the history biases observed are primarily attributable to past responses, not to the physical stimulus features. All data and code to implement and expand on these simulations, including power analyses and our analyses of an empirical dataset are available at: https://github.com/TimCSheehan/historyResponseModeling.

Methods

Generative Model

To better understand how different sources of bias will ultimately shape behavioral responses, we built a model designed to mimic response properties of human observers. First, we generated an independent and identically distributed (IID) stimulus sequence that uniformly sampled a circular 0-180° feature space (e.g. orientation space). When the sequence is encoded, Von Mises distributed perceptual variability is introduced such that the probability of perceiving a stimulus is governed by the following distribution:

$$p_{encoding}(m|\mu,k) = \frac{exp(k\ cos(m-\mu))}{2\pi I_0(k)} \quad [24]$$

where k and μ are the precision and center of the von Mises distribution respectively, and m is the encoded orientation. $I_0(k)$ is the Bessel function of the first kind of order 0. We utilize two

types of encoding processes. The "biased encoder" features both the *oblique effect*, such that precision is higher around vertical and horizontal stimuli

$$\kappa_{oblique} = \kappa_{base}(1 + cos^2(2\theta)) \quad [25]$$

where θ is the stimulus orientation spanning [0, π] and *cardinal bias* such that responses are biased away from the cardinal orientations

$$\mu_{cardinal} = \theta + A \cdot sin(4\theta) \quad [26]$$

where A=10 is the amplitude of the bias (see Figure 1, *Cardinal Bias* for a depiction of both functions). Note that both $\kappa_{oblique}$ and $\mu_{cardinal}$ have two peaks/cycle as the cosine function is squared for the *oblique effect*. The second encoding model, termed the "uniform encoder", has constant precision across feature space ($\kappa_{uniform}=1.5\cdot\kappa_{base}$, equalizing average precision) and is centered on the true stimulus value ($\mu = \theta$).

On each trial, a random draw from the probability distribution $p_{encoding}$ is used to generate a point stimulus estimate m_n which is then used as μ in either the biased or the uniform encoding model. This μ parameter, along with the concentration parameter k of the von Mises distribution, generates a probability distribution function (PDF) that defines the stimulus likelihood function[1]. This likelihood is then multiplied by a Bayesian prior centered on either the previous stimulus ("stimulus bias") or the previous response ("response bias", Figure 2-1, *Bayesian Inference*). This prior is based on measurements of natural videos and is a mixture of a von Mises and a uniform distribution to account for both stable random changes across time (Felsen et al., 2005;

[1]Note that here for simplicity we are equating the shape of the likelihood function, p(θ|m), with the posterior p(m|θ).

van Bergen and Jehee, 2019). The relative influence of stable and random changes is controlled by the parameter p_{stable} such that

$$p_{prior}(m|\mu, k, p_{stable}) = p_{stable}\frac{exp(k\,cos(m-\mu))}{2\pi I_0(k)} + (1-p_{stable})\frac{1}{2\pi} \quad [27]$$

where μ is the stimulus or response on the previous trial and κ is constant (building on previous findings suggesting uncertainty on the previous trial does not appear to shape serial dependence in a Bayesian manner (Fritsche, 2016; Ceylan et al., 2021; Gallagher and Benton, 2022). The maximum value of the resulting posterior

$$resp_n = argmax_m\,(p_{prior} \cdot p_{encoding}) \quad [28]$$

is taken as the Bayes optimal single trial estimate of the stimulus (Figure 2-1, *Bayesian Inference*, sold line). We equate the output of the model with the "perceived" stimulus value that the participant would indicate with a behavioral response.

Behavioral Analysis

Independent Bias Parameterization

To analyze the results from these different encoding and decoding processes, we sorted response errors as a function of the previous stimulus ($\Delta S = stim_{N-1} - stim_N$) or as a function of the previous response ($\Delta R = resp_{N-1} - stim_N$). We visualized the resulting bias for each participant by taking a sliding circular mean of the errors as a function of ΔS or ΔR. To simulate typical trial counts of a psychophysics experiment, we ran experiments of 30 participants completing 360

trials each. The magnitudes of history biases were estimated by fitting a derivative of von Mises (DoVM) function:

$$doVM(x;\ a, w) = a\ wsin(x)\ exp(w\ cos(x))\ /\ (zI_0(w)) \quad [29]$$

with amplitude, *a*, and width, *w* *(Sadil et al., 2021).* These parameters were fit to minimize the RSS errors when *x* corresponds to either ΔS or ΔR. *z* is a normalizing constant such that the amplitude, *a*, corresponds to the height of the resulting function. We additionally performed all analyses using the more commonly utilized derivative of Gaussian function and found similar results, but prefer the DoVM function as it is continuous at ±π.

Long-term Bias Correction

Previous studies have attempted to account for any confounds introduced by *context-independent* biases by subtracting out the average bias from either the responses ($resp_N$) or the errors ($resp_N$ - $stim_N$) (Fritsche, 2016; Pascucci et al., 2019; Sadil et al., 2021; Moon and Kwon, 2022). We perform this correction by first fitting an n=6 parameter Fourier-like decomposition

$$f(\theta;\ a_1, \ldots, a_N)\ = \sum_{n=1}^{N} g(\theta; n, a_n) \quad [30]$$

$$g(\theta; n, a_n) = \begin{cases} sin(\theta n); & n \equiv even \\ cos\big(\theta(n+1)\big); & otherwise \end{cases} \quad [31]$$

to subjects errors as a function of $stim_N$ and subtracting this function from either the responses (*response correction*: $resp_{residual} = wrap(resp_N - f(stim_N))$ or from the resulting errors (*error correction*: $E_{residual} = wrap(resp_N - f(stim_N) - stim_N)$). Note that correcting responses additionally

influences the errors as they are calculated using the modified responses. When analyzing response biases, both corrections impact errors (y-axis) (as correcting responses also corrects errors) while response correction additionally impacts sorting of trials (x-axis). While these two forms of correction ultimately yield similar results, it is important to consider how response correction procedures change the interpretation of any resulting bias (see Discussion).

One concern that arises with analyzing response biases, and a primary motivation for this study, is the presence of 'spurious serial dependence' whereby sorting responses as a function of ΔR can give the appearance of attractive biases to the N+1 stimulus or after shuffling the stimulus sequence (Pascucci et al., 2019). As we do not expect the response on a future or random trial to influence our error on the current trial, the presence of such a bias is concerning and may suggest a bias measured relative to past/future stimuli is an artifact of the analysis procedure. To better understand this phenomenon, we additionally consider our errors relative to both the N+1 stimulus and relative to the N-1 stimulus of a shuffled trial sequence.

Joint Bias Parameterization

Recent studies have simultaneously modeled the impact of the previous stimulus and previous response (Sadil et al., 2021; Moon and Kwon, 2022). We implemented this by parameterizing two DoVM functions modulated by ΔS and ΔR and optimized to minimize the residual SSEs. Specifically, we have two vectors ΔS and ΔR which are inputs to two DoVM functions. The resulting minimization function is

$$min \sum_{\forall i} (E_i - DoVM(\Delta S;\ a_s, w_s) - DoVM(\Delta R;\ a_R, w_R))^2 \quad [32]$$

where E_i corresponds to the actual error, $wrap(resp_i - stim_i)$, on the *ith* trial.

Statistical Analyses

When bias curves are visualized, we include the results of one-sample and paired two-tailed t-tests without correction of the amplitudes of fit DoVM functions.

Power Analysis

We performed power analyses to estimate the probability of detecting a significant effect ($\alpha<.001$) for an experiment conducted with n=30 participants and defined effect sizes and trial counts. For a given experiment, we present the probability of rejecting the null hypothesis that stimulus or response biases are significantly greater and in addition that the magnitudes of the two effects are different from one another.

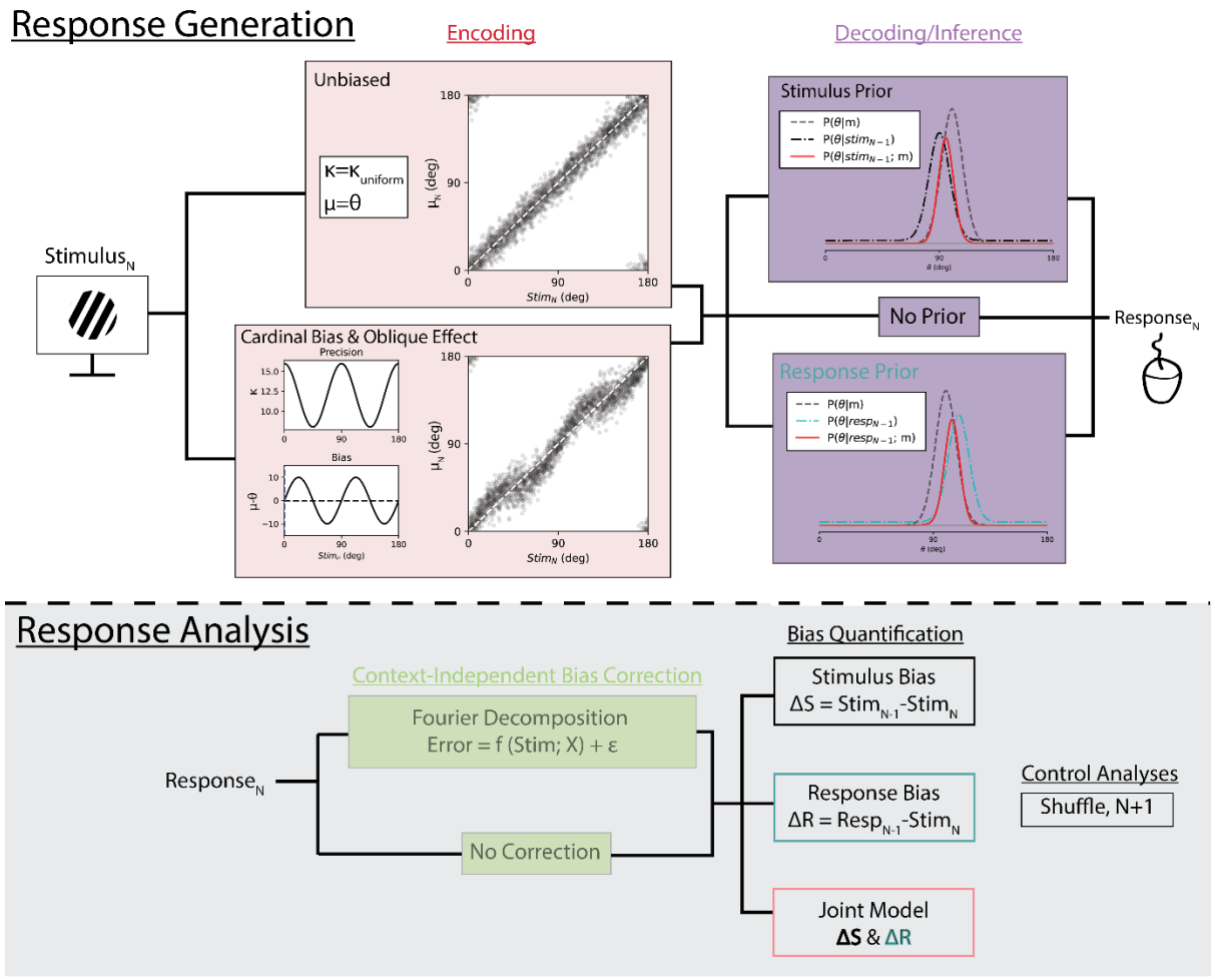


Figure 2-1 Simulated observer model.

Response Generation, on each trial a stimulus is encoded by a biased or unbiased encoder. The encoded representation is interpreted at the inference stage by introducing either a stimulus, response, or no prior for stability. The output from this stage is the response we analyze and used to bias future reports. **Response Analysis**, responses are first corrected (or not) for context-independent biases by fitting a Fourier-like function. We then analyze errors as a function of either the previous stimulus, response, or both. We perform additional control analyses by shuffling trial order or examining the influence of future responses.

Additional controls

Most experimentalists interested in studying serial dependence intentionally utilize stimulus sequences with a roughly uniform distribution of trial-by-trial stimulus transitions (e.g., $P(\Delta S)$ is uniform). For a variety of factors including inadequate randomization due to low trial counts or the introduction of intentional structure into the distribution, this assumption is often violated to varying degrees (He et al., 2010; Chopin and Mamassian, 2012; Maus et al., 2013).

To determine how non-uniform stimulus sequences affect measurements of serial dependence, we additionally simulated an analysis pipeline using sequences that feature positive (+) and negative (-) autocorrelations.

The fundamental concern that motivates including simulations with autocorrelated stimulus sequences is that studies attempting to reveal attractive biases to past stimuli or responses may instead only reveal artifacts of their analysis techniques where no biases are present. To assess these concerns, we additionally generate responses where neither stimulus or response serial dependence were implemented to provide a ground-truth case where no biases should be observed (see Figure 2-1, decoding).

To account for the possibility of a repulsive bias from the stimulus itself, for some experiments we inserted an additive DoVM repulsive bias centered on the previous stimulus with width 1 and variable amplitude.

Psychophysical Study

18 participants completed between 192 and 488 (380 ± 15.2, mean ± SEM) trials of a delayed orientation report task. All participants provided informed consent, had normal or corrected to normal vision, and were compensated either in course credit or at a rate of $10/hour. Participants were instructed to fixate on a black fixation cue that was present at the center of the screen 0.5° (degrees of visual angle) and was visible throughout the entire experiment. The trial began with a 1500 ms ITI featuring only the fixation point. Then, two foveally presented oriented gratings subtending 1.5 to 23° degrees of visual angle were presented in succession separated by a 1000 ms inter-stimulus-interval (ISI). Each stimulus had a randomly oriented grating (2 cycles/°, 0.8 Michelson contrast) that was smoothed by a 2D Gaussian kernel with σ=0.5°. Each stimulus was presented for 1s and reversed phase every 125 ms. Each stimulus was

followed by a 250 ms filtered noise mask [f_{low}=0.25, f_{high}=1.0 cycles/°] that changed once after 125 ms. After the second item, a retro cue (the numbers '**1**' or '**2**') indicated the target most likely to be probed (80% validity). On 1/6th of trials a neutral ('**X**') was presented in lieu of a retro cue (both items equally likely to be probed). The retro cue was followed by a blank delay period 2500 ms. Participants then controlled a black response dial (using the "ASDF" buttons on a standard QWERTY keyboard) and they were given between 500 and 5000 ms to match the orientation of the probed stimulus. After pressing the space bar to confirm their response or timing out, the dial disappeared, and feedback was provided for 2000 ms by displaying the unsigned error in degrees and turning the response dial green if participants were closer than 10° and red otherwise.

Results

Serial Dependence Without Cardinal Bias

We first analyzed responses in a model without *context-independent* biases featuring either stimulus serial dependence, response serial dependence, or no serial dependence (columns *Left, Center, Right* respectively, Figure 2-2). For this simulation, and unless otherwise noted, we use κ_{base} = 8 and therefore $\kappa_{uniform}$ = 12. The first row shows biases relative to the previous stimulus and reveals that trials with true stimulus bias (Figure 2-2A) show a larger stimulus (ΔS, black curve) relative to response (ΔR, teal curve) bias. We additionally observed a larger response bias when the underlying source of the bias is towards the previous response (Figure 2-2B). Together, this suggests that, in the absence of *context-independent* biases, the relative magnitudes of stimulus/response serial dependence is a good proxy for the dominant source of the bias.

Critically, the only artifactual bias occurs when examining ΔR_{N+1} when there was a genuine bias response bias (Figure 2-2E). This demonstrates that cardinal or other history independent biases are not necessary to observe artifacts in analyzing response biases in the presence of true response dependence and suggests that such an artifact is an indicator of a bona fide bias in the data. We explore why this N+1 artifact arises in the next section.

The N+1 response bias artifact

Ensuring that there is no bias toward future responses (i.e. the *N+1 trial*) has been suggested as a valuable control when evaluating response biases (Pascucci et al., 2019). However, as noted above, we find an attractive bias when sorting trials by ΔR_{N+1} when there is a true response-based serial dependence effect. To understand why this bias occurs, we first identified an important distinction between sorting trials based on the past versus future response. Importantly, $resp_{N-1}$ is independent of $stim_N$ and accordingly $P(\Delta R_{N-1})$ is uniform (Figure 2-3A). However, $resp_{N+1}$ is *not* independent of $stim_N$ because it is influenced by a prior centered on either $stim_N$ or $resp_N$ (depending on the source of the bias) resulting in a highly non-uniform distribution (Figure 2-3A, $P(\Delta R_{N+1})$). To explore why the ΔR_{N+1} spurious bias occurs, we considered two possible outcomes on the current trial, an error *CW* or *CCW* relative to the true stimulus. For the purposes of this visualization, we used the average absolute error of our unbiased observer, $\overline{|E|} = 7.8°$. For observers exhibiting response-based history biases, these *CW/CCW* errors generate distinct priors (Figure 2-3B) that differentially shape future responses. These priors shift $P(\Delta R_{N+1})$ towards the current response (Figure 2-3C). The difference in relative probabilities of the previous response error multiplied by the average response error ($\overline{|E|}$) perfectly captures the measured "spurious" response bias (Figure 2-3D, 2-2E). Thus, spurious biases measured by examining the influence of the *N+1* response are *expected* if the

underlying source of the bias is a prior centered on the preceding response. Because of this, examining the *N+1* influence is not a pragmatic control analysis and researchers should instead opt for a shuffled trial sequence which does not exhibit spurious biases when response biases are present in a dataset.

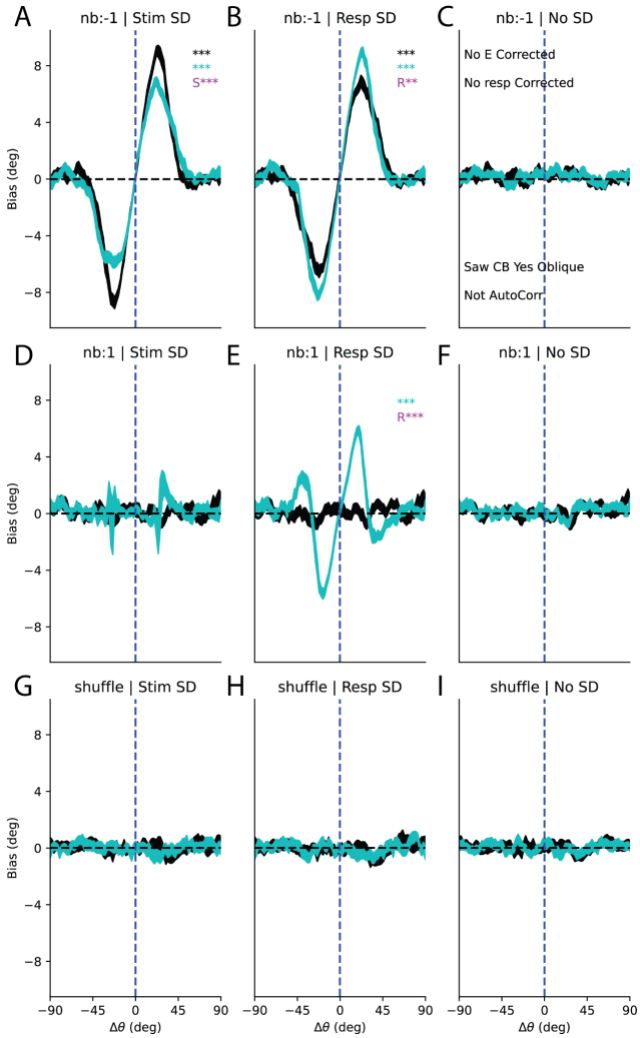


Figure 2-2 Biases of simulated observed without context independent biases.

Stimulus (black) and response (teal) bias curves for all response simulations. (Left, A,D,C) column corresponds to responses generated with an attraction towards past stimuli, (center, B,E,H) column features responses attracted towards past responses, and (right, **C,F,I**) column has no history biases. (Top, **A-C**) row computes Δθ relative to previous trial, (middle, **D-F**) row computes Δθ relative to future trial, and (bottom, **G-I**) row computes Δθ relative to the previous trial after shuffling the stimulus order. Both **A** and **B** show significant attractive biases towards past stimuli and responses with larger attractive biases towards the underlying source of the bias. We additionally observe an attractive bias towards the future response **E** that is an artifact of our sorting procedure. *, p<.05; **; p<.01; ***, p<.01, Bonferroni corrected for 9 stimulus conditions; R, response bias significantly greater than stimulus bias, S, stimulus bias significantly greater than response bias.

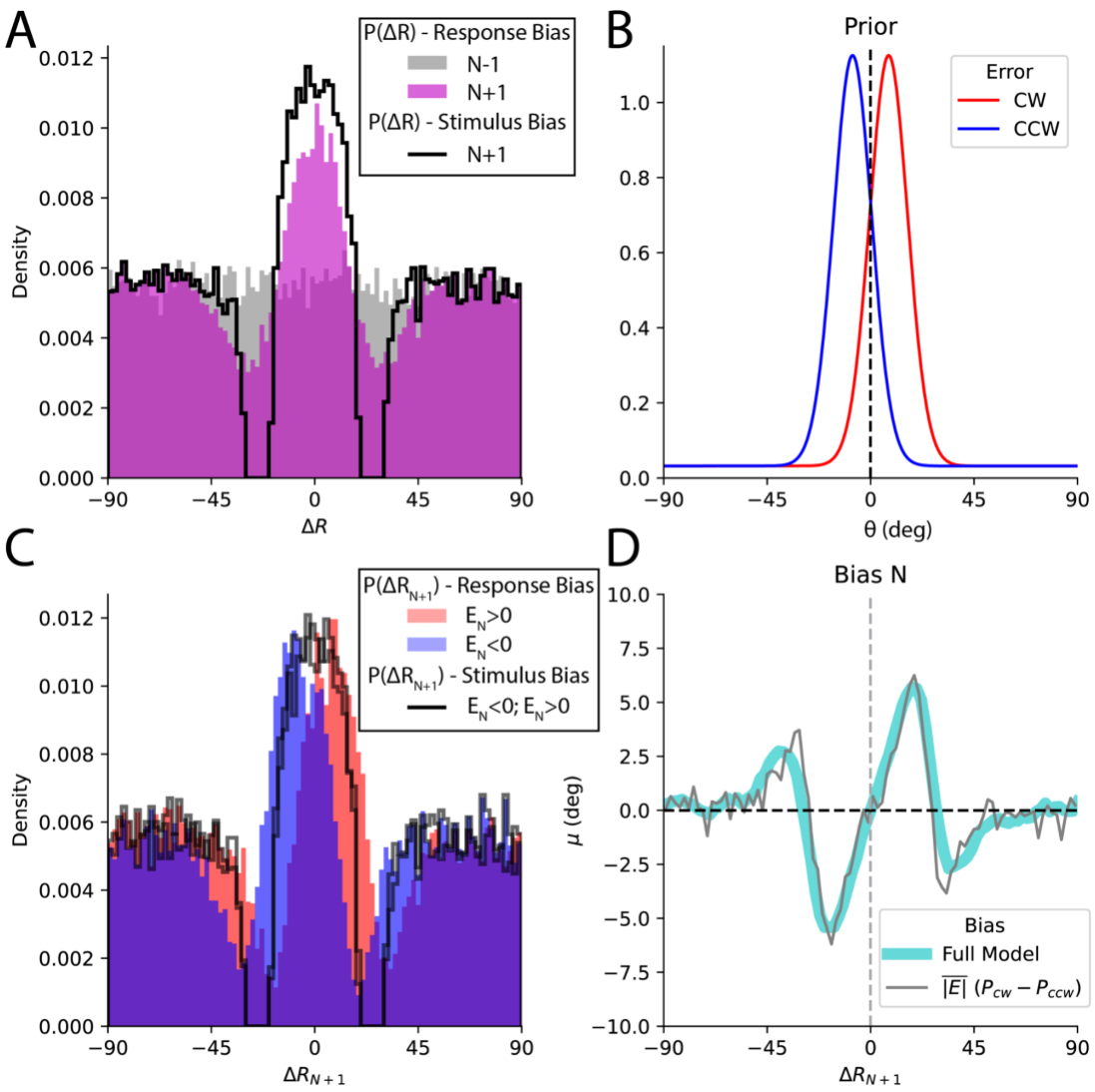


Figure 2-3 The N+1 response bias artifact.

A. P(R_{N-1}), gray, is uniform but P(ΔR_{N+1}), magenta, shows an overrepresentation for small changes. Additionally shown is P(ΔR_{N+1}) for stimulus serial dependance (black trace). **B**. hypothetical priors following a misperception of the average magnitude for our model (7.8°) in the CW or CCW direction. **C**. P(ΔR_{N+1}) on trials with CW or CCW misperceptions are shifted relative to each other. This shifting does not occur when the bias source is the stimulus instead of response (black traces) **D**. The average (unsigned) error multiplied by the difference in the P(ΔR_{N+1}) for CW and CCW responses captures the measured spurious bias.

Serial Dependence with Context Independent Biases

We next analyzed serial dependence after additionally including cardinal bias and the oblique effect at encoding. Both the precision κ and expected value μ were modulated by the

stimulus identity resulting in an encoding process that showed characteristic bias and variance patterns of *cardinal bias* and the *oblique effect* (Figure 2-1). The result of this biased encoding process was then modulated by the same Bayesian prior as used in the previous section. When analyzed, the resulting responses show an increased response bias and a substantial 'spurious' response bias in the absence of any history biases (Figure 2-4A-C) demonstrating that *context-independent* cardinal biases can introduce an artifact as suggested previously (Fritsche, 2016; Pascucci et al., 2019).

This confound is more concerning than the ΔR_{N+1} bias we found in the previous section because an attractive response bias is found even when no underlying serial dependence is present in the generated data (Figure 2-4C) or when trial order is shuffled (Figure 2-8A). Previous studies have tried to address this bias by regressing out the stimulus specific bias from either the errors or the responses. This has generally been achieved by fitting either a higher order polynomial or sinusoidal function to the raw data. For the purposes of this study, we utilized a 6-parameter Fourier like composition of sine/cosine functions of varying frequencies which is more flexible (see eq. 7). Our use of circular functions avoids edge effects found with polynomial fits. We fit this function to errors and subtracted the best-fit function to correct for these biases (Figure 2-4D, *red dotted-line*). This correction substantially reduces any trace of systematic biases (Figure 2-4D, green). We opt to correct errors, but not responses, as this allows ΔR_{N-1} to reflect the relative location of the previous response.

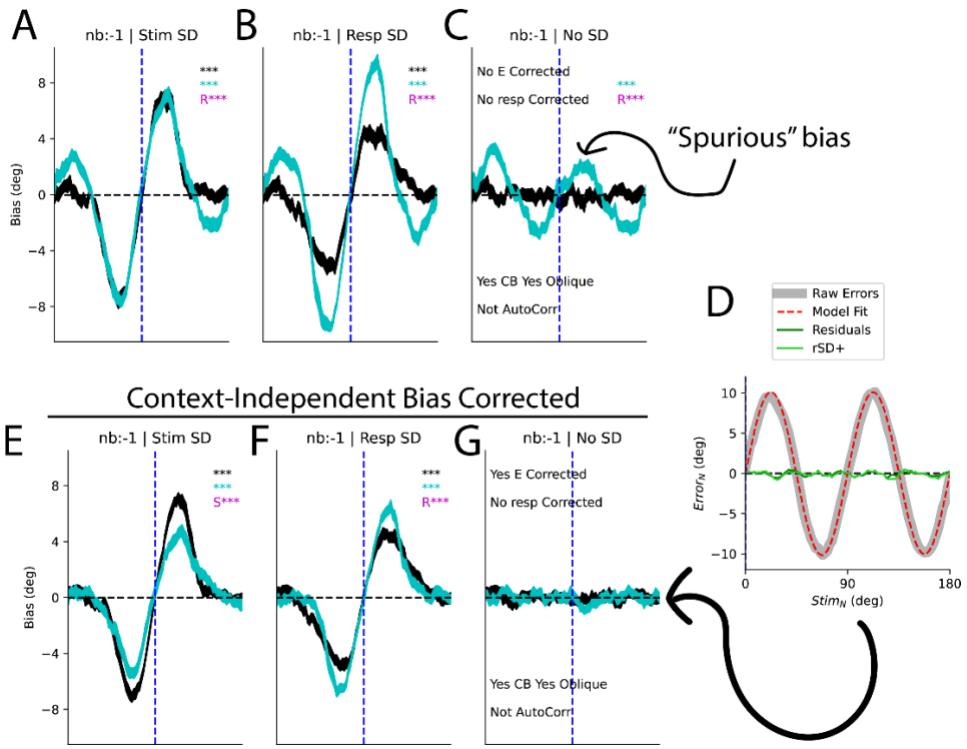


Figure 2-4 Serial dependence in the presence of context independent biases.

A-C. Response/stimulus biases computed using the raw errors results in a spurious response bias (see Fig S1 for all bias curves) **D**. Context-independent biases can be corrected for by fitting a model to responses such that the resulting residuals are not biased as a function of stimulus identity. The light green trace (rSD+) is the residuals when history dependent bias (serial dependance) is present when fitting the history independent bias model. **E-G** Response/stimulus biases computed using the residualized errors.

Correcting for *context-independent* biases in response errors appears to completely remove the presence of spurious biases and returns the relative magnitudes of biases to what is expected given their respective sources (Figure 2-4E-G, See Figure 2-8 for bias curves corresponding to shuffled and N+1 trials). This is critical as this regression-based approach is an effective way to correct for *context-independent* biases and ensure the presence of measured response history biases is not just an artifact. This correction process does nothing to account for differences in variability as a function of the stimulus (the *oblique effect*) but still removes any trace of artifactual responses in the shuffled condition. We separately analyzed the influence of autocorrelations in the sequence of stimuli presented and found no evidence that they introduce new artifacts (Figure 2-9).

Cardinal biases cause spurious response biases

It is not surprising that introducing biased stimulus representations could introduce cofounds. In a general sense, this is because $Error_N$ is dependent on $stim_N$ and furthermore ΔR is no longer independent of the absolute stimulus value. Why this leads to spurious history biases is not particularly intuitive, so we provide a brief demonstration here. First we visualize the joint distribution $P(Stim_N, \Delta R)$ which shows the two variables are clearly not independent (Figure 2-5A). Note that we are not specifying which trial is the inducer (eg. *N-1/ N+1*) as this spurious bias is unchanged even after shuffling trial order. The conditional distributions $P(Stim_N \mid \Delta R)$ for two subsets of ΔR reveal how dramatically $P(Stim_N)$ is interdependent on ΔR (Figure 2-5B). We can then approximate the predicted spurious bias as the dot product of the normalized rows of $P(Stim_N, \Delta R)$ with $\mu_{cardinal}(Stim_N)$ (Figure 2-5C, 2-4A) to get the expected bias

$$Spurious\ Bias\ (\Delta R)\ \approx \sum_{Stim_N} P(Stim_N|\Delta R)\ \mu_{cardinal}(Stim_N)$$

(Figure 2-5D, black). This process captures the "spurious" response bias from the shuffled response distribution (Figure 2-5D, teal). Note that when sorting trials based on the previous stimulus instead of responses, $P(Stim_N|\Delta S)$, is independent and does not give rise to spurious history bias.

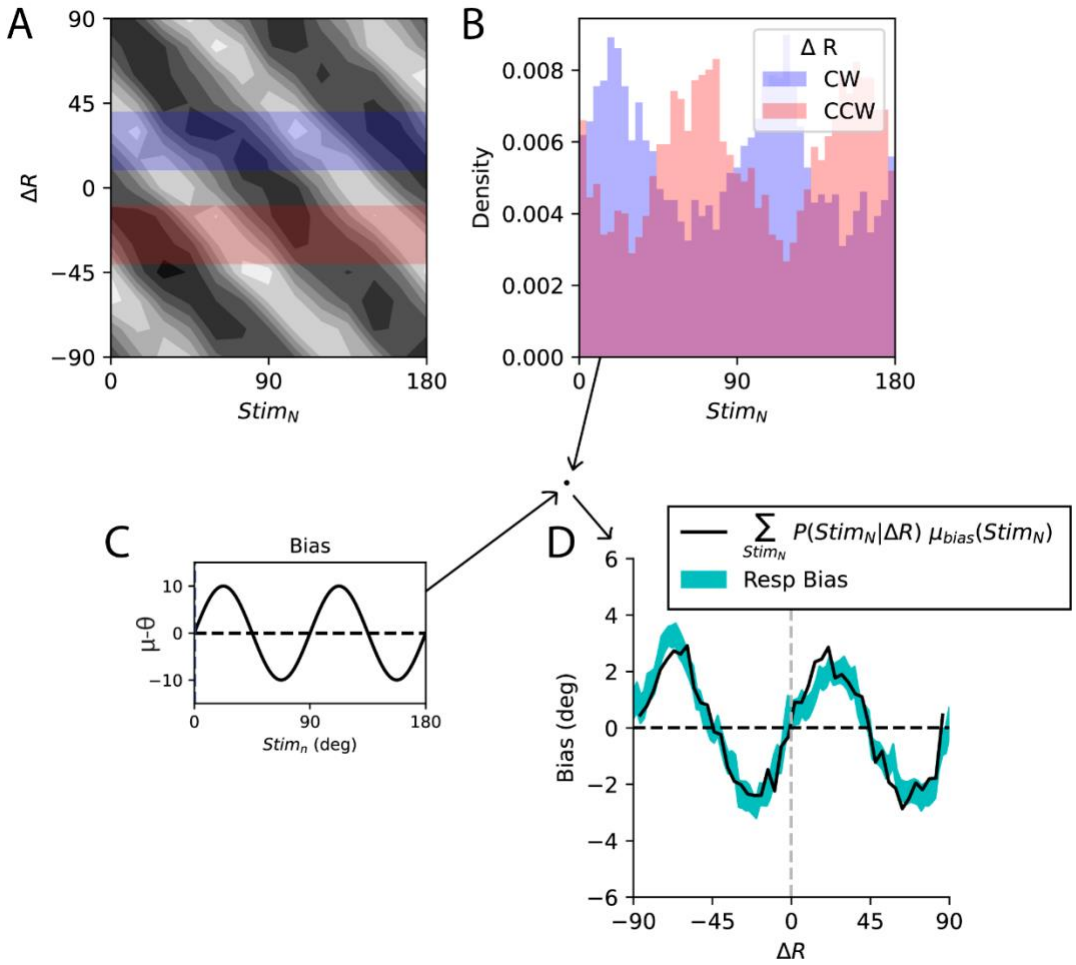


Figure 2-5 Artifactual serial dependence due to context independent biases.

A. The distribution of ΔR is not independent of $Stim_N$. **B**. We illustrate the distribution of $Stim_N$ for the subsets of trials highlighted in (A). **C.** Expected error as a function of $Stim_N$. **D.** Response bias (teal± SEM) is captured by the product of $P(Stim_N|\Delta R)$ and μ(Stim) (black).

Simultaneous modeling of stimulus and response

Two recent studies have tried to disentangle the relative contributions of stimulus and response history biases (Sadil et al., 2021; Moon and Kwon, 2022). Using this approach, the two functions are fit simultaneously instead of fitting a single two parameter function separately to ΔS and to ΔR. Theoretically, this should better disentangle the sources of the bias and the approach has revealed the surprising possibility that stimuli could simultaneously be *repelled* from the previous stimulus but have an even larger attraction to the previous response (Sadil et

al., 2021; Moon and Kwon, 2022). This approach is interesting but may be problematic as the two regressors are highly collinear, which poses a challenge for interpreting the fit parameters. We applied this approach to two simulated datasets, our full model featuring cardinal bias and correction for that bias, and a new model which introduces repulsion from the previous stimulus (see Methods, *Joint Bias Parameterization)*. First, we visualized the average individual fits to our corrected errors (as presented in Figure 2-4C) and note that while our modeling approach correctly captures the predominant bias source, the non-causal source is still of a similar magnitude (Figure 2-6A, left). When we apply our joint fitting procedure to the same data, we are better able to capture the true underlying source of the bias (Figure 2-6A, right). To compare the effectiveness of these alternative approaches, we conducted a power analysis for detecting significant biases while varying trial counts and precision (see Methods). First, we note that our power to distinguish between stimulus and response biases was higher for low precision participants across model types (Figure 2-6B). Critically, however, we note that the independent model consistently detects a significant effect of the non-inducing feature (Figure 2-6B, *top*) while the joint model is much less likely to detect a significant non-causal effect (e.g., Figure 2-6B, *bottom*, ΔS is close to 0% power for the joint model given true response serial dependance). This suggests the joint model is better powered to avoid Type II errors. See Figure 2-11 for a power analysis further broken down by trial count.

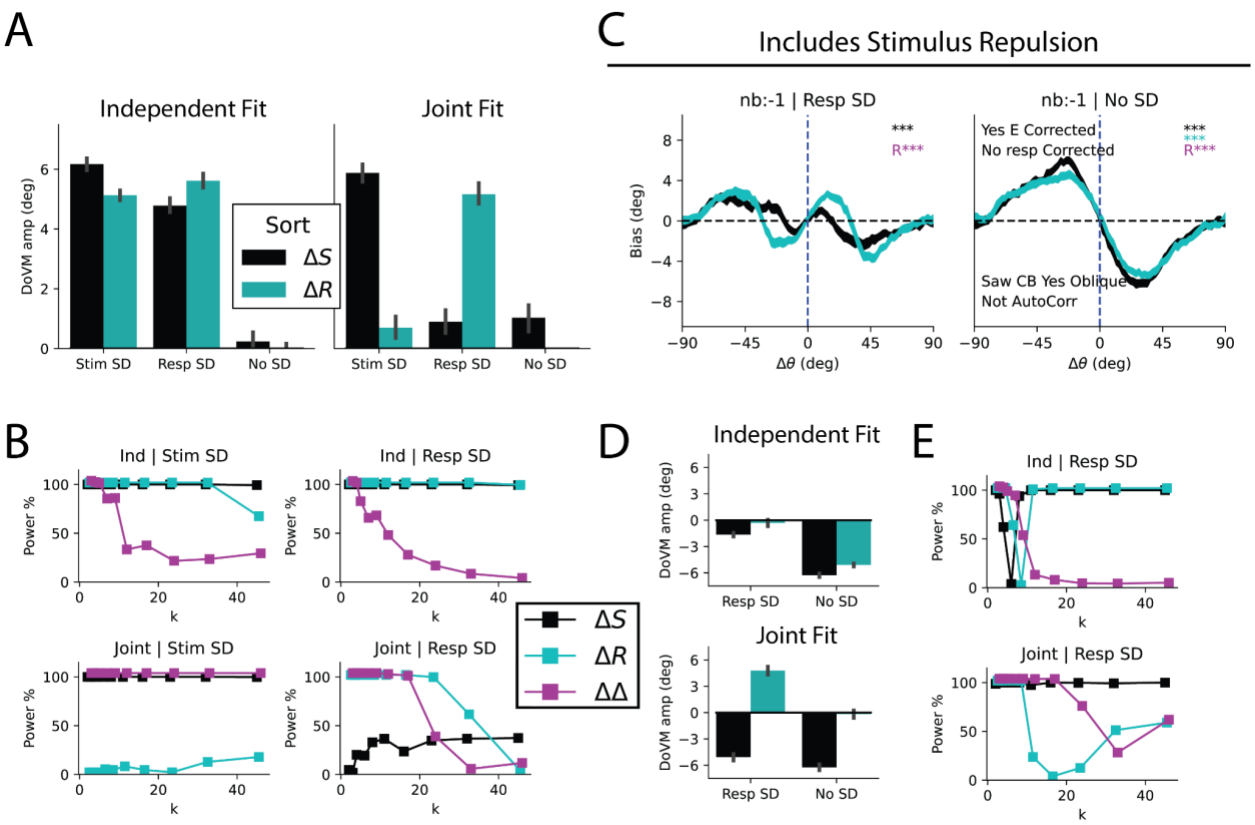


Figure 2-6 Simulated outputs of joint and independent model fits.

A. Fit magnitudes for independent and joint model fits. **B**. Power analysis across a range of k values for independent and joint models. Power is the % chance at finding a significant effect with n=30 participants at α=.001. ΔΔ refers to direct comparison of magnitude of ΔS and ΔR (paired t-test). **C**. Bias curves for an observer featuring stimulus repulsion, additional curves Figure S3. **D**. Joint fit is able to capture magnitudes and signs of true biases while independent model fails to separate the two. **E**. Power analysis reveals challenges in calculating bias magnitudes when the two competing forces are of approximately equal (0 power for ΔR at k=8 for independent model. Expanded power analysis presented in Figure S4.

We next applied the same approach to an observer featuring repulsion from the previous stimulus implemented at encoding to determine how well the joint/independent models captured these opposing effects. This is challenging because stimulus repulsion acts to counteract the influence of response attraction (Figure 2-6C, Figure 2-10). We found the joint model was better able to capture the underlying bias source (Figure 2-6D) and generally had much better power at distinguishing between their influences across a range of conditions (Figure 2-6E, bottom, Figure 2-11). This power analysis revealed an interesting phenomenon that may be common in the serial dependance field. For the independent model, particular values of k led to stimulus and response

biases that largely counteracted one another leading to 0% power (Figure 2-6E, top). Importantly, the joint model was able to reliably detect response biases over this same range (Figure 2-6E, bottom). This idea of opposing attractive and repulsive biases could suggest why null or weak results are common in studies of serial dependance and may provide a new avenue to analyze existing datasets.

Application to Empirical Data

We conclude by applying the techniques and principles developed above to an existing unpublished dataset. Participants (N=18, 6840 trials total) viewed a sequence of two oriented gratings presented foveally in succession and reported one of the stimuli by rotating a response dial with the keyboard after a 3.5s delay period. This experiment included partially valid retro-cues, the full details of which are described in the Methods and schematized (Figure 2-7A, Figure 2-12A). We first noted that responses showed strong *context-independent* biases that were non-sinusoidal (Figure 2-7B, gray). We first attempted to fit *context-independent* biases using a 6 parameter Fourier-like function as with our simulation, but found it was a poor match with large residuals (Figure 2-7B, *light green*). To fully capture the structure, we instead opted for a 12-parameter version which achieved a much tighter fit and smaller residuals (Figure 2-7B, *dark green*). We then examined history biases non-parametrically for the N-1 trial with and without shuffling trial order. For the shuffled responses, the correction procedure removes a spurious response bias seen in the raw responses (Figure 2-7C, bottom). The *In Order* trials show strong stimulus- and response-based biases (Figure 2-7C, top). We next examined stimulus and response biases both separately and using a joint model. To improve our power, we bootstrapped responses by randomly resampling 360 trials with replacement for 1024 surrogate participants. Participants showed strong attractive biases when sorting by ΔS_{N-1} & ΔR_{N-1} (Figure 2-7D, left).

Critically our correction procedure removed the *context-independent* bias artifact (Figure 2-7C, bottom-right). Consistent with our previous simulations, we found that response biases were inflated for all analyses and are significantly greater than 0 after shuffling when we didn't correct for *context-independent* biases (Figure 2-12C). When quantifying history biases independently, both stimulus and response biases were highly significant, but response biases were significantly stronger (Figure 2-7D, left, *In Order*). Importantly, we did not observe any stimulus or response biases for the shuffled trial sequence (Figure 2-7D, left, *Shuffle*). When we applied the joint fitting procedure, we found that only response bias was significantly greater than 0 suggesting that response biases are the dominant source of attractive biases in this data set. We thus demonstrate that our analysis procedure can be applied to empirical datasets and that simultaneously modeling biases can lead to insights otherwise hidden by traditional approaches.

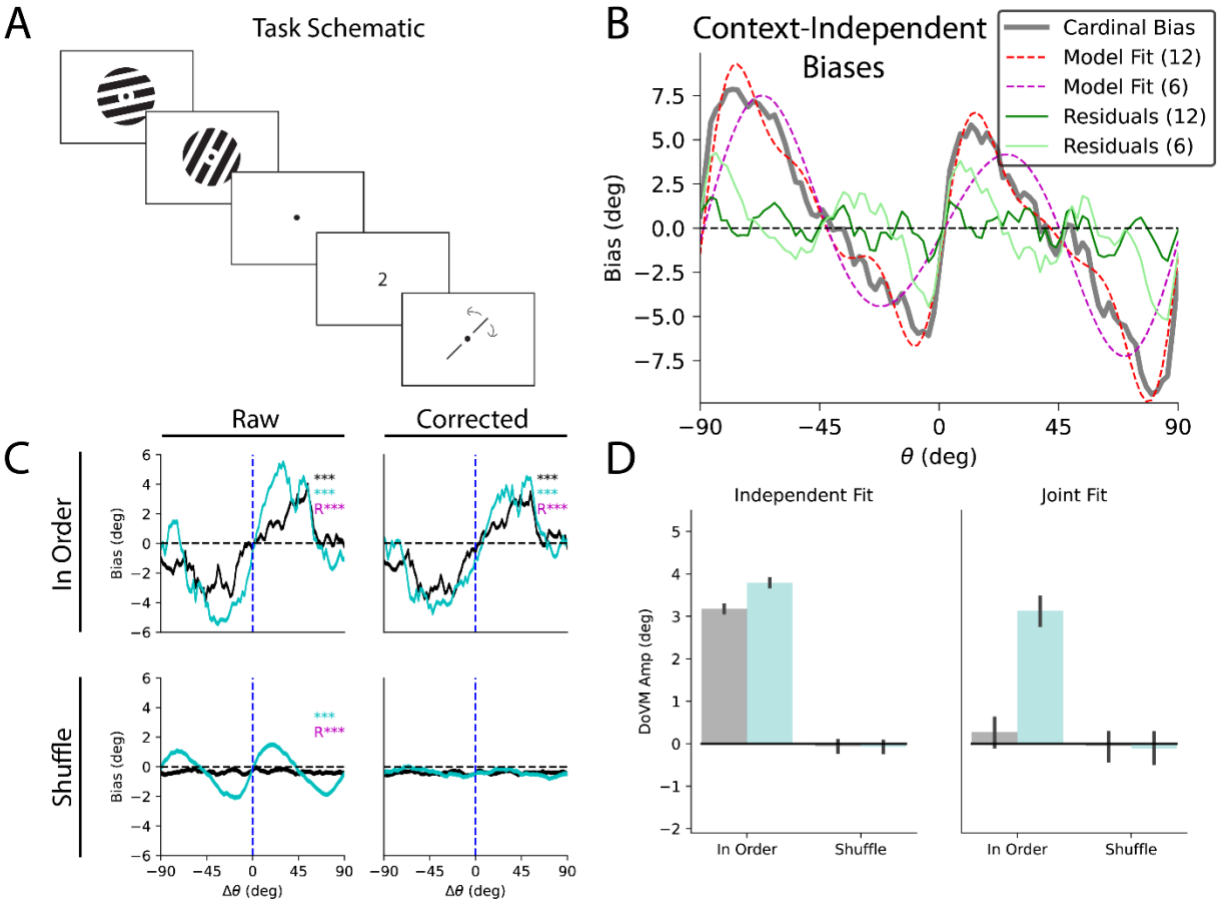


Figure 2-7 Application of joint model to empirical data reveals strong evidence for biases centered on the previous response.

A: simplified task schematic. Participants reported 1 of 2 foveally presented stimuli after a delay. **B**: Responses showed strong context-independent biases (gray). These were corrected by fitting a 12-parameter Fourier based parameterization to the pooled errors (red) resulting in unbiased residuals (green). **C**: Top, N-1, both raw and corrected responses show larger biases when sorting by past responses than stimuli; bottom, shuffle, uncorrected responses show a spurious response bias after shuffling trial order (left) that is eliminated after context-independent correction (right). **D**: While the independent model suggests both stimulus and response biases, joint model reveals bias is driven by responses.

Discussion

The goal of this modeling work was to provide a comprehensive exploration of methods to dissociate stimulus and response biases in the presence of potentially confounding *context-independent* biases such as *cardinal bias*. This work was motivated by an acute interest in analyzing response biases combined with a concern that any bias measured could be an artifact

of the analysis procedure. We first recap the lessons from our simulations and then discuss considerations that need to be made when analyzing such biases in empirical studies. Last, we briefly consider the psychological implications of our own empirical findings and recent related work.

We first identified a spurious future bias that is found specifically when sorting by ΔR_{N+1} (Figures 2-3). This bias is only observed in the presence of true response biases and is found in the absence of (or after correcting for) *context-independent* biases. This phenomenon is a signature of response biases and may be interpreted as evidence for previous responses rather than previous stimuli inducing a history bias (and notably this bias does not emerge under stimulus induced biases, Figure 2-2D). Importantly, there is no analogous spurious future bias after shuffling the trial order before assessing serial dependence (Figure 2-2H). Thus, the analysis of ΔR_{N+1} biases should primarily be used as a confirmatory step for the presence of response biases rather than a control for the influence of *context-independent* biases.

More problematic are artifacts introduced by *context-independent* biases (e.g., *cardinal bias).* These can lead to a spurious attraction between shuffled responses (Figure 2-4C). In our simulations, the spurious response biases were eliminated after regressing out this bias (Figure 2-4D, G). These biases emerge due to the influence of *context-independent* biases on all responses which is why shuffling does not remove them (Figure 2-5). When applying this correction procedure to our empirical dataset, the cardinal biases we observed were much steeper than the sine wave used in our simulation and necessitated additional higher frequency components to achieve truly unbiased residuals (Figure 2-7B). We increased the expressivity of our correction procedure until the errors sorted by $Stim_N$ and $\Delta R_{Shuffle}$ were flat and unbiased (ultimately using a model with 12 free parameters). We were then confident that any response biases were genuine

and not an artifact. Here, we observed a response bias ΔR_{N-1} that was significantly larger than our stimulus bias ΔS_{N-1} (Figure 2-7D, *Independent Fit*).

Lastly, we found promising results utilizing a joint modeling approach that was introduced in a pair of recent studies (Sadil et al., 2021; Moon and Kwon, 2022). Our analysis of simulated data showed that despite stimuli and responses being highly correlated, the joint approach was generally able to capture the true source of the bias (Figure 2-6 A, D). The reliability of this approach was greatly improved when participants were less precise and when there were greater trial counts per participant (Figure 2-6B, E, 2-11). Applying this approach to our empirical dataset revealed strong evidence for a history bias that originated from responses, not stimuli (Figure 2-7D, *Joint Fit*). Surprisingly, this response bias continued back many trials offering a new potential interpretation of past studies that have similarly long-acting biases (Figure 2-12) (Gekas et al., 2019; Fritsche et al., 2020). Our interpretation of this being a response driven bias is strengthened by the fact that other metrics, including the independent fits and the ΔR_{N+1} bias, all aligned closely with metrics observed for our response-driven simulated observer. Thus, simulated observers offer a valuable tool to infer the origin of biases given the outputs of the various metrics we have tested.

Throughout this manuscript, we present stimulus and response driven biases as if they are mutually exclusive. In reality, it is equally, if not more likely, that the inducing feature from the past is the *perceived* stimulus (rather than the response per se). This is supported by past work that has attempted to directly disambiguate perceived from reported orientations (Cicchini et al., 2017) or work that has utilized change detection rather than continuous report paradigms (Fischer and Whitney, 2014; Fritsche et al., 2017; Sheehan and Serences, 2022). That said, others have shown that attraction is not generated unless a stimulus is reported and that attraction may

instead be towards the reported rather than perceived location (Pascucci et al., 2019; Sheehan et al., 2022). In any case, with continuous report paradigms we often don't have any means of directly accessing the identity of the perceived stimulus and so we opt here to use the more general term of "response" throughout this paper as the behavioral response is typically the best/only proxy for the internal perceptual representation. Further disambiguating the physical act of responding (and the associated motor/decisional circuits) from the perception of the stimulus will require careful experimental designs or neural measures that can assess internal representations at different stages of information processing. Thus, finding a bias driven by past responses (rather than physical stimulus identity) as we did primarily suggests that attraction is toward a post-retinal representation or transformation of the stimulus. In retrospect this claim may seem obvious, as the brain has no access to the stimulus per se and will always be relying on internal representations that deviate from the original stimulus feature (Lettvin et al., 1959; Eggermont, 2007; György Buzsáki, 2019).

Now that there are several studies showing strong evidence for response over stimulus driven effects (Sadil et al., 2021; Moon and Kwon, 2022), the goalposts have shifted to further disambiguate exactly which response related components are driving these effects. Change detection paradigms or generally un-correlating responses from perception offer promising avenues to explore this possibility further (Braun et al., 2018; Sheehan et al., 2022; Zhang and Luo, 2022). That said, we argue here that examining biases just as a function of the physical identity of the previous stimulus is ignoring the important role of other biases in shaping the perception of current and past stimuli and may lead to an under and mismatched measurement of the true underlying bias (Pascucci et al., 2019; Sadil et al., 2021).

In the behavioral experiment we report here, there was no direct correlation between the final response and motor action as the probe was initialized in a random location and was controlled by button presses. Thus, we can likely rule out a purely motor origin for the attractive biases that we observed. The nidus of the attractive effect could instead be residual traces tied to memory maintenance, a distinct circuit directly tied to representing sensory history, or plausibly a sensory effect tied to the response or feedback signal presented at the end of the trial (Akrami et al., 2018; Barbosa et al., 2020). Only through additional experiments and analyses that control for these additional possible sources of perceptual biases can we further refine our understanding of these processes.

By demonstrating that the influence of *context-independent* biases can be reliably corrected for – while simultaneously highlighting the concerns raised if they are not – we hope to guide future endeavors to identify the true source of history biases. In our own experiment, we found strong evidence for an attractive bias centered on the previous response rather than the physical identity of the stimulus. We further found evidence for this attraction extending back 6 trials and separate evidence for a repulsion from the physical identity of the stimulus for trials 2, 3, 5 and 6 trials back. This pattern matches prior observations and supports the idea that the stimulus presentation leads to a repulsive bias at encoding while more high-level decisional representations impose a prior of stability (Pegors et al., 2015; Papadimitriou et al., 2016; Braun et al., 2018; Pascucci et al., 2019; Zhang and Alais, 2020; Sadil et al., 2021; Moon and Kwon, 2022; Sheehan and Serences, 2022; Zhang and Luo, 2022). Such a framework additionally fits with general frameworks like efficient encoding and Bayesian inference seen in perception (Wei and Stocker, 2015) and pattern separation and completion seen in various networks across the brain (Cayco-Gajic and Silver, 2019).

Acknowledgments: Chapter 2

Thanks to the Gibbons and juncos that hosted me while I did the of the initial simulations of this work.

Chapter 2, in full, is a reprint of the material under preparation. Sheehan, Timothy C.; Serences, John T. The dissertation/thesis author was the primary investigator and author of this paper.

Works Cited

Akrami A, Kopec CD, Diamond ME, Brody CD (2018) Posterior parietal cortex represents sensory history and mediates its effects on behaviour. Nature 554:368–372.

Anstis S, Verstraten FAJ, Mather G (1998) The motion aftereffect. Trends Cogn Sci 2:111–117.

Barbosa J, Stein H, Martinez RL, Galan-Gadea A, Li S, Dalmau J, Adam KCS, Valls-Solé J, Constantinidis C, Compte A (2020) Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. Nat Neurosci 23:1016–1024.

Barlow HB (1961) Possible principles underlying the transformation of sensory messages. Sens Commun 1:217–234.

Benucci A, Saleem AB, Carandini M (2013) Adaptation maintains population homeostasis in primary visual cortex. Nat Neurosci 16:724–729.

Braun A, Urai AE, Donner TH (2018) Adaptive History Biases Result from Confidence-Weighted Accumulation of past Choices. J Neurosci 38:2418–2429.

Cayco-Gajic NA, Silver RA (2019) Re-evaluating Circuit Mechanisms Underlying Pattern Separation. Neuron 101:584–602.

Ceylan G, Herzog MH, Pascucci D (2021) Serial dependence does not originate from low-level visual processing. Cognition 212:104709.

Chopin A, Mamassian P (2012) Predictive Properties of Visual Adaptation. Curr Biol 22:622–626.

Cicchini GM, Anobile G, Burr DC (2014) Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. Proc Natl Acad Sci 111:7867–7872.

Cicchini GM, Burr DC (2018) Serial effects are optimal. Behav Brain Sci 41:e229.

Cicchini GM, Mikellidou K, Burr D (2017) Serial dependencies act directly on perception. J Vis 17:6.

Corbett JE, Fischer J, Whitney D (2011) Facilitating Stable Representations: Serial Dependence in Vision Zochowski M, ed. PLoS ONE 6:e16701.

Dragoi V, Sharma J, Sur M (2000) Adaptation-Induced Plasticity of Orientation Tuning in Adult Visual Cortex. Neuron 28:287–298.

Eggermont JJ (2007) Correlated neural activity as the driving force for functional changes in auditory cortex. Hear Res 229:69–80.

Felsen G, Touryan J, Dan Y (2005) Contextual modulation of orientation tuning contributes to efficient processing of natural stimuli. Netw Comput Neural Syst 16:139–149.

Fischer J, Whitney D (2014) Serial dependence in visual perception. Nat Neurosci 17:738–743.

Fritsche M (2016) To Smooth or not to Smooth: Investigating the Role of Serial Dependence in Stabilizing Visual Perception. Available at: https://theses.ubn.ru.nl/handle/123456789/3193 [Accessed January 4, 2023].

Fritsche M, Mostert P, de Lange FP (2017) Opposite Effects of Recent History on Perception and Decision. Curr Biol 27:590–595.

Fritsche M, Spaak E, de Lange FP (2020) A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. Neuroscience. Available at: http://biorxiv.org/lookup/doi/10.1101/2020.01.22.915553 [Accessed March 2, 2020].

Gallagher GK, Benton CP (2022) Stimulus uncertainty predicts serial dependence in orientation judgements. J Vis 22:6.

Gekas N, McDermott KC, Mamassian P (2019) Disambiguating serial effects of multiple timescales. J Vis 19:24.

Girshick AR, Landy MS, Simoncelli EP (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nat Neurosci 14:926–932.

György Buzsáki M (2019) The brain from inside out. Oxford University Press.

He BJ, Zempel JM, Snyder AZ, Raichle ME (2010) The Temporal Structures and Functional Significance of Scale-free Brain Activity. Neuron 66:353–369.

He S, MacLeod DIA (2001) Orientation-selective adaptation and tilt after-effect from invisible patterns. Nature 411:473–476.

Henderson M, Serences JT (2021) Biased orientation representations can be explained by experience with nonuniform training set statistics. J Vis 21:10.

Kohn A, Movshon JA (2004) Adaptation changes the direction tuning of macaque MT neurons. Nat Neurosci 7:764–772.

Lettvin JY, Maturana HR, McCulloch WS, Pitts WH (1959) What the Frog's Eye Tells the Frog's Brain. Proc IRE 47:1940–1951.

Manassi M, Liberman A, Kosovicheva A, Zhang K, Whitney D (2018) Serial dependence in position occurs at the time of perception. Psychon Bull Rev 25:2245–2253.

Maus GW, Chaney W, Liberman A, Whitney D (2013) The challenge of measuring long-term positive aftereffects. Curr Biol CB 23:10.1016/j.cub.2013.03.024.

Moon J, Kwon O-S (2022) Dissecting the effects of adaptive encoding and predictive inference on a single perceptual estimation. :2022.02.24.481765 Available at: https://www.biorxiv.org/content/10.1101/2022.02.24.481765v1 [Accessed August 31, 2022].

Mumford D (1994) Pattern Theory: A Unifying Perspective. In: First European Congress of Mathematics: Paris, July 6-10, 1992 Volume I Invited Lectures (Part 1) (Joseph A, Mignot F, Murat F, Prum B, Rentschler R, eds), pp 187–224 Progress in Mathematics. Basel: Birkhäuser. Available at: https://doi.org/10.1007/978-3-0348-9110-3_6 [Accessed January 4, 2023].

Oliver BM (1952) Efficient Coding. Bell Syst Tech J 31:724–750.

Olshausen BA, Field DJ (1996) Natural image statistics and efficient coding. Netw Comput Neural Syst 7:333–339.

Papadimitriou C, White RL, Snyder LH (2016) Ghosts in the Machine II: Neural Correlates of Memory Interference from the Previous Trial. Cereb Cortex:bhw106.

Pascucci D, Mancuso G, Santandrea E, Della Libera C, Plomp G, Chelazzi L (2019) Laws of concatenated perception: Vision goes for novelty, decisions for perseverance Tong F, ed. PLOS Biol 17:e3000144.

Patterson CA, Wissig SC, Kohn A (2013) Distinct Effects of Brief and Prolonged Adaptation on Orientation Tuning in Primary Visual Cortex. J Neurosci 33:532–543.

Pegors TK, Mattar MG, Bryan PB, Epstein RA (2015) Simultaneous perceptual and response biases on sequential face attractiveness judgments. J Exp Psychol Gen 144:664–673.

Sadil P, Cowell R, Huber DE (2021) The Push-pull of Serial Dependence Effects: Attraction to the Prior Response and Repulsion from the Prior Stimulus. Available at: https://psyarxiv.com/f52yz/ [Accessed January 4, 2023].

Sheehan T, Carfano B, Serences J (2022) Serial dependence to prior stimuli and past responses. J Vis 22:4401.

Sheehan TC, Serences JT (2022) Attractive serial dependence overcomes repulsive neuronal adaptation. PLOS Biol 20:e3001711.

van Bergen RS, Jehee JFM (2019) Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. Neuroscience. Available at: http://biorxiv.org/lookup/doi/10.1101/671958 [Accessed March 2, 2020].

Wei X-X, Stocker AA (2015) A Bayesian observer model constrained by efficient coding can explain “anti-Bayesian” percepts. Nat Neurosci 18:1509–1517.

Zhang H, Alais D (2020) Individual difference in serial dependence results from opposite influences of perceptual choices and motor responses. J Vis 20:2.

Zhang H, Luo H (2022) Co-occurrence of past and present shifts current neural representations and mediates serial biases. :2022.06.08.495281 Available at: https://www.biorxiv.org/content/10.1101/2022.06.08.495281v1 [Accessed January 4, 2023].

Supplemental Materials

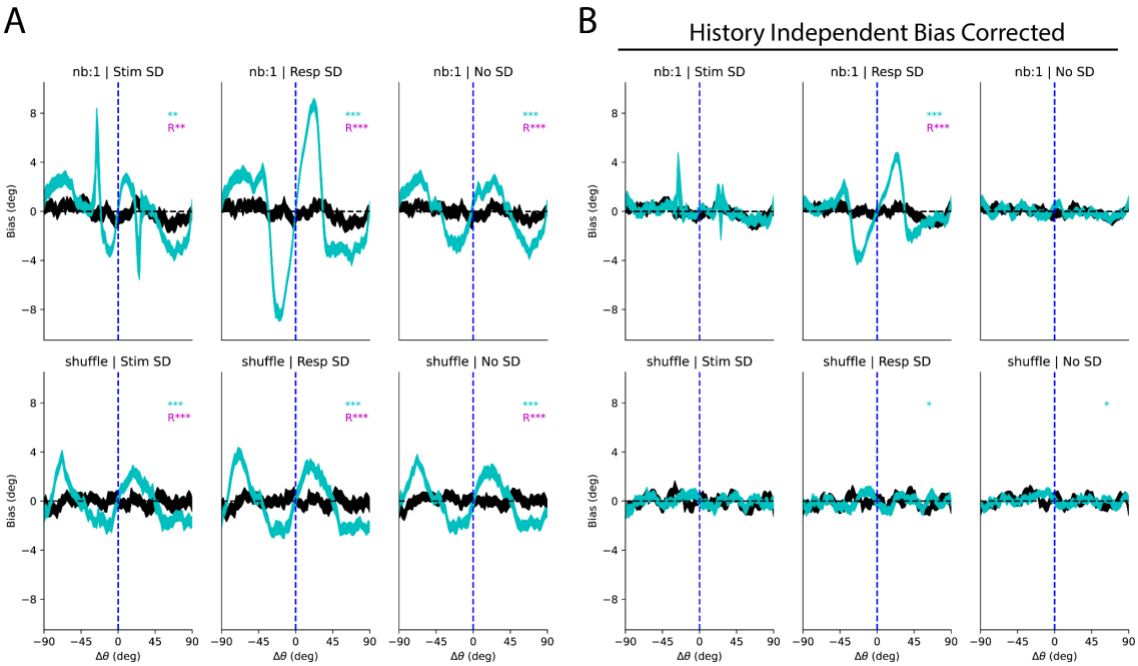


Figure 2-8 Bias curves for N+1 and shuffled distribution for corrected (A) and uncorrected (B) errors from Figure 4.

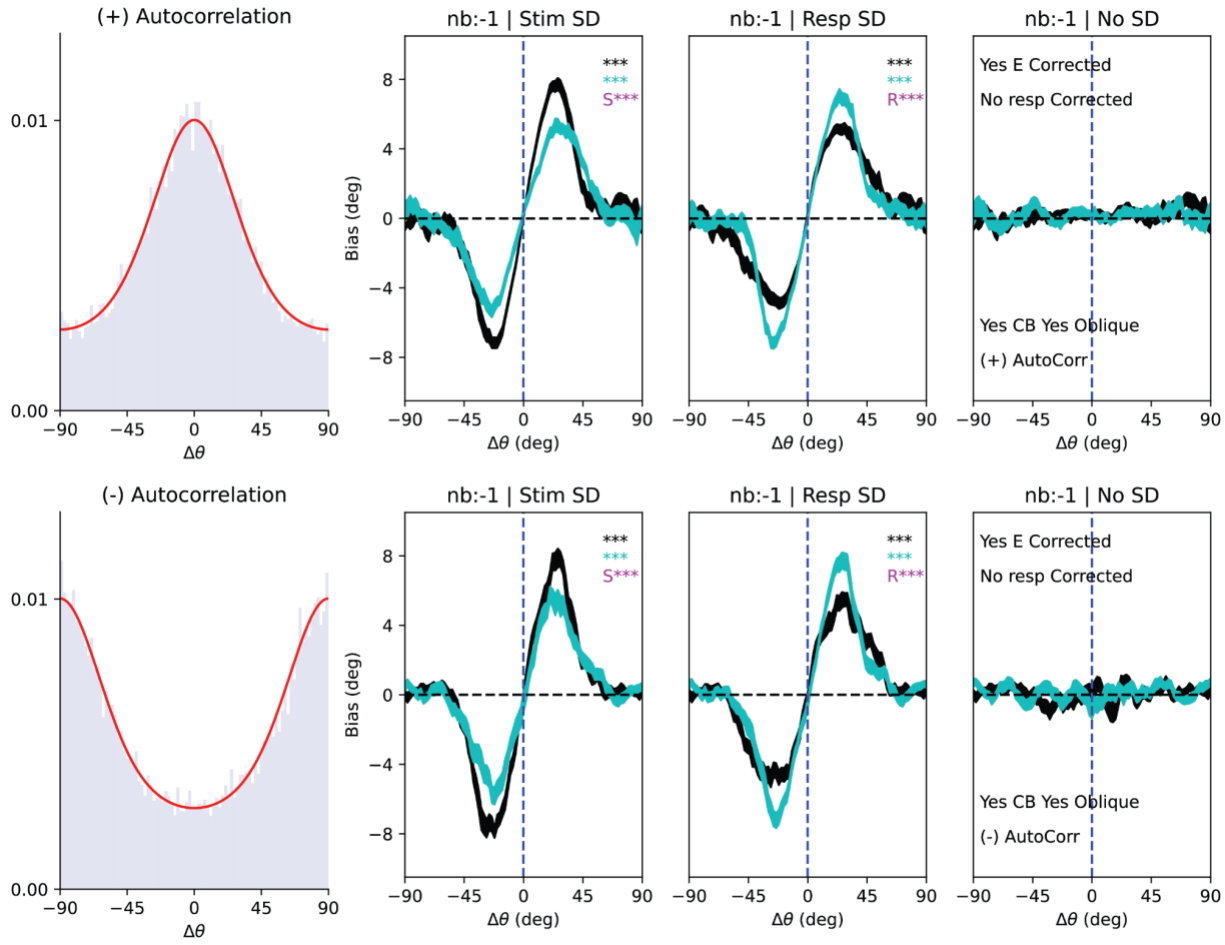


Figure 2-9 Non-independent Stimulus Sequences.

We simulated the analysis of observers where stimulus sequences were non-independent and exhibited strong positive (top left) or negative (bottom left) autocorrelations. Despite the presence of these strong stimulus autocorrelations, their presence alone does not introduce any additional artifacts into our analysis procedure.

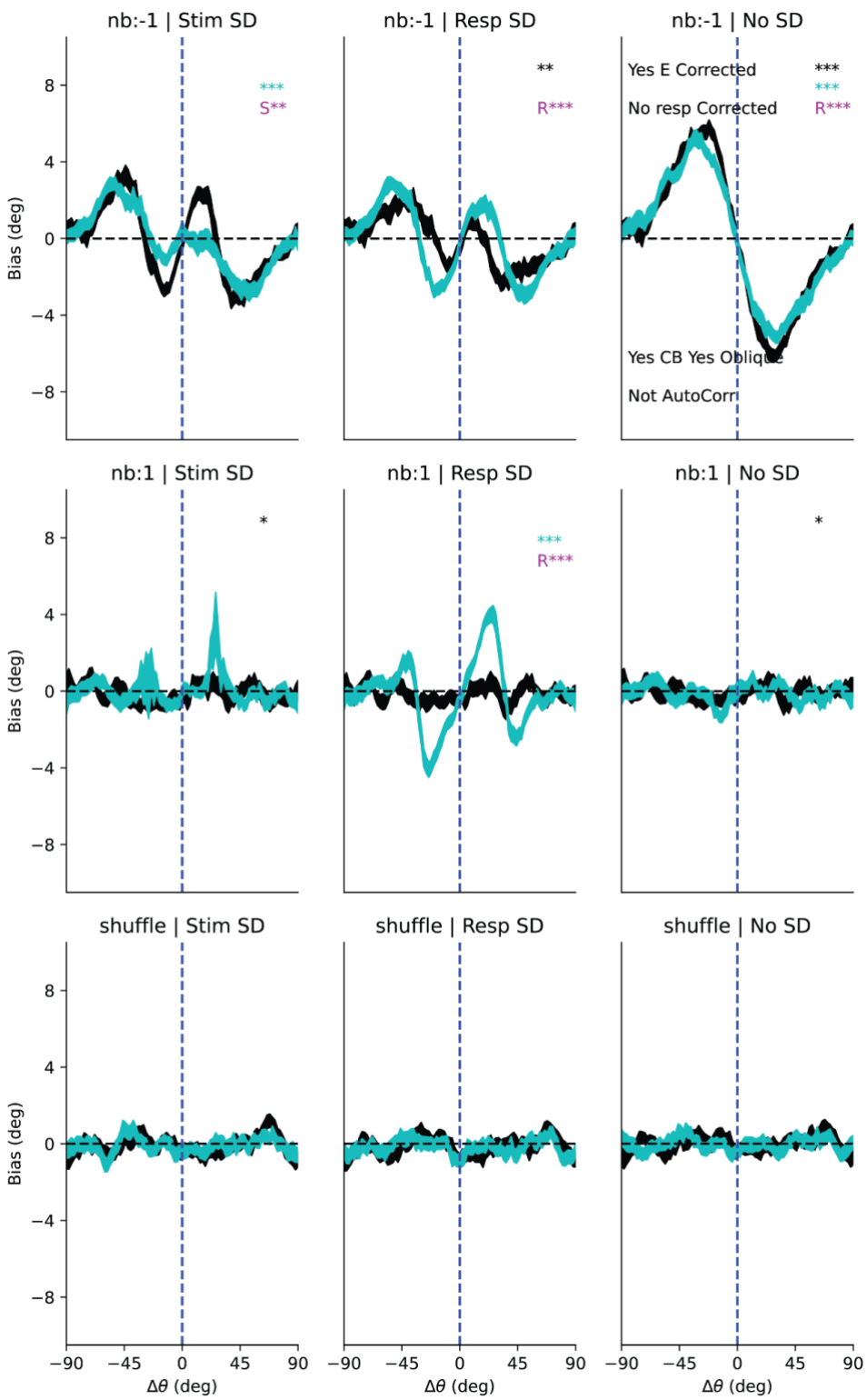


Figure 2-10 All bias curves for observer with stimulus specific repulsion.

Note that the left column is an observer that is both repelled at encoding and attracted at a later Bayesian integration stage (aligning with previously proposed models, Fritsche et al., 2020).

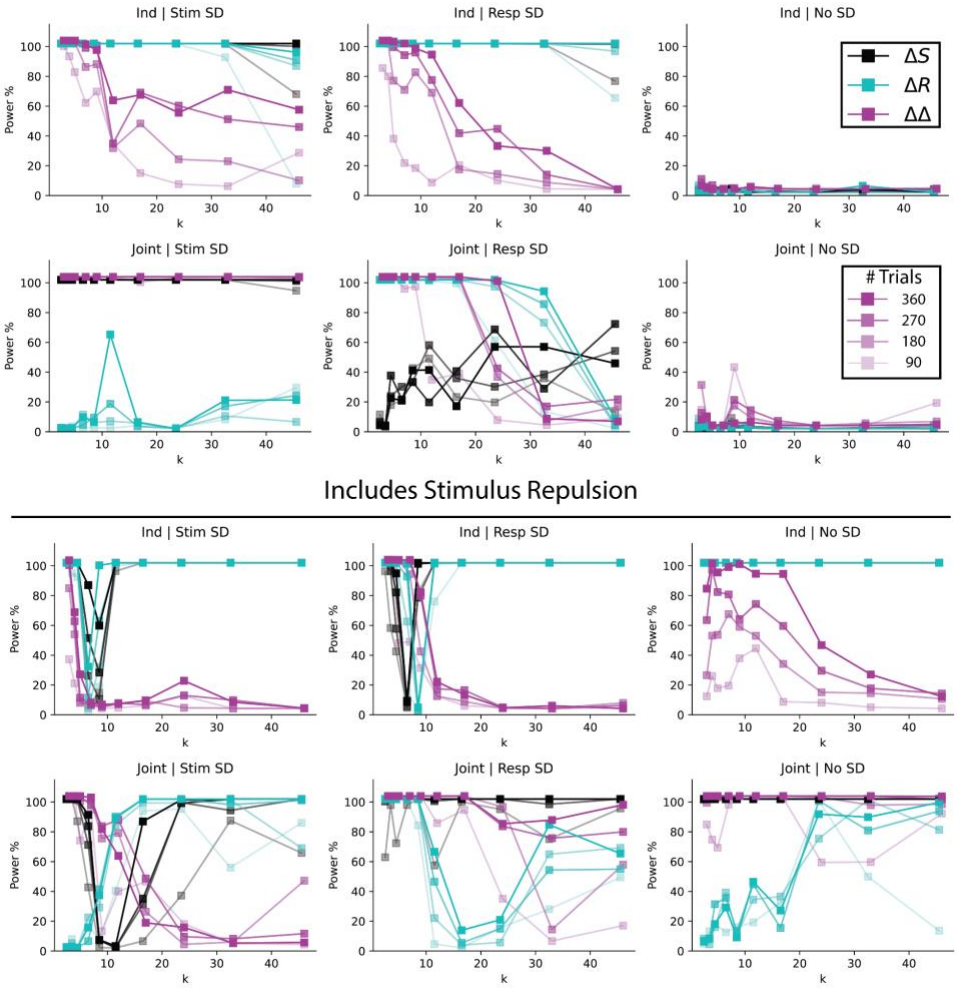


Figure 2-11 Expanded power analysis.
Expanded power analysis for observers without (top) and with (bottom) stimulus repulsion at encoding. Here we split out observers based on the number of trials completed per observer. Power values correspond to α=.001 for an experiment run with 30 participants.

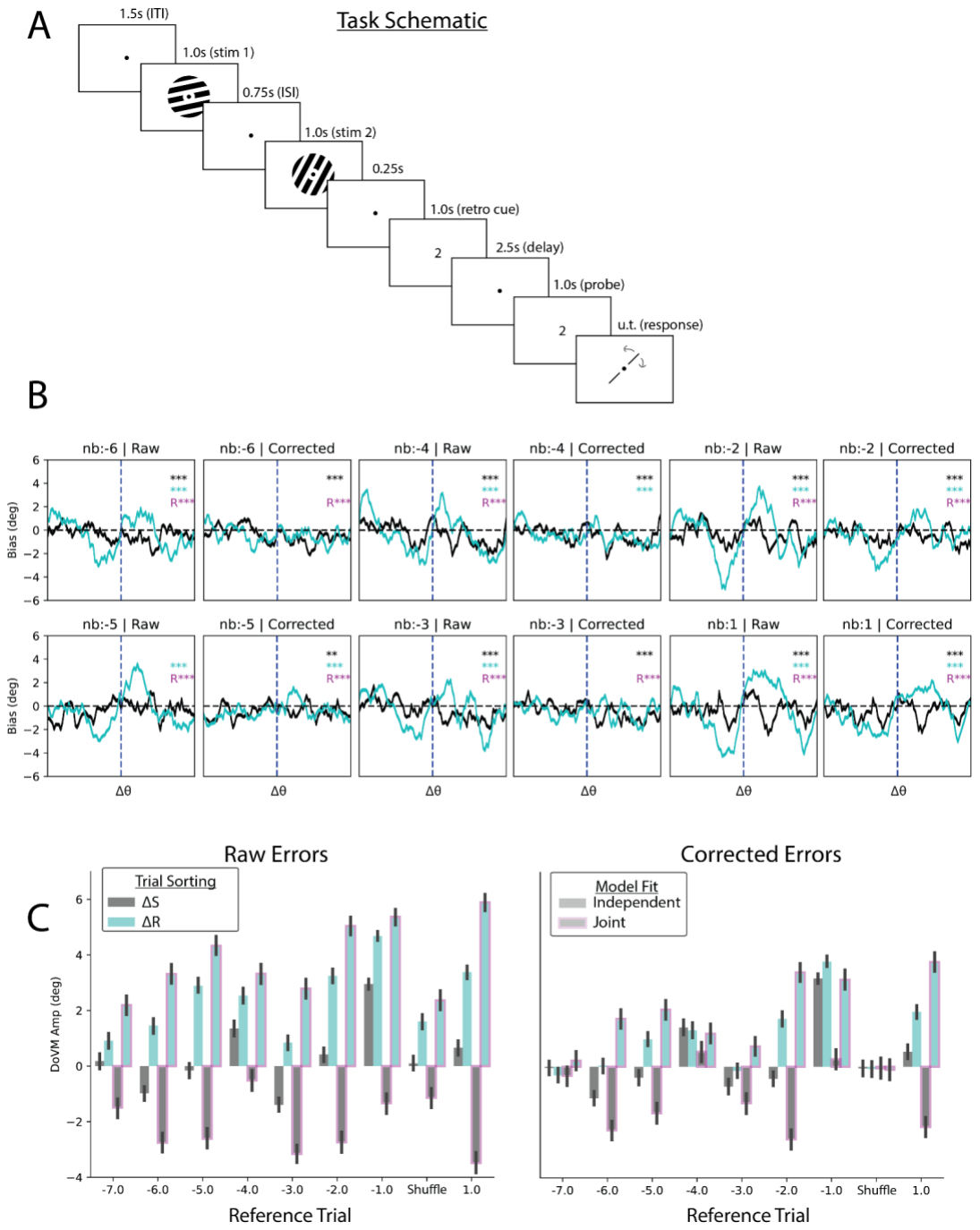


Figure 2-12 Expanded empirical analysis.

A. Full task schematic from delayed report paradigm. A Probabilistic retro-cue (80%) valid was presented immediately after the second item followed by a 100% valid probe and an untimed continuous report task controlled via the keyboard. Probe location initialized to a random location on each trial. **B.** Expanded stimulus and response bias curves for corrected and uncorrected errors for different number of trials back and using shuffled distribution. **C.** Quantified bias fits for both independent (no outline) and joint (magenta outline) models. Correcting errors removes spurious biases in the shuffled distribution (right, shuffle). Joint model reveals attraction to reported stimulus going back several trials.

Chapter 3 Temporal dependencies across perception, decision, and action

Abstract

Perceptual reports across a range of tasks are attracted towards recent stimulus features. This phenomenon is termed serial dependence and could arise from exposure to a natural world that changes slowly over time. Thus, adopting a prior of temporal stability – and allowing recent events to bias the processing of current events – might provide a powerful mechanism to improve the fidelity of information processing (particularly under high uncertainty). Efforts to identify the origin of serial dependance have shown that stimulus strength, attention, working memory, and motor responses all impact the magnitude of serial biases. Thus, rather than a single mechanism, serial dependence may arise due to a canonical prior for stability present across distinct circuits that collectively support different cognitive functions. To test this hypothesis, we systematically manipulated visual stimulation, attention, task relevance, and motor output in a series of working memory experiments. In addition to a standard spatial working memory task, we also used a novel "compass" task that used abstract, semantic cues to indicate the remembered location, critically allowing us to disambiguate memory items from physical environment features. We found robust and generalizable attractive biases towards past responses regardless of visual stimulation, as well as biases towards unreported or attended stimuli. Additionally, the strength and timescale of biases depended on whether they were encoded visually or semantically. Across 4 experiments, we found evidence for concurrent biases that are both visual and decisional in nature. Our results suggest that biases for temporal stability

are not restricted to one level such as early sensory processing. Instead, these biases are likely present at multiple levels of the stimulus-response loop, where they act at different timescales to support domain-specific cognitive operations.

Introduction

Serial dependence is the phenomenon whereby perceptual reports are attracted towards past sensory experiences or actions. This bias allows perception to better match the statistics of naturally occurring stimuli which tend to be stable across time (and timescales) (Dong & Atick, 1995; Felsen et al., 2005; van Bergen & Jehee, 2019). Thus, leveraging the stability of sensory inputs to continuously inform information processing may reduce noise and increase overall efficiency (Cicchini & Burr, 2018; Fischer & Whitney, 2014; Kiyonaga et al., 2017; Sheehan & Serences, 2022). Unlike repulsive effects such as the waterfall illusion or surround tilt illusion, which are generally well explained by stimulus specific adaptation and divisive normalization respectively (Benucci et al., 2013; Clifford, 2014; Schwartz et al., 2009), a consensus on the origins of serial dependence is still lacking.

Identifying the mechanism(s) that support serial dependence is challenging because the phenomenon is observed in many domains and is mediated by a host of cognitive factors: serial dependence arises when processing both low-level and complex features (Fischer & Whitney, 2014; Suárez-Pinilla et al., 2018), in multiple sensory domains (Fornaciai & Park, 2019; Zhang & Luo, 2023), and even when processing semantic or social knowledge (Collins, 2022). This ubiquity across stimulus and cognitive domains has also led to apparent contradictions in the literature. For instance, while serial dependence effects seem to emerge immediately after stimulus presentation in some studies (Cicchini et al., 2017; Fischer & Whitney, 2014; Manassi et al., 2018), serial dependence only emerges after a memory delay in others (Bliss et al., 2017; Papadimitriou et al., 2015),. These seemingly contradictory findings might be explained by relatively subtle changes to stimulus properties or task design. For example, when using low

contrast stimuli that induce high degrees of perceptual uncertainty, existing data suggest that changes in sensory processing are akin to anti-adaptation (Fischer & Whitney, 2014) and resemble 'visual persistence' effects that might contribute to attraction. On the other hand, for high contrast stimuli that minimize perceptual uncertainty, data suggest that serial dependence is more closely tied to activity-dependent plasticity in memory circuits (Barbosa et al., 2020; Bliss & D'Esposito, 2017) or to shifts in attentional gain (Papadimitriou et al., 2016).

In addition to the influence of sensory uncertainty and memory maintenance, the relative contribution of other factors has also been debated. For instance, some work suggests that serial dependence is largely due to changes in perceptual processing (Collins, 2020; Fischer & Whitney, 2014; Fornaciai & Park, 2018; Liberman et al., 2014; Murai & Whitney, 2021; Suárez-Pinilla et al., 2018) whereas others suggest a decision or response-level account (Ceylan et al., 2021; Feigin et al., 2021; Moon & Kwon, 2022; Pascucci et al., 2019; Zhang & Alais, 2020) see also (Cicchini et al., 2021). Similarly, although attention has been shown to enhance serial dependance (Bae & Luck, 2020; Fischer & Whitney, 2014; Fritsche & de Lange, 2019; Makovski & Jiang, 2008; Suárez-Pinilla et al., 2018), attractive biases have also been observed for passively viewed or even task irrelevant features (Fornaciai & Park, 2018; Murai & Whitney, 2021). Lastly, the neural basis of serial dependence are under debate with different studies pointing to sensory (Ranieri et al., 2022), association (Akrami et al., 2018), working memory (Barbosa et al., 2020), motor (Neto & Bartels, 2021), executive (Schwiedrzik et al., 2014), or more generally post perceptual (Hajonides et al., 2023; Papadimitriou et al., 2016; Sheehan & Serences, 2022) origins.

Thus, many attempts have been made to understand how serial dependencies arise with a focus on identifying a specific level of the processing stream and which factors do or not matter

for a specific experimental paradigm. Rather than explaining how a single process could give rise to all of these observations, here we explore the possibility that serial dependence is a canonical feature of information processing across perceptual and cognitive domains. Natural visual experience is quite stable across time, and these statistical dependencies are internalized from the earliest stages of development (Blakemore & Cooper, 1970; Dong & Atick, 1995; Felsen et al., 2005; Mayer, 1977; van Bergen & Jehee, 2019). Additionally, long term behavioral goals require the application of self-directed attention and working memory to further stabilize and hold constant neural representations of relevant features. Thus, internal representations of the world are apt to be even more stable than the outside world. In addition to perceptual and mnemonic processing, motor responses also exhibit a high degree of stability: even the simplest mobile organisms show a preference for stability in heading direction (e.g. Levy flights) to optimize foraging patterns given limited prior information (Deneubourg et al., 1990; Hays et al., 2011) for ants and jellyfish respectively. Therefore, organisms have a preference and fitness pressure towards stability at the levels of sensation, internal experience, and action. Temporal dependencies are a natural feature of neural information processing, arising from local properties such as tonic firing, short term synaptic plasticity, and recurrent feedback across areas. Serial dependence may be the result of these underlying mechanisms operating at multiple levels of processing (Bliss & D'Esposito, 2017; Burr & Cicchini, 2014; Urai & Donner, 2022; Zhang & Luo, 2023).

We sought to formalize this framework of stability being a universal property across several levels of processing. As experimenters, we typically only have access to the stimuli and the responses made by the individual (Figure 3-1, *Observed States*). Thus, to properly model all cognitive steps utilized in a typical perceptual-report task (Figure 3-1, *Internal States*) a

challenging inference problem is created. A stimulus is encoded by sensory circuits and relayed to executive areas which hold contextual information (e.g., task objectives) and direct future actions. Depending on task demands, this stimulus information may need to be held in memory and converted to a useful representation or plan baked on task contingencies. Finally, participants make some sort of motor response which may require integrating new sensory information with existing representations of task variables. Thus, any effects of sensory history could emerge from any of these prospective levels of information processing (recursive arrows, Figure 3-1, *Internal States*) unless task manipulations specifically target a given level of processing.

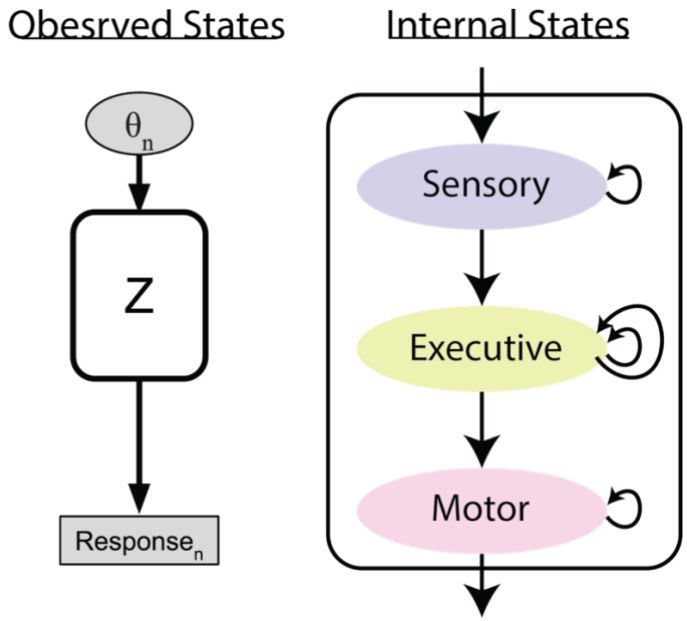


Figure 3-1 Inverse problem of behavioral inference.

As researchers interested in the internal workings of perception and decision making, we can devise experiments to present stimuli to participants and observe their responses. While such an approach is powerful, the observed states are quite limited with respect to the levels of internal states. For example, for a given perceptual report task, a participant's behavior will likely draw on their encoding of the stimulus through sensory circuits, maintaining and restructuring of this sensory signal to a task relevant format, and planning and executing the appropriate response to achieve good performance. This processing stream will utilize many conceptual levels of processing including "sensory", "executive", and "motor". In the context of measuring serial dependence, it is possible that autocorrelations across responses could arise from any of these levels unless task designs carefully control around this.

The goal of our modeling framework is to better understand which processes are likely to contribute to an observed serial bias. To do so, we examined serial biases using a novel paradigm that mixes typical low-level stimuli with abstract representations. This allowed us to disentangle biases that could emerge from low-level encoding from those requiring a less direct mapping to target feature. We found that serial dependencies occur for stimuli encoded with semantic cues ruling out a purely sensory origin of the effects, but also found effects are stronger for low-level features suggesting a contribution of sensory factors as well. We separately examined the role of past actions/reports and again found that while attraction was driven largely by the previous report, history effects were still detectable in some cases when no report was made, or the inducing stimulus was task irrelevant. Last, we found that history effects for more abstract stimuli have a longer time constant, a sensible consequence of biases emerging at multiple levels of the sensory-response loop. In sum, we take these findings as evidence that serial biases emerge from integration at many stages across the processing stream and conflicting findings may share a conceptual motivation but not a neural instantiation.

Results

We devised a "compass" working memory task that allowed us to disentangle the role of bottom-up sensory drive from the feature being remembered and reported. Put simply, if serial dependence exclusively relied on actions occurring at encoding, then one should not see attractive biases operating on stimuli encoded through wholly different processes. On standard "dot" trials, participants encoded and reported the location of a briefly presented black dot. In this condition there was a direct correspondence between stimulus and reported feature. On semantic "compass" trials, the to-be-remembered location was indicated by one of 16 possible abbreviations corresponding to a direction on a compass rose (Figure 3-2A). The compass

stimulus corresponds to an abstract representation of the relevant feature – a spatial location – that has no physical relationship with the target location. By including stimuli that must be encoded and processed in vastly different ways but ultimately map onto the same representational space (physical location on a ring) we aimed to better identify components giving rise to serial dependencies. For example, if serial dependencies are only observable between "dot" trials, this would argue that low level sensory encoding is a necessary component for serial biases. Alternatively, if there were no discernible differences in serial dependencies between the stimulus types, then this would suggest that only overlap in decisions or response plans is relevant. Across five experiments, we manipulated stimulus type, task timing, and the method of report. We found evidence consistent with *both* sensory and non-sensory contributions to serial dependence and argue that serial dependence can emerge at many levels of the stimulus-response loop.

Measuring serial dependence

We first examined classic serial dependence effects in our dataset, determined by how the identity of past stimuli impacted responses on future trials. To examine serial dependence, we quantified errors on the current trial as a function of the previous stimulus ($\Delta\theta=\theta_{n-1} - \theta_n$). In experiment 1, participants (n=19) completed 8.46±1.9 blocks of 96 trials. On each trial, the color of the compass cue indicated whether the "dot" location (red) or "compass" coordinates (white) should be reported after the delay (Figure 3-1A). Consistent with prior work, we found that spatial memory reports were attracted towards previously encoded spatial positions (Figure 3-1B, light blue) (Bliss et al., 2017; Papadimitriou et al., 2015). This attraction was well parameterized by a derivative of Von Mises function (DoVM, see Methods) and had an amplitude that was significantly greater than 0 (amp=1.75±0.327, t(18)=5.35, p=3.633e-05).

Serial dependence for semantic stimuli and across stimulus types

We next examined reports of "compass" stimuli. Overall accuracy was not different between the two stimulus types and responses were significantly more precise and quicker around cardinal directions (North, South, East, West) (Figure 3-7C-D). Responses also showed large stable biases as a function of target angle (generally repulsion from these cardinal directions) that differed between stimulus type but were consistent across individuals (Figure 3-7A). As we are primarily interested in quantifying how short-term history impacts responses, we fit a Fourier-like function to the distribution of errors made by each participant as a function of stimulus type and direction. All analyses are performed on the residualized errors (Figure 3-7B).

We observed a significant attractive serial dependence between consecutive compass trials suggesting serial biases do not only operate on purely sensory driven representations (amp=0.798±0.271 t=2.95, p=0.008296, Figure 3-2C). A separate cohort completed the same task but with compass and dot trials separated by blocks rather than determined by a change in the color of the compass cue (Experiment 1b "blocked"). Under these conditions, we continued to observe an attractive bias for both dot (amp=1.48±0.18, $t(12)=8.03$, p<1E-5) and compass (amp=1.13±0.19, $t(12)=5.97$, p<1E-4, Fig 3-8) stimuli. Next, we asked whether there was significant serial dependence across stimulus types (dot, compass). While attraction has been found between stimuli of different visual types (eg. orientation of a Gabor vs. the symmetry line in a dot array, (Ceylan et al., 2021)) they have generally not been found when the attended feature is of a different class altogether (eg. orientation vs. size, (Fritsche & de Lange, 2019); but see (Van der Burg et al., 2021)). Here, the visual encoding of our two stimulus classes (dot and compass) has absolutely no cross correspondence (as the compass must be processed semantically) but the imagined locations occupy the same space. Despite the differences in

sensory processing, we observed robust serial attraction between both cross stimulus sequences (Dot->Compass, amp=1.13±0.233, $t(18)=4.85$ p=0.0001114 and Compass->Dot, amp:1.94±0.222, $t(18)=8.75$ p=4.343e-08, Figure 3-2B-D). A mixed effects linear model revealed a significant effect of the current trial type such that the dot stimuli exhibited larger biases (coefficient = 1.14±0.38, p=.003, Figure 3-2D) with no effect of the previous stimulus type or interaction between the two. Thus, it seems that serial dependence can be observed towards recently attended features regardless of the manner in which they are encoded but tend to be stronger for stimuli that directly indicate the encoded feature rather than require semantic processing.

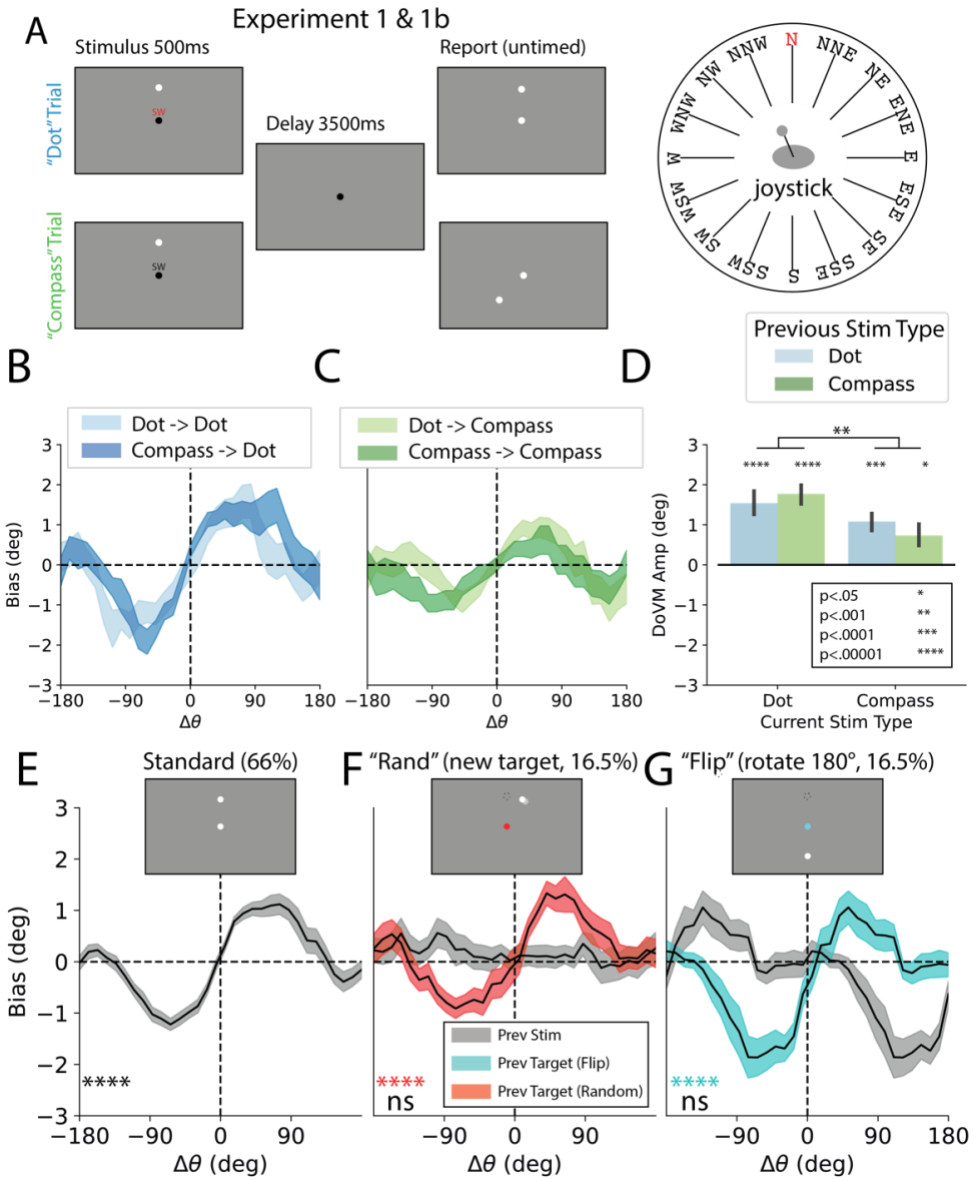


Figure 3-2 Serial dependence tracks previous responses.

A. Experiment 1 & 1b task schematic. Participants are presented with two stimuli on each trial: a concrete "dot" stimulus where the dot corresponds to the target location and a semantic "compass" stimulus were coordinates indicate the target location. Trial type is indicated by the color of the compass coordinates and were intermixed (Expt. 1) or blocked (Expt. 1b). After a delay, responses were made using a joystick on a videogame controller. B. Average bias on "dot" trials as a function of the previous of the previous angle shows clear serial dependence. Shading indicate SEM across participants. C. Same as B for "compass trials. D. Raw errors were parameterized on a single trial level and showed significant attractive biases in every condition. Biases elicited on "dot" trials were significantly stronger however. Error bars correspond to SEM. E. Average bias curve collapsing across all current and previous trial types (dot and compass) following "standard" responses. F. Average bias curve following "random" responses sorting trials by the relative angle of both the stimulus (gray) or the new target location (red). Responses show a strong attraction towards the previously reported location but not the previously encoded stimulus. G. Average bias curves following "flip" responses. Strong bias is seen by the previously reported location (teal) but not the originally encoded location (gray).

Serial dependence driven by previous response

In experiment 1, on 1/3rd of trials participants were presented with a cue at the end of the delay period that instructed them to not report the stimulus presented. Instead, they were to report either a location rotated 180° from the target location ("flip" trials) or disregard the remembered stimulus and instead move the response cursor to a new faintly illuminated location (direction independent of remembered location, "random" trials, Figure 3-2F-G). Only a single non-standard response cue (either 'flip' or 'random') was included in a given block of trials to reduce task complexity. Participants were slightly less accurate on ($p<1e-5$, Figure 3-9A) and slower on "flip" trials ($p=<1e-6$, Figure 3-9B).

This manipulation allows us to determine if the attraction towards the previously reported stimulus is driven by processes related to the encoding and maintenance of the target, or instead processes related to the the act of reporting it. As such, we evaluated the degree to which attractive biases depended on the previously remembered stimulus or on the report made. For simplicity, we only analyze trials where the current trial required a veridical response (e.g., not a "flip" or "random" trial and note that the ordering of trial type is independent of past stimuli). To maximize power, we collapse across all current and previous trial types (dot/compass) to get a single serial dependence measure following these different manipulations. Following standard responses, the previous stimulus and target location are equivalent and we observe an attractive serial dependence as reported previously (Figure 3-2E, amp=1.31±.13, $t(17)=9.7$, p=1.4E-8). Following "random" trials participants did not show a significant attraction towards the previous stimulus (amp =0.10±0.28, p=.73) and instead showed a significant attraction towards the new probed location (amp=1.36±0.20, $t(17)=6.68$, p=.0000029, Figure 3-2F). We observed a similar behavior on "flip" trials where participants were attracted towards the previous target

location (1.76±0.26, $t(17)=6.68$, p=.0000029) but not towards the previously encoded stimulus (-0.44±0.3, p=.18, Figure 3-2G).

These findings are surprising in light of previous stimulus-response manipulations which found attraction centered on the previous stimulus, not response when they were dissociated(Cicchini et al., 2017). However, they generally fit with work suggesting it is previous decisions (rather than the previously remembered item) that drives serial biases (Pascucci et al., 2019). Even when the stimulus is used in a transformed state (as in flip trials), the remembered location does not seem to induce any residual attractive bias (Figure 3-2F, gray). One interpretation of this result is that any serial bias towards the previously remembered stimulus is completely mediated by that memory being utilized during the previous response.

It is somewhat surprising that the "random" response is able to elicit an attractive bias. This condition is reminiscent of the 0s delay condition in previous studies (Bliss et al., 2017; Manassi et al., 2018) and work showing negligible impact of the previous delay period (Papadimitriou et al., 2015). However, this "random" condition in the present task doesn't require the inducing trial to utilize working or iconic memory as the target remains on the screen for the entire response period, leaving the potential drivers of the induced effect to either seeing/attending to the response/target or the act of manipulating the response controls. We next sought to determine the role of these respective components by manipulating the method of response and visual experience during the reporting period.

Serial dependence across response type

In experiment 1, serial dependence clearly tracked the reported location. This bias could be driven by a high-level representations of decisions, or alternatively be related to the specific motor actions taken (path of thumb) or the visual experience during the response itself. Prior

work has largely shown a dissociation between attractive serial dependence effects and motor responses, but these approaches have almost exclusively been in a 2AFC design where the standard response was already unrelated to stimulus identify per se (Feigin et al., 2021; Zhang & Alais, 2020; Zhang & Luo, 2023). We ran a follow up study to see if we could identify which components of the previous response gave rise to the attractive bias while maintaining the fundamental structure of our task. In experiment 2, the correlation between motor action and reported stimulus was abolished by utilizing a separate response mechanism with a random starting point (Figure 3-3A). Under these conditions serial dependance persists regardless of the current or previous stimulus type (Figure 3-3A-C, all amplitudes $p<.05$), suggesting a minimal role for pure motor hysteresis. In experiment 3, the stimulus was not reported but rather compared with a probe stimulus (CW/CCW judgment, Figure 3-3D, Figure S4C). We continued to observe attractive serial dependence on dot trials (Figure 3-3E, $p<.005$) but there was no significant effect on compass trials (Figure 3-3F, $p>.05$). However, when we collapsed across previous stimulus types, we did observe a significant attractive effect of compass trials ($p=0.013$) suggesting that - as in experiment 1 - the effect is smaller but not abolished for the semantically encoded compass stimuli. Notably, we also observed motor biases irrespective of the previous stimulus identity that were overall repulsive but also exhibited a "win-stay lose-switch" tendency (note that stimulus non-specific feedback was provided, see Methods) often seen in 2AFC designs (Figure 3-10E, (Abrahamyan et al., 2016; Akrami et al., 2018)). Together, experiments 2-3 demonstrate that the attractive effects towards the previous response cannot be reduced to either the motor action alone. Thus, biases likely operate on a more general representation of the previous decision irrespective of the method of report.

We separately examined the role of the visual experience of the response. While previous studies have typically tried to make the method of response divergent from the remembered stimulus, there is still typically substantial overlap in the feature of interest (e.g., orientation, (Ceylan et al., 2021)) or more generally between the visual experiences of the response per se. Thus, many effects attributed to the previous stimulus could easily also be explained by the visual experience of the response (which is highly correlated). To better control for this aspect, we sought to remove any potential visual confounds from our response period. Specifically, while the compass stimulus did not have a visual correlation with the target location, the response mechanism in experiments 1-2 (and experiment 3 to a lesser degree) elicited visual stimulation similar to the presentation of a dot stimulus. To avoid this factor, in experiment 4 we trained participants on using a response joystick without seeing their response on the screen (Figure 3-3G). Under these conditions we observed significant serial dependence in the compass followed by dot trial condition (Figure 3-3H, $p=.001$) but not in the other conditions (but note that attractive effects for compass stimuli do emerge from trials further back, Figure 3-12). Thus, the attractive effects we observe towards the previous stimulus cannot be solely attributed to motor activity (experiments 2&3), past memory maintenance (experiment 1, "random" response), or even visual experience (experiment 4, following compass stimuli). Notably however, across experiments attraction was significantly stronger on "dot" than "compass" trials (Figure S5, $t(89)=2.82$, $p=0.006$) potentially signaling sensory drive could still be a contributing factor.

Our analysis additionally revealed some notable effects worth reporting. First, in Experiment 2 we observed an attractive bias towards the starting location of the probe. Note that this bias was tuned (only attraction was seen when the starting position was within 90°) and that

participants were required to move the dial before advancing (Figure 3-10A, p=.0002). This bias could be a general 'laziness' tendency to under-rotate the probe, but more likely reflects a subtle visually induced bias towards the response probe onset. In experiment 3 we additionally observed a motor/response bias with participants reflecting a trend to alternate responses regardless of the current stimulus offset (β=-0.037±0.013, p=.006, Figure 3-10E). Participants were provided feedback after each response (correct/incorrect) and we additionally observed a "win-stay lose-switch" tendency (β=0.043±0.015, p=.005, (Abrahamyan et al., 2016; Akrami et al., 2018)). Thus, we found a coexistence of an attractive bias towards previously remembered stimuli with a repulsive bias towards past responses. Lastly, in both Experiments 2&3 we observed larger effects on trials with longer delays for dot but not compass stimuli (Figure 3-10 B,D). This finding matches previous findings for spatial working memory (Bliss et al., 2017; Papadimitriou et al., 2015) and suggests a role for either sensory uncertainty at the end of the trial or memory maintenance circuits in mediating our observed biases, at least for dot trials.

In sum, our results so far have been largely consistent with a "decisional level" bias as proposed in an account by Pascucci et al., 2019 with separate evidence that sensory and/or memory circuits play an important role. That said, we have yet to examine if any additional biases persist towards stimuli that were not attended or reported.

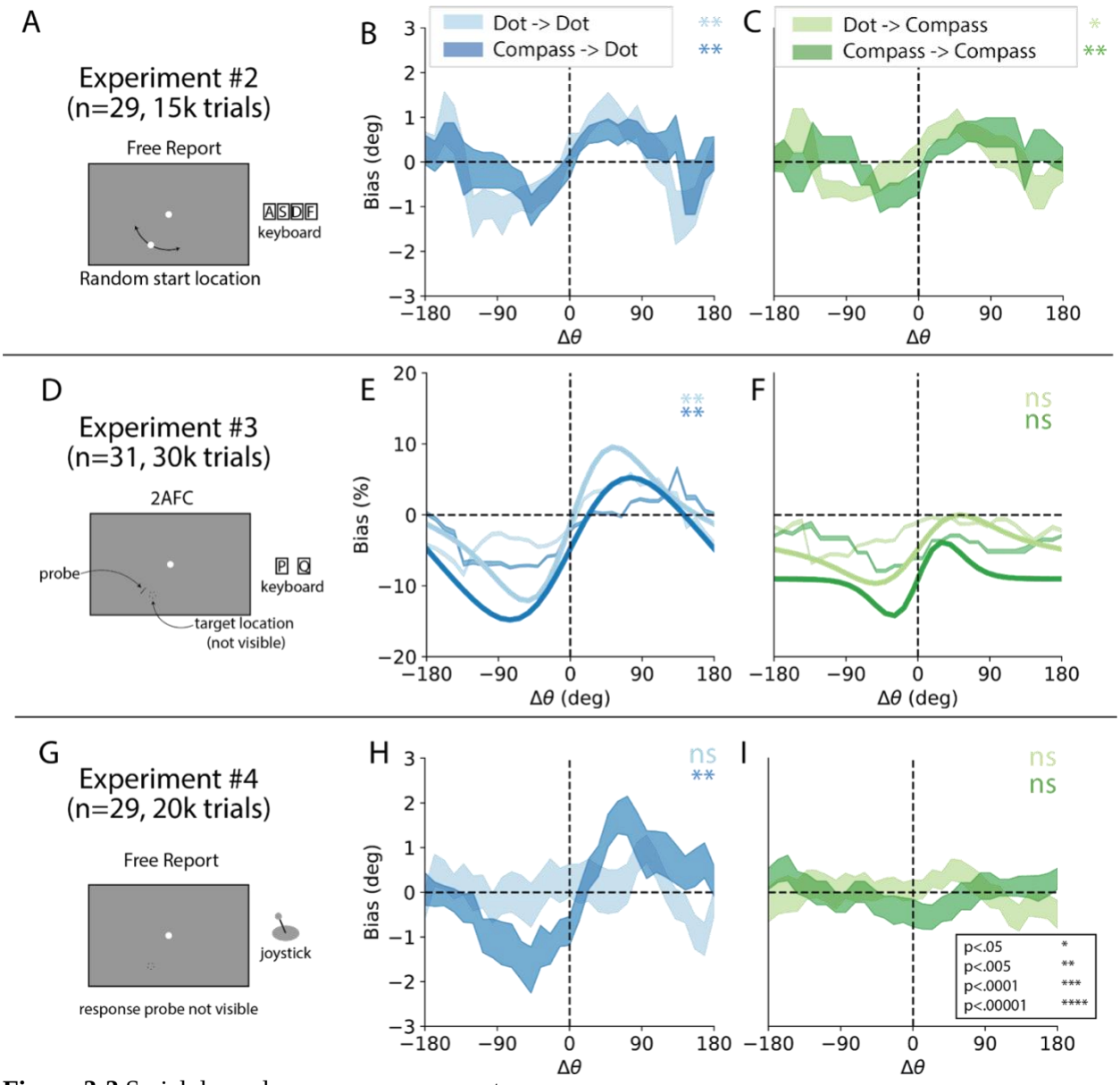


Figure 3-3 Serial dependence across response type.

A. Experiment 2 task schematic. Experiment 2 utilized a response using timed keyboard presses rather than a joystick such that motor action was no longer directly correlated with response. B-C. Bias curves for serial dependence across conditions. Shading indicates SEM across participants. All cross conditions are significant when parameterizing amplitude with a DoVM function. D: Experiment 3 task schematic. Participants judge whether a response probe is CW or CCW relative to the presented stimulus. This design eliminates any motor correlation with the target stimulus. E. Average % CCW responses as a function of the relative angle of the previous stimulus for "dot" trials. These bias curves suggest a significant attractive bias when parameterized. F. Same as E but for "compass" trials. No significant attractive bias was found in these two conditions. G. Task Schematic Experiment 4. Participants used a joystick to report the previous stimulus but were unable to see the response dot while they were responding. H. Participants exhibited clear attraction to previous "dot" stimuli but only following "compass" trials. I. Participants showed no attractive biases on "compass trials.

Biases towards unreported stimuli

In experiments 1-4, we show robust attraction towards the previous stimulus largely independent of how it was encoded or reported. When we disentangled the stimulus from the response experimentally or manipulated how the response was made, we consistently found an attraction towards whatever stimulus was reported. This is true even if the stimulus was never associated with a visual experience correlated with the stimulus (as in experiment 4 following compass trials). We followed up by asking whether we could observe any attraction to past stimuli when they were not explicitly attended or reported.

We first examined trials following "drop" trials. Experiments 2-4 featured drop trials on 20-33% of trials where the fixation point turned red at the end of the delay period. On these trials participants simply pressed a button to advance to the next trial (Figure 3-4A). We observed an attractive bias towards the previous stimulus following dot stimuli in Experiment 1b ($p=.008$, Figure 3-4B-C) but notably not following compass stimuli or in any other experiment. Note that owing to low trial counts/ subject for drop trials, for analyzing trials following drop trials we utilize a bootstrap approach pooling across participants (see Methods). It is worth noting that the drop condition in Experiment 1b was a much simpler design than our other tasks such that on dot blocks, participants could immediately prepare their motor plan towards the dot location. Thus, the presence of an attractive bias in this condition without a previous response could likely be related to movement preparation. Conversely, we observed a repulsive bias from unreported dot stimuli in experiment 4 in which no motor preparations could be made as the target stimulus was not known until the end of the delay period (Figure 3-4B,D). Thus, we find attraction towards a previously unreported stimulus only when a straightforward motor plan can be made early in the

trial and there are no intervening actions suggesting attractive biases even in the absence of responses may be related to response preparation.

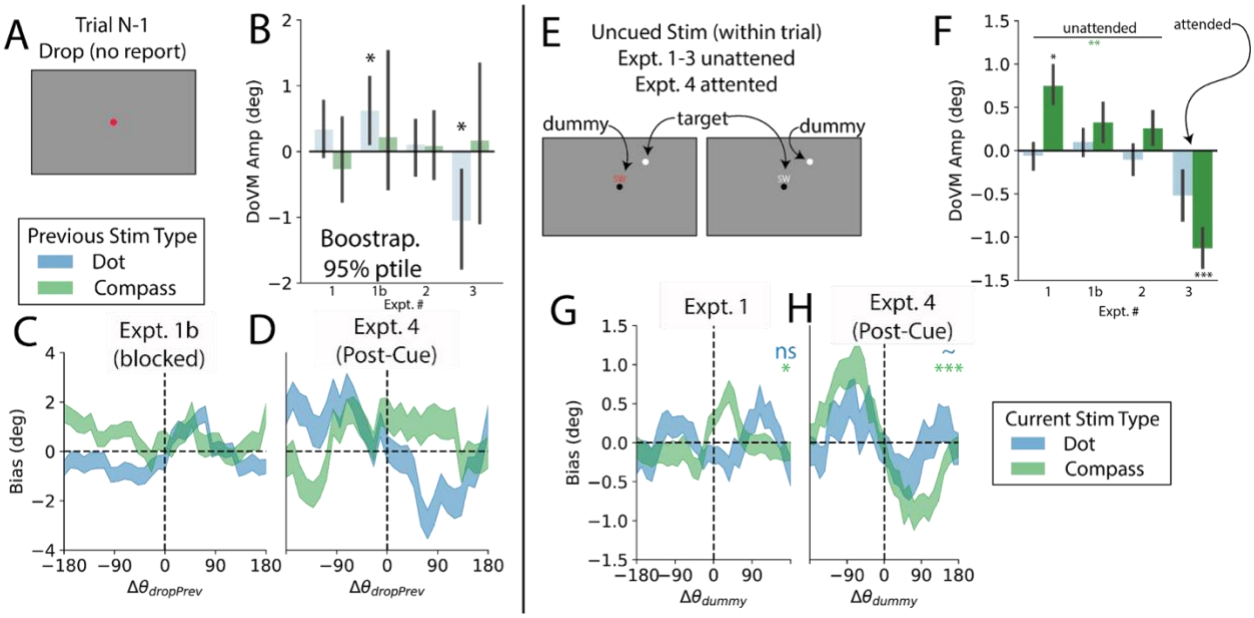


Figure 3-4 Serial dependence towards unreported stimuli.

A-D The influence of unreported stimuli from the previous trial ("drop" trails). A. Schematic of what participants see at the end of the delay on drop trials. They must press A or SPACE to advance to the next trial and are discouraged from making any movement by a warning message. B. Average biases across experiment and pervious stimulus type. Values are from bootstrapped participants and errorbars correspond to the 95 percentiles of bootstrapped distribution. C. Significant attraction was observed towards previously dropped dot stimuli in Expt. 1b. D. Significant repulsion was observed towards previously dropped dot stimuli in Expt. 4. E-H The influence of unreported stimuli within a trial. E. Schematic indicating "dummy" stimuli on "dot" and "compass" trials respectively. F. Average biases relative to dummy stimuli for "dot" or "compass" trials. Note that for green bars, the reported stimulus is "compass" and the inducing stimulus is "dot". Significant attractive biases are seen for compass trials both in Expt. 1 and when collapsing across Expts. 1-2 while repulsive biases are seen in Expt. 3. G. Bias curves for dummy stimuli in Expt. 1. H. Bias curves for dummy stimuli in Expt. 4.

We next examined a control present in all our experiments meant to match visual input across trials. On each trial in addition to the target stimulus, there was an unattended item (either the dot location on "Compass" trials, or vice versa) that was ignored on either a block (experiment 1b) or trial basis (experiments 1,2-4, Figure 3-2A, Figure 3-4E). These unattended "dummy" stimuli were neither attended not reported and so offer a useful marker on what

elements are sufficient or necessary to trigger an attractive bias. We first examined the influence of "dummy" stimuli on the current trial.

In experiment 1 we observed a significant attraction on compass trials to the unattended dot stimulus (amp = 0.82 ± 0.27, $t(18)=3.0$, p=.008) but apparently no impact of the unattended compass cue on dot trials (-0.08 ± 0.14, p=0.6, Figure 3-4F-G). This suggests that unattended and unreported stimuli can induce attractive biases in an automatic fashion. We found similar effects when pooling across experiments 1-2 (Figure 3-4F, $t(60)=3.2$, p=.002). This suggests that perceptual reports can be automatically biased towards past visual experiences even when those items are task irrelevant and encoded in a different format from.

Interestingly, in experiment 4 the target item was not cued until the end of the delay period such that both stimuli had to be held in memory. Under this condition, we observed repulsion of the items from each other (dot amp = -0.54 ± 0.31, $t(28) = 1.74$, p=0.09; compass amp = -1.13 ± 0.23, $t(28)=4.91$ p=.0004, Figure 3-4H). Thus, in stark opposition to serial biases, attention seems to prevent inter-item biases within a trial. More generally, we find while attractive biases towards unreported or unattended items are substantially weaker and less reliable than reported items on the previous trial, they are significantly more pronounced for low-level dot stimuli (relative to compass coordinates). This suggests that low level properties may be critical for some components of attractive serial dependence.

Semantic and spatial history traces have distinct time-constants

Up until now we have only examined the impact of the immediately preceding (N-1) stimulus. Past work has shown that history effects can extend back several trials and in some instances lead to repulsion from more distant stimuli (Braun et al., 2018; Collins, 2020; Fischer & Whitney, 2014; Fritsche et al., 2020). We extended our analysis of history effects back many

trials across all our studies and noted a consistent trend. Sensory dot stimuli showed a strong attractive bias on the first trial followed by a repulsive bias extending back several trials (Figure 3-5D). By contrast, the bias on compass trials is typically smaller on the first trial (Figure 3-5A, p=.007) but extends back much further in time (Figure 3-5B-D, p=.000095). To improve power, we additionally examined the collective influence of trials 2-5 back and observed a graded effect influenced by the stimulus type (Figure 3-5E). The same trend is observed across individual experiments (Figure 3-12). Thus, it appears that both the largest attractive (N-1) and largest repulsive (N=-2 through -5) biases are found when utilizing low-level sensory stimuli.

Finally, we examined the impact of this expanded history bias on both uncued (dummy) and unreported (drop) stimuli. We found in both cases that unreported dot stimuli elicited robust repulsive biases exclusively on other dot stimuli (uncued: Figure 3-5F, $t(76)=-4.01$, p=.0001 and drop: Figure 3-5G, $t(76)=-3.78$, p=.0003). This suggests the repulsive bias observed following distant is related to an automatic form of perceptual adaptation that is driven by low-level sensory representations.

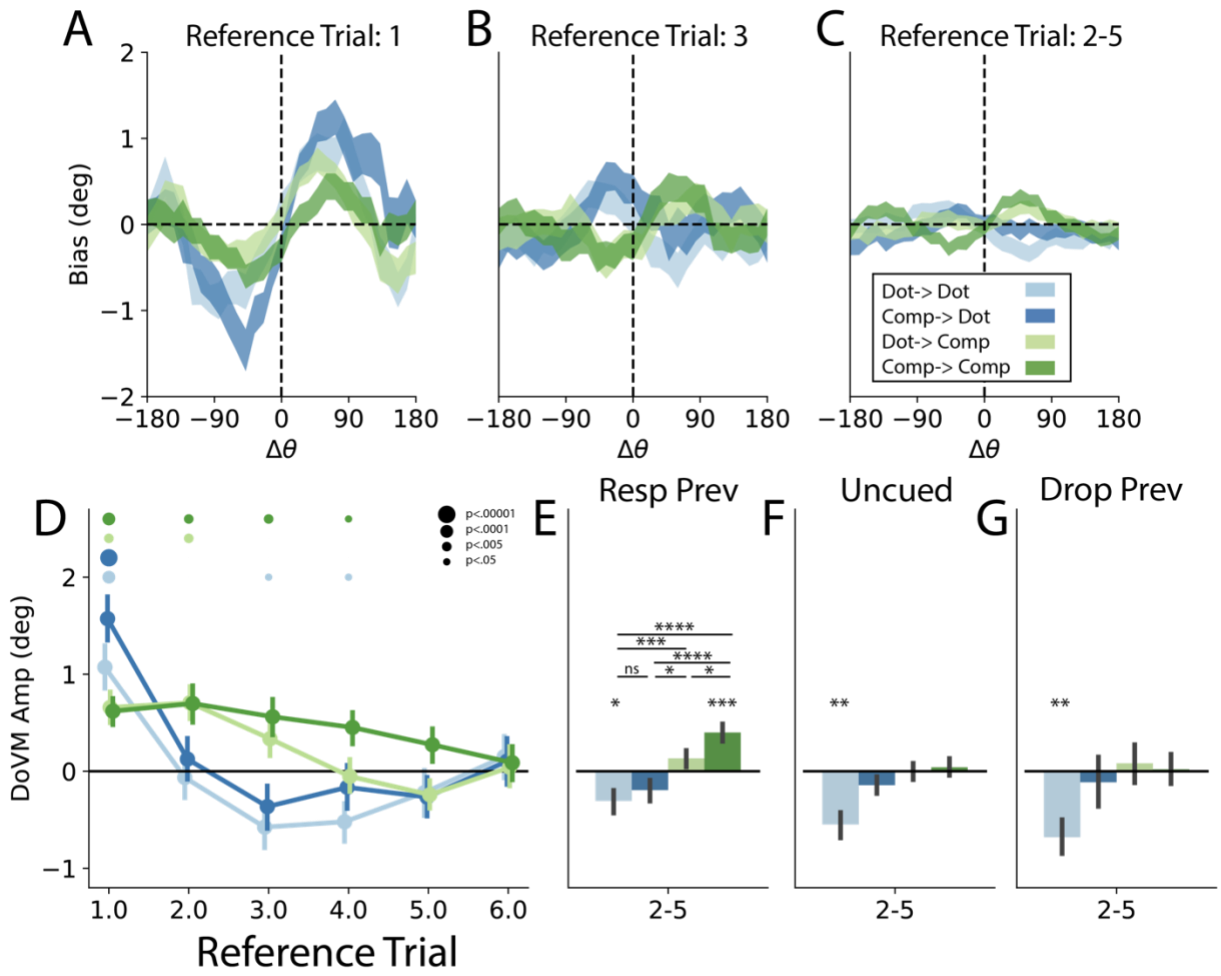


Figure 3-5 Timescales of serial dependence.

A. Serial dependence pooled across all tasks for all combinations of current and previous stimulus type for the immediately preceding (n-back=1) trial. B. Same as A for n-back=3. C. Same as A-B but pooling across NB=2-5. D. Bias amplitude across all task conditions for trials 1 through 6 back. Dots size above lines indicates 2-tailed significance test. E. Response bias pooling across NB=2-5 for all conditions. Repulsive biases are seen between dot trials while attractive biases are seen between compass trials. F. Same as D but previously uncued "dummy" stimuli. Dot-> dot stimuli show a significant repulsive bias. Note that for this panel only we have adjusted our color scheme such that the colors correspond to the previously unreported stimulus. G. Same as E-F but for previously attended but unreported stimuli on dot trials.

Discussion

We examined serial dependence across a range of stimulus types and response conditions. We found ample evidence supporting a bias operating at a post-perceptual or decisional level. First, serial dependence persisted regardless of whether the stimulus was

encoded in a visual (dot) or semantic (compass) manner or if consecutive stimuli were of the same type. In addition, we found that attraction primarily followed the previous response rather than the previously encoded and remembered location (Figure 3-2 F-G). This response-driven attraction was observed regardless of motor (Figure 3-3A-F) or visual (Figure 3-3 D-I) correspondence between stimuli and response. Thus, a principal finding of our study is that serial biases emerge from a decisional level representation of task relevant information that are not solely attributable to low-level sensory or motor components (Braun et al., 2018; Feigin et al., 2021; Moon & Kwon, 2022; Pascucci et al., 2019; Urai & Donner, 2022; Zhang & Alais, 2020; Zhang & Luo, 2023).

That said, we also find evidence for a critical role of low-level stimulus properties. Most centrally, "dot" trials – where a stimulus was physically present – are more biased by past stimuli than "compass trials" both within and across tasks (average Δamp = 0.56 ± 0.20, $t(89)=2.823$, $p=0.0059$, Figure 3-3, 3-11). This is not explained by a change in overall uncertainty between the two stimulus classes (Figure 3-7C) and instead suggests part of the attractive bias we observe is implemented during the feedforward sweep of sensory encoding. Dot stimuli additionally induce long lasting repulsive sensory adaptation (Figure 3-5E). Interestingly, this bias appears to be automatic as it is equally present following uncued or "dropped" stimuli on previous trials (Figure 3-5F-G), matching previous findings (Pascucci et al., 2019) and aligning with the related finding that repulsive peripheral bumps were not impacted by attention (Fritsche & de Lange, 2019). Thus, the stronger (N-1) attractive bias seen on dot trials is additionally overcoming this repulsive tendency (note the difference between N-1 compass->dot and dot->dot trials, Figure 3-5D). In addition, we observe within trial attraction to unattended dot stimuli, suggesting low-level stimulation at encoding, regardless of attention, can shape perceptual reports in our

paradigm. By contrast, compass stimuli exhibit long lasting attractive biases and are minimally impacted by the repulsive effects of past stimuli. Thus, in addition to the presence of decisional level biases, we also see strong evidence for short lived attractive biases emerging as the result of sensory encoding of low-level features (Fornaciai & Park, 2018; Liberman et al., 2014, 2014; Manassi et al., 2018).

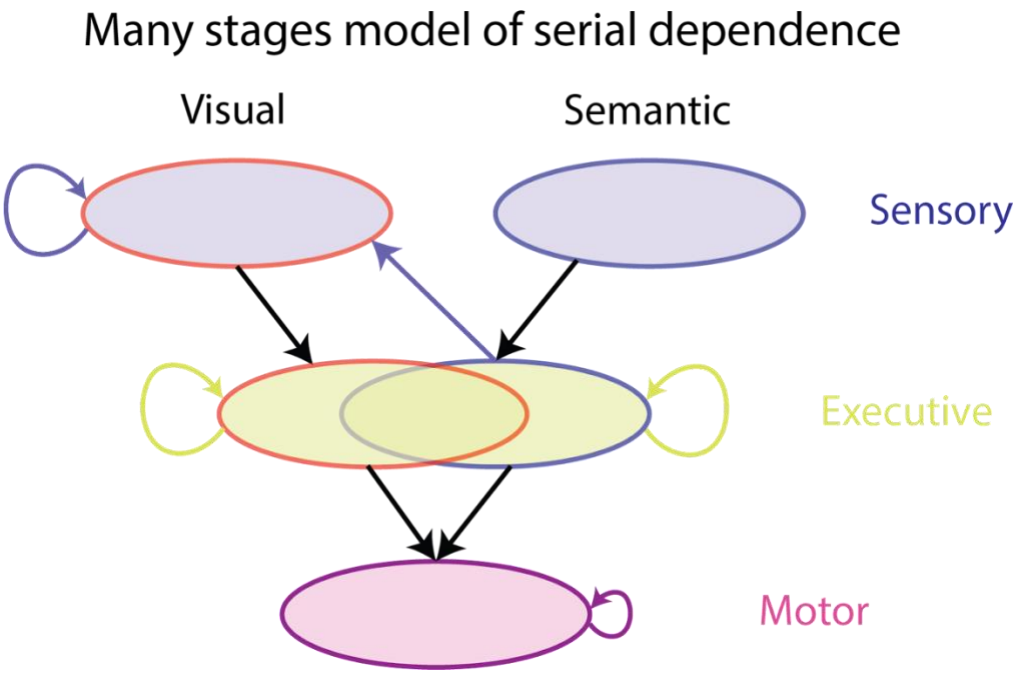


Figure 3-6: Conceptual model.

Prior studies have generally attributed serial dependence effects to either sensory (purple) or post-sensory functions (eg. decisions or working memory maintenance, yellow). In either case, history information somehow feeds into processing at the level. In the many stages account, we instead propose that history integration occurs across several levels of information processing with signs and magnitudes of biases at different stages depending on timing, uncertainty and other task contingencies.

Thus, these data collectively support the hypothesis that serial dependance, as typically measured in the lab, arises from multiple interacting levels of the stimulus-response loop (Figure 3-6). Such an account is consistent with the findings in our study as well as the diversity of findings in other neural and behavioral studies of serial dependence. Notably, prior work has disentangled the influence of previous stimuli, choices, and responses and found simultaneous contributions of all elements (Braun et al., 2018; Zhang & Luo, 2023). It follows that different

conditions will elicit varying forms of biases. For instance, the uncertainty associated with a dot stimulus will emerge from both encoding, memory maintenance, and the response and thus serial dependencies (or less 'intentional' neural bleed over; Luekmann et al., 2016) are apt to operate at both levels. Separately, on compass trials uncertainty is more likely to emerge at a decisional stage and thus history biases are likely to act at a later in the processing stream. Under more general circumstances, the relative amounts of stimulus, choice, and motor uncertainty could influence representational hysteresis at each stage and in turn impact responses and measured biases in non-additive ways.

A related but distinct account was recently proposed by Cicchini and colleges (Cicchini et al., 2021). In their paradigm, they measured both attractive serial dependence across stimuli and repulsive biases due to the surround tilt illusion. Critically, they found that biases introduced by spatial context influenced how the present stimulus influenced future responses, but this same spatial context did not influence serial biases on the current trial. They thus argued that serial dependencies were *induced* spatial context was incorporated and then fed-back to influence representations at an early stage on subsequent trials. This account is consistent with our findings of attraction generally tracking previous perceptual reports rather than just the previously seen dot stimulus (Figure 3-2 F-G, Figure 3-3). However, it is not obvious how an early bias would operate on compass stimuli. One possibility is that compass coordinates are recoded into a sensory format during our task and that this sensory representation supports biases on subsequent trials. This type of recoding from an abstract-to-sensory format has recently been found in the context of long-term associative memory tasks (Sutterer et al., 2019; Vo et al., 2022) . It is also consistent with our findings of inter-trial interference between stimulus types, including from compass stimuli when they are attended (Figure 3-4F, Expt. 4). That said, this recoding would

require much slower feedback induced activity in sensory cortex and would completely miss the feedforward sweep which could be critical for inducing these biases. This difference in how stimuli are encoded into sensory representations may explain the relative difference in bias strength between dot and compass stimuli. Critically, while the proposed decisional to sensory account of serial dependence is consistent with many of our findings, it is unable to account for the larger biases seen for longer delay periods (Figure 3-10 B,D; (Bliss et al., 2017; Papadimitriou et al., 2015) Bliss et al. 2016; Papadimitriou et al., 2015)) nor the failure of observed biases following drop trials in most experiments. To do so, one must also consider history biases that act on post sensory representations.

We are not the first group to propose that different serial dependence findings arise from distinct mechanisms. Early work in this area readily noted that the ubiquity of attractive effects implied a general mechanism, perhaps fundamental to all neural processing (Burr & Cicchini, 2014; Fritsche & de Lange, 2019; Kiyonaga et al., 2017). That said, most models of serial dependence rely on action at a single level of information processing. The notable exceptions are models assuming repulsive adaptation is implemented at encoding and attractive biases are due Bayesian integration at a later stage (Moon & Kwon, 2022; Pascucci et al., 2019; Sheehan & Serences, 2022) see also (Papadimitriou et al., 2016) but here still, assimilative biases are only implemented at a single, later stage. In contrast, our "many stages" model explicitly assumes that stability priors may be present at all stages of information processing – depending on stimulus/behavioral requirements – and that any observed serial dependance will thus arise from biases implemented at multiple stages.

One assumption of our proposed model is that assimilative biases can operate on stimulus representations at an early sensory stage. While in line with some behavioral data, neural

evidence is generally lacking, particularly at the time scales over which behavioral serial dependence is observed. Notably, visual persistence has been shown to show concurrent attractive biases in both perception (Di Lollo & Dixon, 1988) and early visual areas (Benucci et al., 2009; Coltheart, 1980; Duysens et al., 1985) but these effects have been shown to only persist for hundreds of milliseconds, 2 orders of magnitude shorter than typical serial dependence effects. Instead, the most consistent findings to emerge are those showing a reinstantiation of the previous stimulus with the emergence of the current trial (Bae & Luck, 2019; Ranieri et al., 2022; Zhang & Luo, 2023), but notably in an inverted or orthogonal code relative to the currently encoded stimulus (Hajonides et al., 2023; Luo & Collins, 2023). Thus, it may be that perceptual serial dependence affects are instead due to sensory readout (Sheehan & Serences, 2022) either utilizing the lingering sensory representation of the previous stimulus (Zhang & Luo, 2023) or a separate store that encodes stimulus-specific information as a function of trial history (Akrami et al., 2018).

In addition to sensory and decision-level biases, motor biases have been shown to be present in perceptual report or judgment tasks. Interestingly, these follows typically follow a repulsive pattern (Pape & Siegel, 2016; Zhang & Alais, 2020; Zhang & Luo, 2023) but with large variability across participants (Urai & Donner, 2022) and occasionally more complex dynamics such as “win-stay, lose-switch” (Abrahamyan et al., 2016; Akrami et al., 2018; Lueckmann et al., 2018). While most of these studies have utilized a 2AFC design, a challenging coincident timing task showed participants a tendency to respond too slowly following slow trials and vice versa (Neto & Bartels, 2021). Critically, applying TMS to a motor area (but not a visual area or vertex) between trials significantly reduced inter-trial dependency, pointing to a direct role of motor cortex and inter-trial dependencies. Related work has even found a direct

mechanism for motor response alternation (beta rebound) that seems to track motor biases even on a single trial level (Pape & Siegel, 2016; Urai & Donner, 2022). Thus, we may have substantially more and more consistent evidence for direct neural instantiation of motor biases than ones operating at perceptual or decisional levels suggesting it may be a very important pathway for maintaining stable behavior across time.

In this study we propose that serial dependence arises due to general priors for temporal stability across perception, choices, and motor actions. Across a series of experiments, our results are inconsistent with a single origin of serial dependence and instead suggest that it emerges from multiple sources (e.g. sensory and decisional). We secondarily found that the timescales of serial dependence differ depending on the stimulus type, suggesting longer timescale (and presumably later) representations play a larger role for semantic over visually encoded stimuli.

Methods

Experimental Design, General

All participants received informed consent as to the scope of the study and were compensated monetarily for their participation as approved by the UC San Diego IRB (approval number 180067). Experiments were performed in dim private rooms with participants seated with a chin rest to stabilize viewing 50 cm from a 39 by 29 cm CRT monitor (1600 × 1200 px) with a visual angle of 42.6° (screen width). Participants completed experiments over the course of 1-3 90-minute sessions.

In all experiments, participants were required to remember and report on the angle of an encoded stimulus. This stimulus could either be a spatial cue, a black circle (0.75 degrees of visual angle, dva) presented at the target location, or a semantic cue, 1-3 letter string presented at

fixation corresponding to 1 of 16 possible compass coordinates {'N', 'NNE', 'NE', …, 'NNW'} (Figure 3-2A). Participants were cued as to which stimulus was relevant on a trial or block basis. Responses were made in a continuous report fashion by moving a dot along a ring to match the remembered stimulus using either a joystick or the keyboard. Both dot stimuli and responses were centered on an imaginary circle with a diameter of 18 dva.

All participants completed practice blocks of the experiment with various aids to make the task easier (e.g., fewer conditions, shorter delay period) until they reached a performance criterion (generally <10° |error|). Trials from these practice blocks were not subsequently analyzed. We excluded participants from subsequent analysis if they completed <200 trials or had an average unsigned error of >25° (except in experiment 3 which had a 2AFC design).

Experiment 1

In experiment 1, participants (n=22, Figure 3-2A) were exogenously cued to the relevant feature by the color of the compass stimulus at encoding (500ms, red= attend dot, white = attend compass). After a delay period (3500 ms) they used the left joystick on a video game controller (Xbox 360) to report the target feature of the attended stimulus (untimed) and pressed "A" to advance to the next trial. To minimize the role of hysteresis in the position of the joystick between trials, the next trial would not initiate until the joystick was moved back to the center. The task was completed in blocks of 96 trials, 2 participants only completed two blocks and 1 participant had an average error exceeding our threshold and were excluded from subsequent analyses. The remaining n=19 participants completed 8.46±1.9 [6, 12] blocks of trials.

On 1/3rd of trials participants were cued to not report the remembered stimulus but either a new, random location faintly illuminated ('random") or rotated 180° from the target stimulus

("flip"). The type of response was indicated by the color of the fixation cue at the end of the delay period and was restricted to only a single type of drop trial for a given block of trials.

Experiment 1b

Experiment 1b (n=14) was a simpler version of experiment 1 where dot/compass trials were presented in a blocked manner (64 trials/block). Additionally, for 9 participants we had participants withhold responses on 25% of trials ("drop") with participants simply pressed "A" to advance. One participant was excluded from future analysis as their average error was >20° and one was excluded for completing <200 trials leaving a total n=12 who completed 10.8±5.3 [4, 20] blocks of trials.

Experiment 2

Experiment 2 (n=33, Figure 3-3A) was similar to experiment 1 except 1) delay time varied across trials (2000 or 5000ms), 2) ITI varied across trials (1000 or 4000ms) 3) responses were made using the keyboard to rotate a dot randomly positioned on a ring, and 4) For a subset of participants (n=17) on ~20% of ("drop") trials participants were not presented with a probe stimulus but instead were instructed to choose either report randomly when the fixation point turned red. The response dot was controlled using the ASDF keys which rotated the dot CW/CCW with course and fine adjustment. Due to a bug in the code, delay times were not randomized but rather alternated across trials (2000 ms always followed by 5000ms) for some (XX) sessions (% of total trials). Because of this, any analysis of the role of current and previous trial delay time is partially co-dependent. Four participants were excluded for having an average error >25° leaving a total n=29.

Experiment 3

Experiment 3 (n=21, Figure 3-3D) had a similar encoding period to experiments 1-2. Like experiment 2, delay (2000 of 5000ms] and ITI (1000 or 2000ms) duration varied randomly across trials. At the end of the delay period, participants were presented with a 2AC probe, a thin white line (length 3.6 dva) was presented between within 12° CW/CCW of the remembered location and participants reported whether that line was CW/CCW using a button press ("P"-CCW, "Q"-CW). The average offset was adjusted for each participant to target (80% accuracy, 0.79 ± 0.01 achieved). For each participant, 3 different offset values were used that were closer and further than the thresholded offset. On ~10% of all trials, participants were not presented with a probe stimulus but instead were instructed to choose press either report button randomly when the fixation point turned red. Feedback was given following every response by turning the fixation point green (correct) or red (incorrect) for 250ms before initiating the inter trial interval.

Experiment 4

Experiment 4 (n=29, Figure 3-3G) was similar to experiment 1 with but the relevant stimulus being revealed at the end of the delay period with a post-cue. The fixation point would change one of two colors (yellow/purple) at the end of the initial delay period (3500ms) indicating whether the compass or dot stimulus should be reported. Participants would then report the target stimulus using the joystick. Critically, responses in this version were blind such that they had to use a remembered correspondence between the joystick and the response location. Participants had to reach threshold performance (<15° error) before advancing to the main experiment and additionally had their response be made 'visible' approximately every 10 trials to minimize drift.

Removing systematic biases

Responses for both stimulus types exhibited systematic biases which we did not want to shape our measures of history bias or overall performance. We thus fit a 12-parameter Fourier like function to each participants errors separately for our two stimulus conditions (dot and compass) and performed subsequent analyses on the residualized errors. We have previously described a similar procedure for residualizing errors from an orientation report task (Sheehan & Serences, 2023) and include the equation here:

$$f(\theta;\ a_1,\ldots,a_N)\ = \sum_{n=1}^{N} g(\theta; n, a_n) \quad [33]$$

$$g(\theta; n, a_n) = \begin{cases} sin\left(\frac{1}{2}\theta n\right); & n \equiv even \\ cos\left(\frac{1}{2}\theta(n+1)\right);\ otherwise \end{cases} \quad [34]$$

where θ corresponds to the presented angle for a given trial in radians [-π, π].

Quantifying serial dependence

To measure the magnitude, shape and assess the significance of serial biases, we parameterized participants errors as function of the relative angle between the current trial's stimulus and an inducer. This inducer is typically a previously reported stimulus, eg. $\Delta\theta_{N-1} = \theta_{N-1} - \theta_N$ where θ_{N-1} corresponds to the presented stimulus on the previous trial. In this way, positive values of $\Delta\theta$ correspond to inducers that are CW relative to the current stimulus being reported, matching convention. Note that for all operations on circular variables, we imply that resulting angles are wrapped between [-π, π] radians. We can separately examine the impact of other inducers including other stimuli further back in time θ_{N-x}, previous response targets when they

differed from previous stimuli (eg. following “random” or flip trials), stimuli that were attended but not reported, and distracting “dummy” stimuli on presented on the same trial. In any case we examined serial dependence in two manners. In the first, we took a sliding circular mean (window ±32°) of errors as a function of the relative angle of the inducing stimulus. This sliding mean was used exclusively for visualization purposes. We separately fit a derivative of von Mises function (DoVM) to parameterize the shape and magnitude of any observed bias. The basic form of the DoVM function is:

$$doVM(x;\ a, w) = a\ wsin(x)\ exp(w\ cos(x))\ /\ (zI_0(w)) \quad [35]$$

with amplitude, *a*, width, *w*, and where *x* corresponds to the relative orientation of the reference stimulus, Δθ (Sadil et al., 2021). *z* is a normalizing constant such that the amplitude, *a*, corresponds to the height of the resulting function.

As many of our participants seemed to exhibit systematic biases (e.g. on average responding CCW relative to the true stimulus value), we included an extra offset parameter to better model response errors. Thus, our final model of errors has 3 free parameters and is:

$$\hat{E} = f(x; a, w, offset) = doVM(x; a, w) + offset \quad [36]$$

Optimal parameters were estimated using a bounded minimization algorithm (scipy.optimize.minimize), restricting amplitude ([-15°,15°], initial=0), precision ([.1, 5], initial=1), and offset ([-10°,10°], initial=0). This initialization was consistently able to converge to a solution.

This procedure was modified for experiment 3 which utilized a 2AFC rather than a continuous report design. Here we coded the responses to -1 for CW responses and +1 for CCW responses. As reports indicate the perceived relative angle of the probe stimulus, a more positive overall value would correspond to a CW shift in the perception of the original stimulus. We then fit a modified DoVM function to try and estimate the average proportion of CW/CCW responses with the bounds of amplitude (±0.5) and offset (±0.1) adjusted to fit reasonable ranges for response biases. We multiplied the resulting amplitude coefficient by 100 to get report any serial dependence as a % bias.

For analyzing biases across multiple previous trials, we concatenate the errors and relative orientation differences for each offset included and then fit our DoVM model in the same manner.

Statistics

For most analyses, and whenever not specified otherwise, we fit models to subsets of trials within individual participants. When measuring serial dependence or other subject specific measures, we exclude trials on which no response was made (drop trials) or where the unsigned error was >30°. As we were primarily interested in quantifying the strength and general presence of serial dependence, we evaluated the fit amplitudes for a given condition using 1 sample 2-tailed t-tests on the fit amplitude parameters. To ensure observed effects were not an artifact of our task design or analysis procedure we repeated the same analyses after shuffling trial order.

For analyzing responses follow drop-trials (which have low trial counts) or data from experiment 3 (2AFC design) our ability to accurately fit models to individual participants was severely limited. To combat this limitation, we pooled trials across participants are performed a bootstrapped analysis where we randomly resampled all trials with replacement 1000 times and

used the resulting distribution to estimate p-values non-parametrically. For a given bootstrapped distribution, the 2x the lower proportion of trials above or below zero corresponds to the 2-tailed p-value. We repeated this procedure using shuffled trial orders to ensure our results were not an artifact of our analysis procedure.

All statistical tests are 2-tailed t-tests unless otherwise specified. We do not correct for multiple comparisons when we simultaneously display tests across experiments as such an approach would be drastically more conservative in estimating effect sizes than is the norm in this literature. When we fit linear models with multiple factors (as we do for examining the influence of both the current and previous stimulus type on bias amplitude, as well as for examining the influence of delay period across stimulus type in experiment 3, we used a mixed effects linear model including subject ID as a random factor). We report the coefficient and p-values for the factors of interest.

Acknowledgments: Chapter 3

Thanks to Kirsten Adam for providing incredibly helpful feedback on the manuscript.

Chapter 3, in full, is a reprint of the material under preparation. Sheehan, Timothy C.; Carfano, Ben; Richmond, Dianthe; Serences, John T. The dissertation/thesis author was the primary investigator and author of this paper.

Works Cited

Abrahamyan, A., Silva, L. L., Dakin, S. C., Carandini, M., & Gardner, J. L. (2016). Adaptable history biases in human perceptual decisions. *Proceedings of the National Academy of Sciences*, *113*(25), E3548–E3557. https://doi.org/10.1073/pnas.1518786113

Akrami, A., Kopec, C. D., Diamond, M. E., & Brody, C. D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, *554*(7692), 368–372. https://doi.org/10.1038/nature25510

Bae, G.-Y., & Luck, S. J. (2019). Reactivation of Previous Experiences in a Working Memory Task. *Psychological Science*, *30*(4), 587–595. https://doi.org/10.1177/0956797619830398

Bae, G.-Y., & Luck, S. J. (2020). Serial dependence in vision: Merely encoding the previous-trial target is not enough. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-019-01678-7

Barbosa, J., Stein, H., Martinez, R. L., Galan-Gadea, A., Li, S., Dalmau, J., Adam, K. C. S., Valls-Solé, J., Constantinidis, C., & Compte, A. (2020). Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. *Nature Neuroscience*, *23*(8), 1016–1024. https://doi.org/10.1038/s41593-020-0644-4

Benucci, A., Ringach, D. L., & Carandini, M. (2009). Coding of stimulus sequences by population responses in visual cortex. *Nature Neuroscience*, *12*(10), 1317–1324. https://doi.org/10.1038/nn.2398

Benucci, A., Saleem, A. B., & Carandini, M. (2013). Adaptation maintains population homeostasis in primary visual cortex. *Nature Neuroscience*, *16*(6), 724–729. https://doi.org/10.1038/nn.3382

Blakemore, C., & Cooper, G. F. (1970). Development of the Brain depends on the Visual Environment. *Nature*, *228*(5270), 477–478. https://doi.org/10.1038/228477a0

Bliss, D. P., & D'Esposito, M. (2017). Synaptic augmentation in a cortical circuit model reproduces serial dependence in visual working memory. *PLOS ONE*, *12*(12), e0188927. https://doi.org/10.1371/journal.pone.0188927

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports*, *7*(1), 14739. https://doi.org/10.1038/s41598-017-15199-7

Braun, A., Urai, A. E., & Donner, T. H. (2018). Adaptive History Biases Result from Confidence-Weighted Accumulation of past Choices. *Journal of Neuroscience*, *38*(10), 2418–2429. https://doi.org/10.1523/JNEUROSCI.2189-17.2017

Burr, D., & Cicchini, G. M. (2014). Vision: Efficient Adaptive Coding. *Current Biology*, *24*(22), R1096–R1098. https://doi.org/10.1016/j.cub.2014.10.002

Ceylan, G., Herzog, M. H., & Pascucci, D. (2021). Serial dependence does not originate from low-level visual processing. *Cognition*, *212*, 104709. https://doi.org/10.1016/j.cognition.2021.104709

Cicchini, G. M., Benedetto, A., & Burr, D. C. (2021). Perceptual history propagates down to early levels of sensory analysis. *Current Biology*, *31*(6), 1245-1250.e2. https://doi.org/10.1016/j.cub.2020.12.004

Cicchini, G. M., & Burr, D. C. (2018). Serial effects are optimal. *Behavioral and Brain Sciences*, *41*, e229. https://doi.org/10.1017/S0140525X18001395

Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. *Journal of Vision*, *17*(14), 6. https://doi.org/10.1167/17.14.6

Clifford, C. W. G. (2014). The tilt illusion: Phenomenology and functional implications. *Vision Research*, *104*, 3–11. https://doi.org/10.1016/j.visres.2014.06.009

Collins, T. (2020). Serial dependence alters perceived object appearance. *Journal of Vision*, *20*(13), 9. https://doi.org/10.1167/jov.20.13.9

Collins, T. (2022). Serial dependence occurs at the level of both features and integrated object representations. *Journal of Experimental Psychology: General*, *151*, 1821–1832. https://doi.org/10.1037/xge0001159

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, *27*(3), 183–228. https://doi.org/10.3758/BF03204258

Deneubourg, J.-L., Aron, S., Goss, S., & Pasteels, J. M. (1990). The self-organizing exploratory pattern of the argentine ant. *Journal of Insect Behavior*, *3*(2), 159–168. https://doi.org/10.1007/BF01417909

Di Lollo, V., & Dixon, P. (1988). Two forms of persistence in visual information processing. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 671–681. https://doi.org/10.1037/0096-1523.14.4.671

Dong, D. W., & Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems*, *6*(3), 345–358. https://doi.org/10.1088/0954-898X_6_3_003

Duysens, J., Orban, G. A., Cremieux, J., & Maes, H. (1985). Visual cortical correlates of visible persistence. *Vision Research*, *25*(2), 171–178. https://doi.org/10.1016/0042-6989(85)90110-5

Feigin, H., Baror, S., Bar, M., & Zaidel, A. (2021). Perceptual decisions are biased toward relevant prior choices. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-020-80128-0

Felsen, G., Touryan, J., & Dan, Y. (2005). Contextual modulation of orientation tuning contributes to efficient processing of natural stimuli. *Network: Computation in Neural Systems*, *16*(2–3), 139–149. https://doi.org/10.1080/09548980500463347

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743. https://doi.org/10.1038/nn.3689

Fornaciai, M., & Park, J. (2018). Attractive Serial Dependence in the Absence of an Explicit Task. *Psychological Science*, *29*(3), 437–446. https://doi.org/10.1177/0956797617737385

Fornaciai, M., & Park, J. (2019). Serial dependence generalizes across different stimulus formats, but not different sensory modalities. *Vision Research*, *160*, 108–115. https://doi.org/10.1016/j.visres.2019.04.011

Fritsche, M., & de Lange, F. P. (2019). Reference repulsion is not a perceptual illusion. *Cognition*, *184*, 107–118. https://doi.org/10.1016/j.cognition.2018.12.010

Fritsche, M., Spaak, E., & de Lange, F. P. (2020). *A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.01.22.915553

Hajonides, J. E., Ede, F. van, Stokes, M. G., Nobre, A. C., & Myers, N. E. (2023). Multiple and Dissociable Effects of Sensory History on Working-Memory Performance. *Journal of Neuroscience*, *43*(15), 2730–2740. https://doi.org/10.1523/JNEUROSCI.1200-22.2023

Hays, G. C., Bastian, T., Doyle, T. K., Fossette, S., Gleiss, A. C., Gravenor, M. B., Hobson, V. J., Humphries, N. E., Lilley, M. K. S., Pade, N. G., & Sims, D. W. (2011). High activity and Lévy searches: Jellyfish can search the water column like fish. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1728), 465–473. https://doi.org/10.1098/rspb.2011.0978

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial Dependence across Perception, Attention, and Memory. *Trends in Cognitive Sciences*, *21*(7), 493–497. https://doi.org/10.1016/j.tics.2017.04.011

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial Dependence in the Perception of Faces. *Current Biology*, *24*(21), 2569–2574. https://doi.org/10.1016/j.cub.2014.09.025

Lueckmann, J.-M., Macke, J. H., & Nienborg, H. (2018). Can Serial Dependencies in Choices and Neural Activity Explain Choice Probabilities? *Journal of Neuroscience*, *38*(14), 3495–3506. https://doi.org/10.1523/JNEUROSCI.2225-17.2018

Luo, J., & Collins, T. (2023). The representational similarity between visual perception and recent perceptual history. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.2068-22.2023

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in visual working memory. *Memory & Cognition*, *36*(1), 43–52. https://doi.org/10.3758/MC.36.1.43

Manassi, M., Liberman, A., Kosovicheva, A., Zhang, K., & Whitney, D. (2018). Serial dependence in position occurs at the time of perception. *Psychonomic Bulletin & Review*, *25*(6), 2245–2253. https://doi.org/10.3758/s13423-018-1454-5

Mayer, M. J. (1977). Development of anisotropy in late childhood. *Vision Research*, *17*(6), 703–710. https://doi.org/10.1016/S0042-6989(77)80006-0

Moon, J., & Kwon, O.-S. (2022). *Dissecting the effects of adaptive encoding and predictive inference on a single perceptual estimation* (p. 2022.02.24.481765). bioRxiv. https://doi.org/10.1101/2022.02.24.481765

Murai, Y., & Whitney, D. (2021). Serial dependence revealed in history-dependent perceptual templates. *Current Biology: CB*, *31*(14), 3185-3191.e3. https://doi.org/10.1016/j.cub.2021.05.006

Neto, R. M. de A., & Bartels, A. (2021). Disrupting Short-Term Memory Maintenance in Premotor Cortex Affects Serial Dependence in Visuomotor Integration. *Journal of Neuroscience*, *41*(45), 9392–9402. https://doi.org/10.1523/JNEUROSCI.0380-21.2021

Papadimitriou, C., Ferdoash, A., & Snyder, L. H. (2015). Ghosts in the machine: Memory interference from the previous trial. *Journal of Neurophysiology*, *113*(2), 567–577. https://doi.org/10.1152/jn.00402.2014

Papadimitriou, C., White, R. L., & Snyder, L. H. (2016). Ghosts in the Machine II: Neural Correlates of Memory Interference from the Previous Trial. *Cerebral Cortex*, bhw106. https://doi.org/10.1093/cercor/bhw106

Pape, A.-A., & Siegel, M. (2016). Motor cortex activity predicts response alternation during sensorimotor decisions. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms13098

Pascucci, D., Mancuso, G., Santandrea, E., Della Libera, C., Plomp, G., & Chelazzi, L. (2019). Laws of concatenated perception: Vision goes for novelty, decisions for perseverance. *PLOS Biology*, *17*(3), e3000144. https://doi.org/10.1371/journal.pbio.3000144

Ranieri, G., Benedetto, A., Ho, H. T., Burr, D. C., & Morrone, M. C. (2022). Evidence of Serial Dependence from Decoding of Visual Evoked Potentials. *Journal of Neuroscience*, *42*(47), 8817–8825. https://doi.org/10.1523/JNEUROSCI.1879-21.2022

Sadil, P., Cowell, R., & Huber, D. E. (2021). *The Push-pull of Serial Dependence Effects: Attraction to the Prior Response and Repulsion from the Prior Stimulus*. PsyArXiv. https://doi.org/10.31234/osf.io/f52yz

Schwartz, O., Sejnowski, T. J., & Dayan, P. (2009). Perceptual organization in the tilt illusion. *Journal of Vision*, *9*(4), 19. https://doi.org/10.1167/9.4.19

Schwiedrzik, C. M., Ruff, C. C., Lazar, A., Leitner, F. C., Singer, W., & Melloni, L. (2014). Untangling Perceptual Memory: Hysteresis and Adaptation Map into Separate Cortical Networks. *Cerebral Cortex*, *24*(5), 1152–1164. https://doi.org/10.1093/cercor/bhs396

Sheehan, T. C., & Serences, J. T. (2022). Attractive serial dependence overcomes repulsive neuronal adaptation. *PLOS Biology*, *20*(9), e3001711. https://doi.org/10.1371/journal.pbio.3001711

Sheehan, T. C., & Serences, J. T. (2023). *Distinguishing response from stimulus driven history biases* (p. 2023.01.11.523637). bioRxiv. https://doi.org/10.1101/2023.01.11.523637

Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). Serial dependence in the perception of visual variance. *Journal of Vision*, *18*(7), 4. https://doi.org/10.1167/18.7.4

Sutterer, D. W., Foster, J. J., Serences, J. T., Vogel, E. K., & Awh, E. (2019). Alpha-band oscillations track the retrieval of precise spatial representations from long-term memory. *Journal of Neurophysiology*, *122*(2), 539–551. https://doi.org/10.1152/jn.00268.2019

Urai, A. E., & Donner, T. H. (2022). Persistent activity in human parietal cortex mediates perceptual choice repetition bias. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-33237-5

van Bergen, R. S., & Jehee, J. F. M. (2019). *Probabilistic representation in human visual cortex reflects uncertainty in serial decisions* [Preprint]. Neuroscience. https://doi.org/10.1101/671958

Van der Burg, E., Toet, A., Brouwer, A.-M., & Van Erp, J. B. F. (2021). Serial Dependence of Emotion Within and Between Stimulus Sensory Modalities. *Multisensory Research*, 1–22. https://doi.org/10.1163/22134808-bja10064

Vo, V. A., Sutterer, D. W., Foster, J. J., Sprague, T. C., Awh, E., & Serences, J. T. (2022). Shared Representational Formats for Information Maintained in Working Memory and Information Retrieved from Long-Term Memory. *Cerebral Cortex*, *32*(5), 1077–1092. https://doi.org/10.1093/cercor/bhab267

Zhang, H., & Alais, D. (2020). Individual difference in serial dependence results from opposite influences of perceptual choices and motor responses. *Journal of Vision*, *20*(8), 2. https://doi.org/10.1167/jov.20.8.2

Zhang, H., & Luo, H. (2023). Feature-specific reactivations of past information shift current neural encoding thereby mediating serial bias behaviors. *PLOS Biology*, *21*(3), e3002056. https://doi.org/10.1371/journal.pbio.3002056

Supplementary Figures

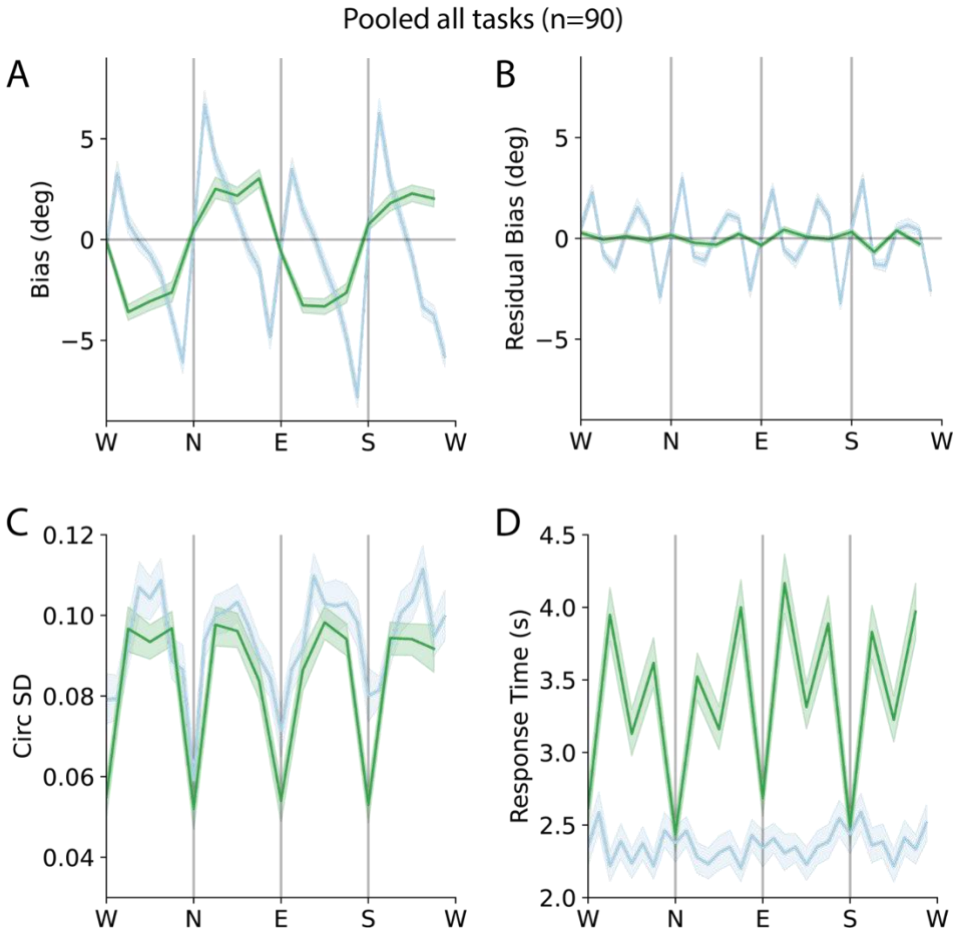


Figure 3-7 Context independent biases

A. Stable biases associated with compass (green) and dot (blue) stimuli respectively. The biases show a strong dissociation across stimulus type. Shading is SEM across participants, pooling data across all experiments. B. Residuals of a correction procedure for the two stimulus types. C. Similar changes in response precision as a function of stimulus angle is observed across stimulus types with participants more accurate around cardinal axes. D. Responses on compass trials are much slower on compass trials. Within compass reports, responses were fastest for cardinal coordinates (eg. N), slower for secondary coordinates (eg. NE), and slowest for tertiary coordinates (eg. NNE).

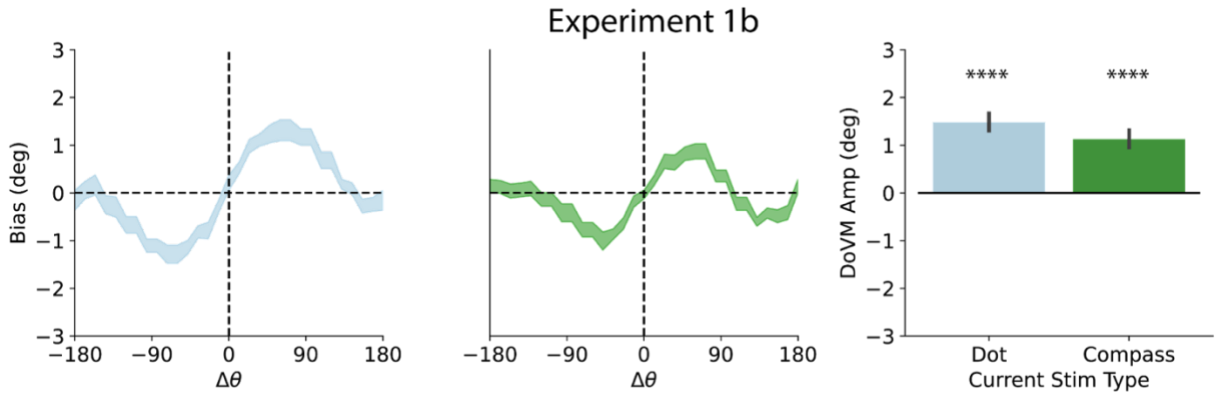


Figure 3-8 Response biases for Experiment 1b.

Left and Central panels show biases for dot and compass trials respectively. Right panel shows biases are significantly attractive in both conditions.

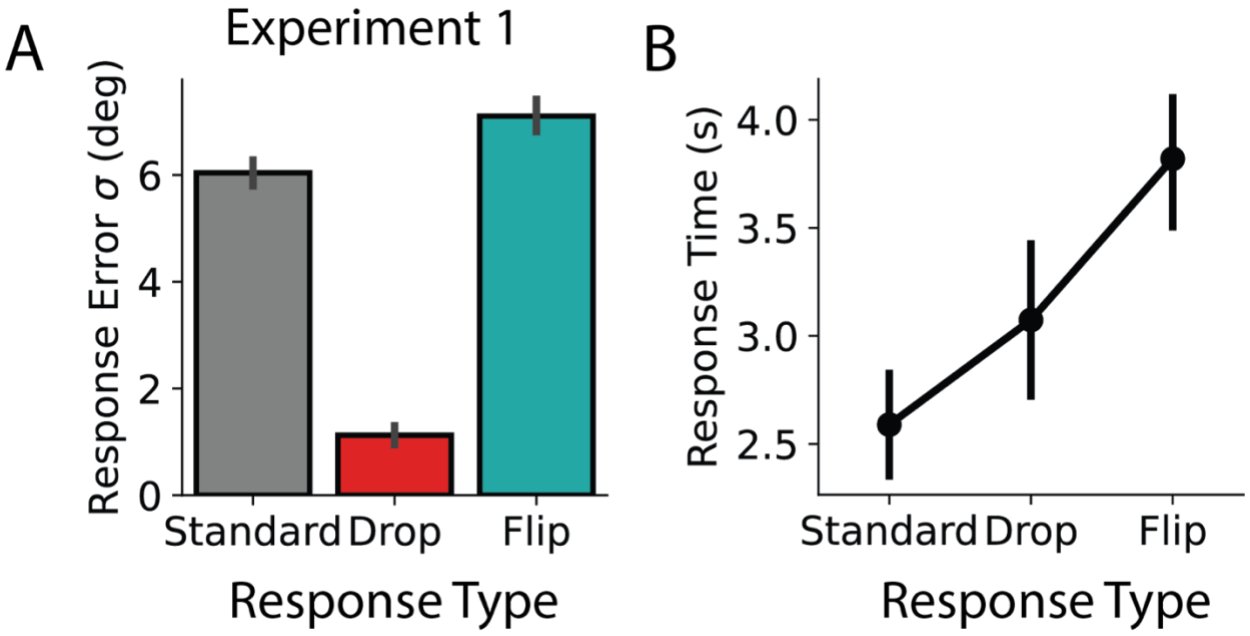


Figure 3-9 Performance and RT across response type.

A. Participants were slightly less accurate for "flip" relative to standard reports. Participant was very low for "random" responses and this is a good estimate of the floor of motor error. B. Response were significantly slower for "flip" relative to standard reports.

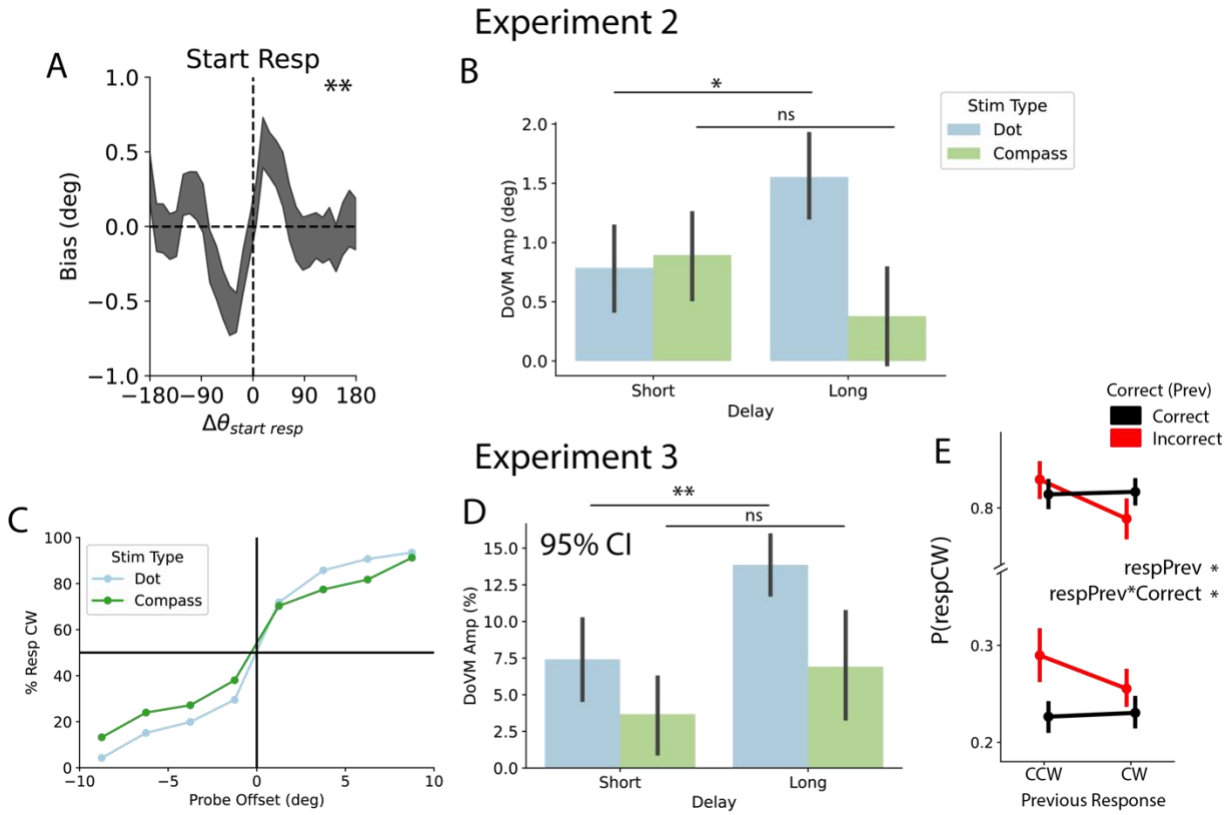


Figure 3-10 Supplemental analyses for experiments 2 and 3.

A-B Expt. 2, C-E Expt. 3. A. Response are significantly attracted to the start point of the response dial but only within ±90°. B. Bias was generally stronger for long delay trials for dot but not compass trials. Note that we use a 1-tailed test for this comparison. C. Average psychometric curves across participants for both stimulus types. The curve is notably steeper for dot over compass stimuli. D. Bias was significantly stronger on long delay trails for dot but not compass trials. Bootstrap test for comparing means. E. Response biases in Expt. 3. Splitting responses based on previous response regardless of the stimulus offset shows a general propensity to switch responses across trials.

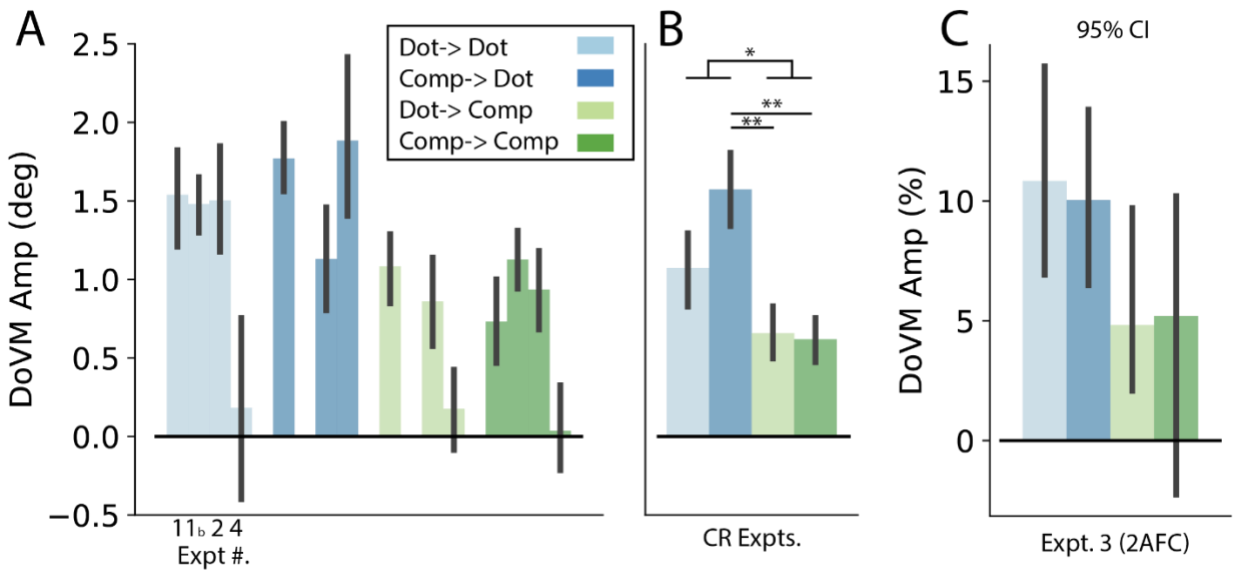


Figure 3-11 Bias towards the previous (N-back=1) stimulus across experiments and conditions. A. Effects split out for experiments 1, 1b, 2 and 4. B. Effects combing data across all continuous report experiments and reveals a global effect of dot trials exhibiting larger serial dependence. Error bars are SEM. C. Same for % responses in Expt. 3. Error bars are 95 percentiles of bootstrapped distribution.

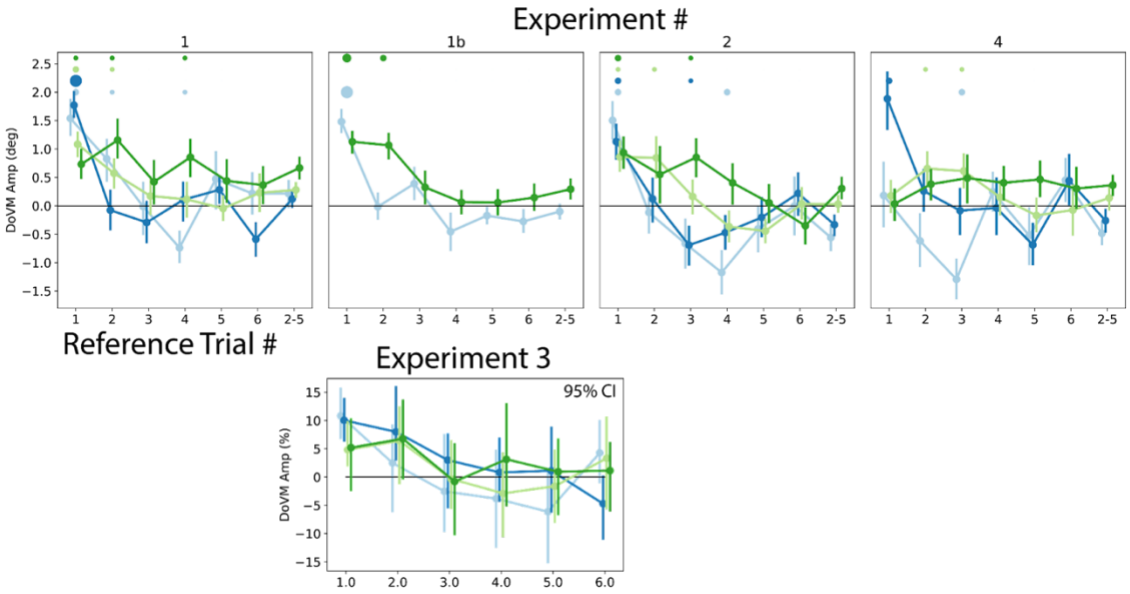


Figure 3-12 DoVM amplitude across condition and reference trial for each individual experiment. Patterns are generally consistent with our pooled analyses.

GENERAL CONCLUSION

This work examined the underpinnings of serial dependence through behavioral, neural, and in-silico models. Early work on serial dependence was largely focused on the central role of low-level vision and the effects of stimulus position and feature similarity. While this work is certainly important, our work has highlighted the more high-level and invariant nature of this bias. First in chapter 1, we reported the novel finding that attractive serial dependence can occur even when early sensory representations are dominated by repulsive adaptation. Thus, while attractive biases at encoding could play a role under some conditions, it is not a necessary condition. We further proposed a few competing models including one where adaptation occurs at encoding but is accounted for in post-sensory readout. A novel idea arising from this model is that serial dependence occurs in spite of – rather than instead of – neural adaptation.

In Chapter 2 we further explored the implications of the competing forces of repulsive and attractive biases in serial dependence. In simulations, we found that under many conditions attractive and repulsive biases cancel each other out perhaps explaining the great diversity of serial dependence findings, including the role of individual differences. Finally, we applied a technique we verified with simulations to disentangle stimulus vs. response driven biases in a delayed report task. Here, consistent with the post-perceptual account in chapter 1, we found consistent evidence for an attraction towards previous responses, not stimuli.

In chapter 3, we applied more experimental techniques to determine the source of the serial dependence. While we found strong evidence for a decisional/response centered bias, we also found examples of low-level stimulus driven biases in the same responses. We reconcile this and our other findings by proposing a more general account of serial dependence arising from a canonical function for stability across many stages of information processing. Such biases may

only arise under a small range of experimental conditions for each level of information processing, leading to often competing interactions across layers.