12-18-2020

Uncovering the Neural and Behavioral Factors That Underlie Changes in Processing Visual Orientation

Patrick Sadil
University of Massachusetts Amherst

UNCOVERING THE NEURAL AND BEHAVIORAL FACTORS THAT UNDERLIE CHANGES IN PROCESSING VISUAL ORIENTATION

A Dissertation Presented

by

PATRICK S. SADIL

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2020

Psychological and Brain Sciences

UNCOVERING THE NEURAL AND BEHAVIORAL FACTORS THAT UNDERLIE CHANGES IN PROCESSING VISUAL ORIENTATION

A Dissertation Presented

by

PATRICK S. SADIL

Approved as to style and content by:

Rosemary A. Cowell, Chair

David E. Huber, Member

Michael A. Cohen, Member

Erik G. Learned-Miller, Member

Caren Rotello, Department Head
Psychological and Brain Sciences

ACKNOWLEDGMENTS

Thanks go first to my advisor Rosie Cowell. I did not start as her lab manager intending to become a graduate student, but I am grateful that she encouraged me to do so. I also thank Dave Huber for becoming a *de facto* second advisor. I appreciate that both of them have allowed me to roam through a range of projects. Their deep knowledge, rapid insights, kind advice, and pushes toward clarity have helped me grow as a scientist. In this dissertation, many of the most interesting ideas can be traced back to them. I am grateful that they patiently gave me the space to implement them.

As a member of both the computational Memory and Perception Lab and the Cognitive Experiments Models and Neuroscience Lab, I have benefited from many talented and hardworking research assistants and lab managers. For the experiments in this dissertation, thanks go especially to Mar Nikiforova for both recruiting many participants and participating in many pilots.

Much of my enjoyment of UMass comes from the post-docs and other graduate students. They were people with whom I could gripe about inscrutable software, discuss new analyses, organize co-working sessions, drink, bike through the towns, celebrate weddings, watch movies. I feel fortunate to have gone through graduate school with them.

Finally, I owe a lot to my second community in Western Mass, a small but impactful Capoeira group: Jahmal, Tara, Bruno, Simone, Ian, John, Sen, the many others who have dropped by for only a few classes, and those from this last year who are sticking around. Yes, this group has offered a helpful distraction from the

occasionally stressful UMass, offering somewhere that I can escape to when my experiments fail. But this group has also been enriching: inspiring me to be healthy, exemplifying focus, and providing a space for playing with other people right at the edge of everyone's comfort—stretching but not straining.

ABSTRACT

UNCOVERING THE NEURAL AND BEHAVIORAL FACTORS THAT UNDERLIE CHANGES IN PROCESSING VISUAL ORIENTATION

SEPTEMBER 2020

PATRICK S. SADIL

B.A., REED COLLEGE

M.S., UNIVERSITY OF MASSACHUSETTS, AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Rosemary A. Cowell

From moment to moment, the visual environment appears stable; despite prolonged scrutiny, the edge of a desk is not perceived to change. But this apparent stability emerges from perceptual and decisional systems that undergo continuous modulation. In two chapters, I focus on two different kinds of modulation to the processing of visual orientation (i.e., the tilt of an edge). In both chapters, the form of modulation is latent, obscured by standard analyses. To detect those latent changes in perceptual decisions, I develop in this dissertation new statistical tools, at both behavioral and neural levels.

In the first chapter, I consider modulations to behavior in an orientation judgment task. Viewing and responding to an orientation causes systematic errors in

subsequent responses (Fischer & Whitney, 2014; Gibson & Radner, 1937): the orientation reported on one trial can appear to be biased either toward (attracted to) or away (repelled) from recent orientations. I performed a meta-analysis of the literature on attractive biases, finding a wide variety of effect sizes, with no experimental variable clearly explaining this variation. I show that this variation likely arises from a mixture of attraction to the last response and repulsion from the last stimulus; both forces affect every response, and for any experiment the relative mixture can result in on-average behavior that is only repulsive, only attractive, or neither. I developed two complementary techniques for disentangling this mixture and demonstrate their effectiveness as applied to both a new experiment and previously published experiments.

In the second chapter, I developed a technique for identifying how orientation "tuning" functions change with experimental manipulations (e.g., high/low contrast). These tuning functions and their modulation have been observed with single-cell electrophysiology in animals, but there are no non-invasive methods for identifying them in humans. Using functional magnetic resonance imaging, individual voxels exhibit tuning despite arising from the combined responses of hundreds of thousands of neurons. My technique models the distribution of neurons contributing to each voxel and uses model comparison to identify the most likely form of neuromodulation. I validated this technique with a new neuroimaging experiment.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

THE SERIAL DEPENDENCE EFFECT IN ORIENTATION JUDGMENTS IS BOTH AN ATTRACTION TO PRIOR RESPONSES AND A REPULSION FROM PRIOR STIMULI

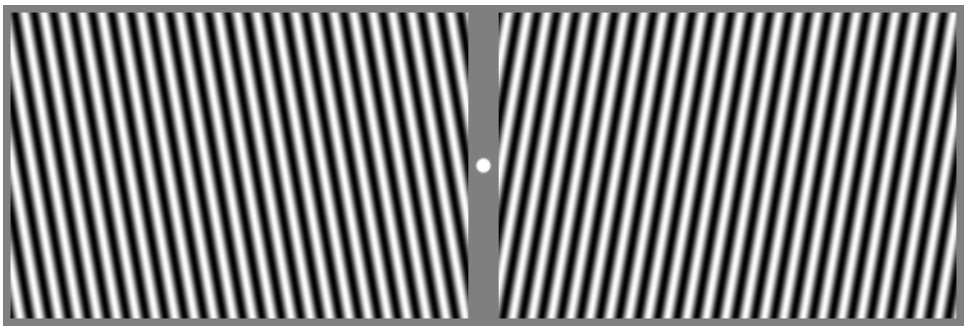
1.1 Abstract

Context affects how humans process visual information, of even low-level information. A classic example is the tilt aftereffect, in which viewing one orientation repels perception of a second orientation. In contradiction to this, the more recently identified "serial dependence" effect is a bias toward (i.e., attraction to) recently seen stimuli. However, serial dependencies range from strongly attractive to weakly repulsive. To make sense of this variability, I performed a replication experiment and used it to develop two new analysis techniques. I then applied the techniques to four published datasets. In both the replication and re-analyses, the techniques reveal that the serial dependence effect reflects the combined influence of two latent factors, one that attracts to prior responses and one that repels from prior stimuli. Variability in serial dependence likely emerges from these two factors canceling each other out to different degrees in different experiments. In all cases, the analyses revealed a lack of evidence that serial dependence implies a novel attraction to recently encountered visual information.

1.2 Introduction

Fixate on the white dot at the center of Figure 1.1A for about 20–30 seconds. While fixating, your vision may appear "normal". However, staring at that dot –

(A)



(B)

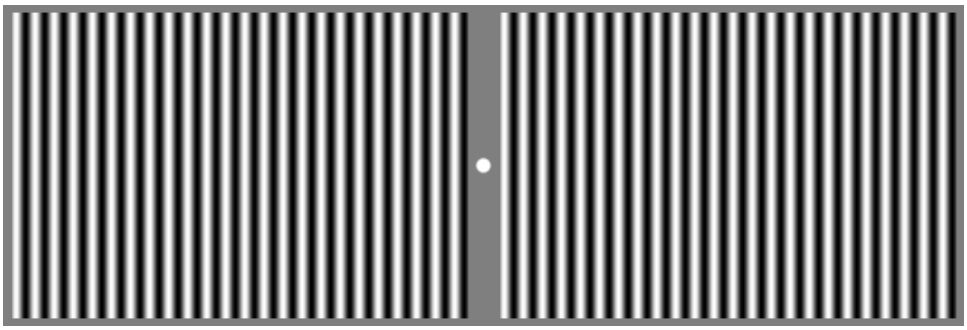


Figure 1.1: Tilt Aftereffect (Gibson & Radner, 1937). **(A)** Sample Inducing Stimulus. **(B)** Sample Target Stimulus. After fixating for 20–30 seconds on A, the top of the squares in B may appear to tilt inwards. To enhance the effect, it may help to zoom in so that the stimuli fill the screen.

allowing regions of your visual field to be bombarded with orientation information – will profoundly alter your ability to perceive orientation. The alteration can be experienced by quickly shifting your gaze to the white dot at the center of Figure 1.1B. How do the two squares on either side of that dot appear? After fixating on Figure 1.1A, the squares in Figure 1.1B may appear to lean slightly inward at the top. But such leaning would be an illusion. The squares in Figure 1.1B are plumb, the lines vertical. This is an old, well-known effect called the tilt aftereffect (Gibson & Radner, 1937).

Figure 1.2A summarizes the tilt aftereffect. The left side of the subfigure depicts a typical trial in a study of the effect, including both the stimuli presented to a

participant and how that participant might respond. First, one stimulus is presented, which I will refer to as the "inducer" or inducing stimulus (e.g., Figure 1.1A would be the inducer). After the inducer, another stimulus is presented, which I'll refer to as the "target" stimulus (e.g., Figure 1.1B). Finally, the participant reports the orientation of the target. The right side of the subfigure shows a standard way of plotting the data. The x-coordinate of the plot is set by the relative orientations of the target and inducer. In Figure 1.2A, the inducer is oriented clockwise relative to the target, and so the depicted trial would be in the right half of the graph. The y-coordinate is set by the relative orientations of the target and response. The depicted response is counterclockwise to the target, so the resulting trial lands in fourth quadrant. Data falling in this quadrant will be called "repulsive", in reference to how the inducer appears to have pushed the response past the target. Data in the second quadrant will also be called repulsive. Data in a study of the tilt aftereffect tend to fall in these two quadrants.

One way to talk about the tilt aftereffect is that recently-viewed orientations can alter perception. The tilt aftereffect can occur after even only brief exposure to the inducer (e.g., under 10 ms Sekuler & Littlejohn, 1974), does not require awareness of the inducer (Kanai et al., 2006), is sensitive to a range of low-level features for which early visual neurons are also sensitive (Greenlee & Magnussen, 1987; Harris & Calvert, 1985, 1989; Magnussen & Johnsen, 1986; Morant & Mikaelian, 1960; Parker, 1972), can occur across different screen location and spatial frequencies (Jacob et al., In Preparation), and mechanisms that could give rise to the tilt aftereffect have been observed in single-cell recordings of early visual neurons (Clifford et al., 2000; Dragoi et al., 2000; Gutnisky & Dragoi, 2008; Patterson et al., 2013; Wissig & Kohn, 2012). Combined, the research suggests that the tilt aftereffect – like visual aftereffects in general – is caused by merely viewing the inducing stimulus and does not require

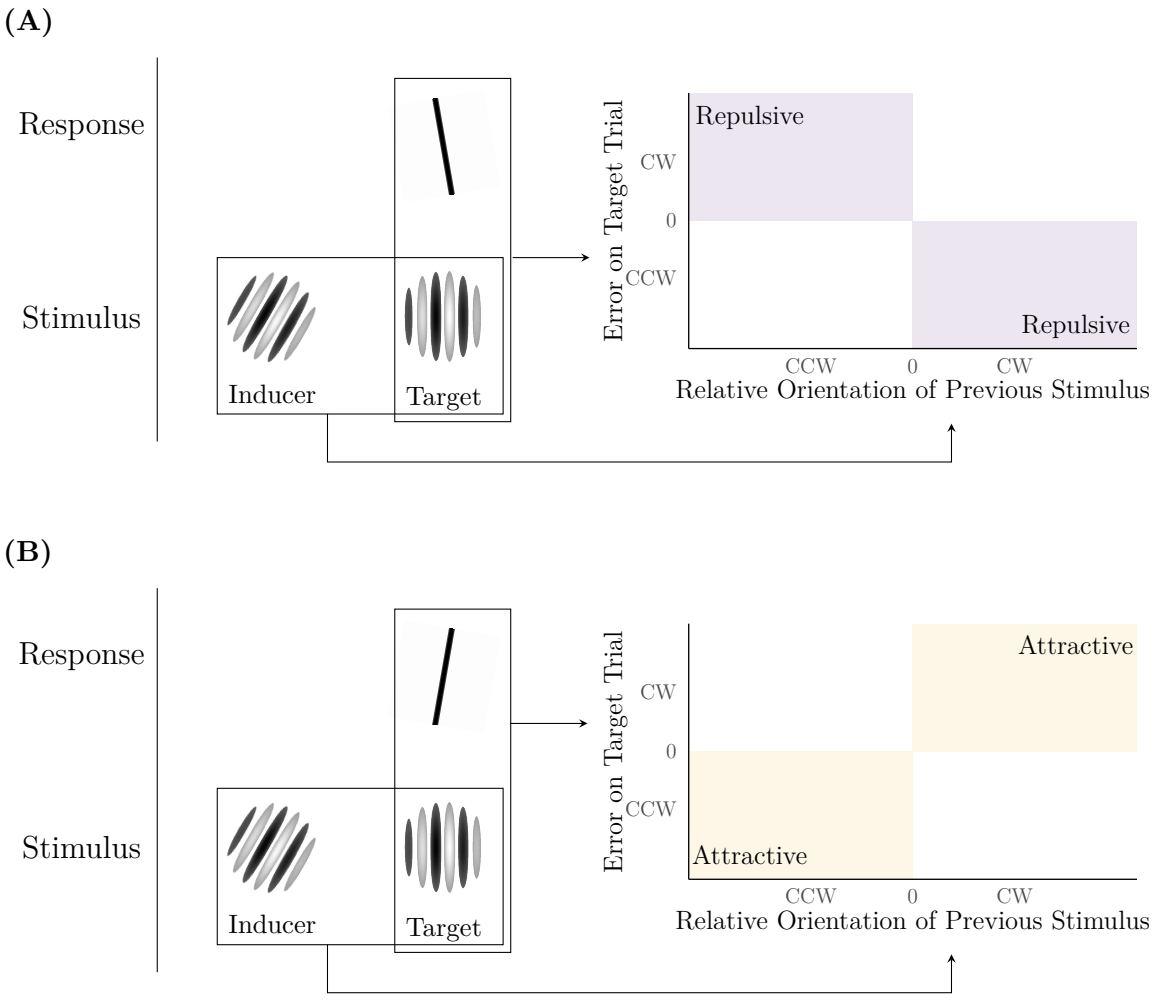


Figure 1.2: Schematic for Repulsive and Attractive Dependencies. In both subfigures, the left portion shows an example trial, and the right portion shows how the resulting trial could be plotted. **(A)** The tilt aftereffect tends to produce errors whose sign does not match the relationship between the inducer and the target. **(B)** The serial dependence effect labels the opposite tendency, errors whose sign match the relationship between the inducer and target. Figure adapted from Fischer and Whitney (2014).

higher-level processing of the inducer (which is not to say that aftereffects are immune to higher-level processes; e.g., Thompson & Burr, 2009).

A new effect has recently been identified that occurs in experimental setups similar to those that cause repulsive aftereffects, but which produces the opposite pattern

of data (Cicchini et al., 2014; Fischer & Whitney, 2014; Kiyonaga et al., 2017). That is, information from the recent past can also attract (Figure 1.2B). When participants are presented with the same kind of stimuli repeatedly (e.g., a grating whose orientation differs on each trial; Figure 1.3A), their responses to the current stimuli can be systematically biased toward previously encountered stimuli. This attractive effect has been labeled serial dependence, and it may be caused by a new mechanism – a "continuity field" – that works in opposition to aftereffects (Fischer & Whitney, 2014). Like aftereffects in general, attractive serial dependencies have been observed for a wide range of stimulus classes, including orientation, spatial location, motion direction, numerosity, timing, identity, gaze direction, ensemble statistics, attractiveness, and gender (Alais et al., 2018; Fischer et al., 2020; Fornaciai & Park, 2018b; Liberman et al., 2014; Suárez-Pinilla et al., 2018; Taubert, Alais, et al., 2016; Taubert, Van der Burg, et al., 2016; Xia et al., 2016). Given this wide range of stimulus classes, the continuity field could be essential to how perceptual-decisions leverage information from the recent past.

Despite its ubiquity, the magnitude of serial dependence is highly variable–even considering only serial dependencies in judgments of orientation. In orientation judgments, the quadrant of the data determines whether the data are attractive or repulsive (Figure 1.2). Beyond tending to land in certain quadrants, the magnitude of the errors tend to vary along the x-axis. To capture this variability, the serial dependence effect is often modeled with the derivative of a Gaussian function (Figure 1.3B). This is a convenient model because the derivative has an amplitude parameter that corresponds to the maximum average error elicited by prior stimuli, which is taken as the magnitude of the serial dependence effect. Across studies, estimated amplitudes have ranged from 11° to −5° (Figure 1.3C), where a negative value indicates repulsion from a recently encountered orientation (akin to a tilt aftereffect). For scale, the just-noticeable difference for orientation information presented, in a representa-

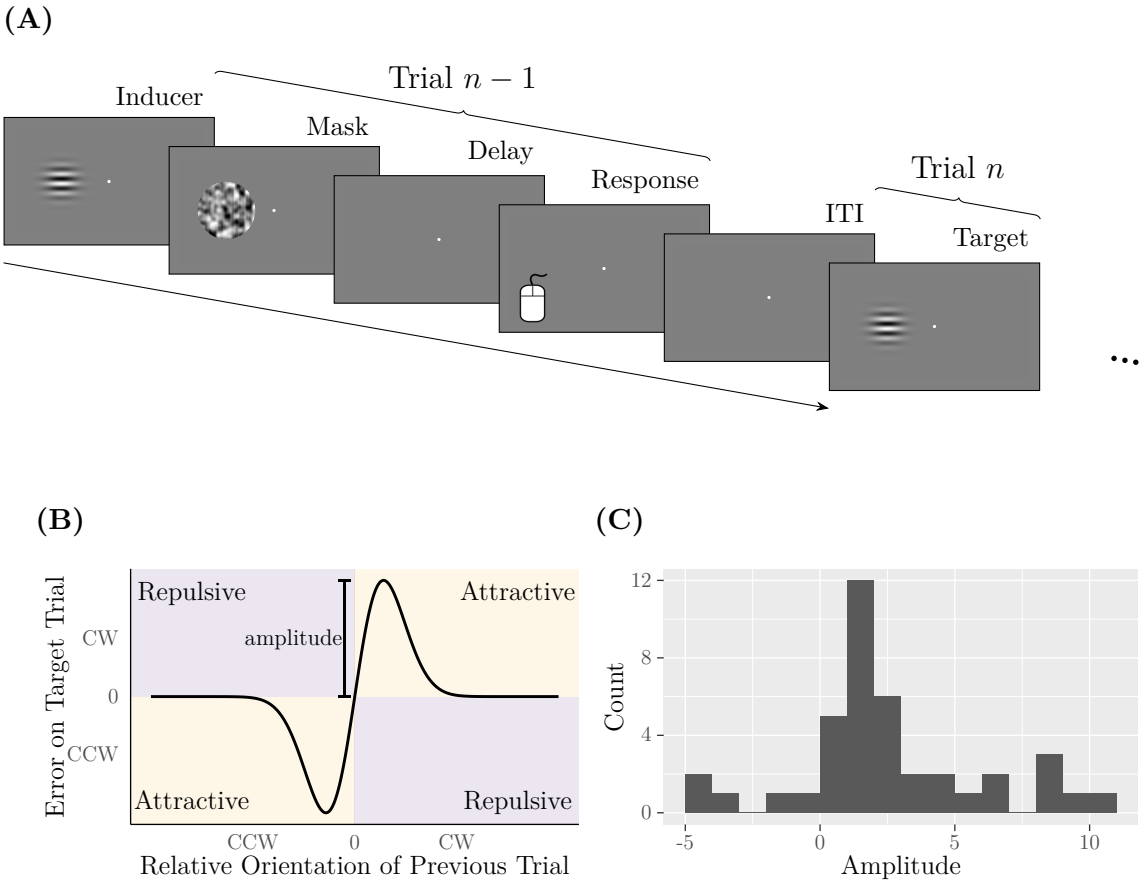


Figure 1.3: Review of serial dependencies in judgments of Orientation. **(A)** Typical Serial Dependence Paradigm. Trials in typical task consist of five parts: stimulus presentation, mask, a delay period, the response, and then an inter-trial interval (ITI). The stimulus on trial $n-1$ serves as the inducing stimulus, and the stimulus on trial n is the target. **(B)** The Derivative of a Gaussian Function. The derivative has been used as a descriptive model for errors. The derivative has two parameters, a width and an amplitude. The sign of the amplitude determines whether the dependence is attractive or repulsive (displayed here as positive, meaning attractive). **(C)** Histogram of Observed Amplitudes. The amplitudes are from the studies listed in Table 1.1.

tive serial dependence experiment, is $\sim 5.4^{\circ}$ (Fischer & Whitney, 2014). This means that serial dependencies can vary by approximately three just-noticeable differences across experiments, and, importantly for theoretical interpretations, the range of effects straddles the zero point. That is, across different studies, the effect changes

direction qualitatively from positive (attractive) to negative (repulsive). Individual studies have identified factors that contribute to this variability, including attention, participants' confidence, the spatial frequency of the stimuli, eccentricity, the time spent observing each stimulus, and whether participants respond on every trial (Cicchini et al., 2018; Fischer & Whitney, 2014; Fritsche & de Lange, 2019a; Fritsche et al., 2017; Liberman et al., 2016; Pascucci et al., 2019; Samaha et al., 2019; van Bergen & Jehee, 2019). However, no single factor, nor specific combination of factors, has yet been shown to account for the full range of reported amplitudes.

To better understand the factors contributing to the variability, I first consider the ongoing debate about what causes serial dependence (Bae & Luck, 2020; Cicchini et al., 2017; Fischer & Whitney, 2014; Fornaciai & Park, 2018a, 2018b; Fritsche & de Lange, 2019a; Fritsche et al., 2017; Manassi et al., 2018; Pascucci et al., 2019). The debate hinges on what kind of processing the inducing stimulus must undergo to induce serial dependence. For example, the original account of serial dependence claims that attending to the inducing stimulus engages the continuity field, that the continuity field is sufficient to cause the attraction, and that further higher-level processes are not required (i.e., no high-level processes beyond attention; Cicchini et al., 2017; Fischer & Whitney, 2014; Manassi et al., 2018). An alternative account claims that merely attending to the inducing stimulus is insufficient to elicit an attraction, and that attractive serial dependencies appear only after the inducer has undergone post-perceptual processing (Alais et al., 2017; Bae & Luck, 2020; Fornaciai & Park, 2019; Fritsche & de Lange, 2019a; Fritsche et al., 2017; Pascucci et al., 2019). I will refer to these two accounts, the original and alternative, as a perceptual and post-perceptual account, respectively.

The post-perceptual account challenges the original interpretation of the serial dependence effect by pointing out that multiple stages of a perceptual-decision contribute to attractive dependencies. For example, in two-alternative forced choice

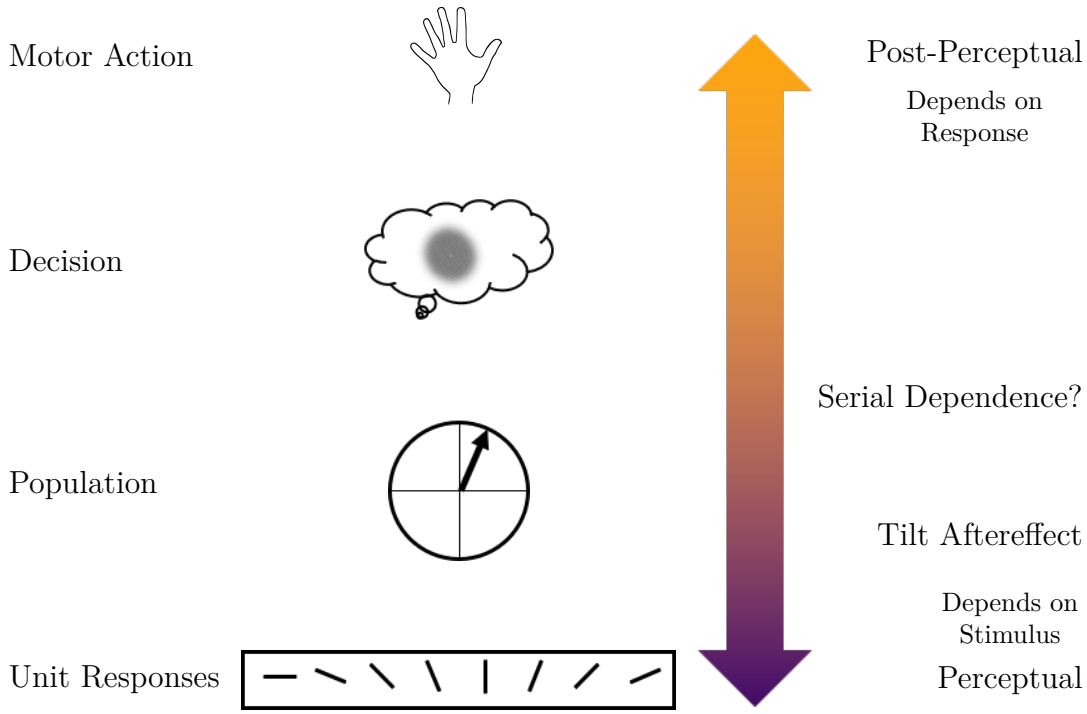


Figure 1.4: Serial Dependence could emerge at multiple levels of processing. An inducing stimulus causes a cascade of processing. The tilt aftereffect likely requires only the earliest kinds of perceptual processing (e.g., unit responses in visual cortex). However, dependencies – including those that underlie the serial dependence effect – may exist throughout this continuum. Dependencies caused by processes early in this cascade are assumed to be coupled more closely with the veridical stimulus being sensed, whereas dependencies caused by later processes are assumed to be coupled more closely with the final behavioral response.

tasks, participants tend to repeat their responses (e.g., Garner, 1953; McKenna, 1984), and this kind of a tendency can be caused by making a decision about the inducer, rather than simply viewing the inducer (Akaishi et al., 2014). Such tendencies have generally been called "sequential effects", but also, emphasizing either a processing stage or the behavior, "decisional inertia" and "response hysteresis". A response hysteresis has been observed in even tasks that require participants to respond to an unchanging stimulus (Gilden, 2001; Gilden et al., 1995). Importantly, a factor like decisional inertia implies that attractive dependencies can arise between participants' decisions themselves, and this attraction can exist without regard to the

stimulus to which they were responding[1]. Even beyond a decision, myriad processes intervene between sensing the inducer and making a response (Figure 1.4), and so the inducer might cause multiple effects, including low-level perceptual effects (e.g., a repulsive tilt aftereffect) as well as high-level effects (e.g., an attractive response hysteresis). Each of these effects could contribute to the observed serial dependencies (see also Jones et al., 2013).

But although multiple induced effects is central to the post-perceptual account (Alais et al., 2017; Bae & Luck, 2020; Fornaciai & Park, 2019; Fritsche & de Lange, 2019a; Fritsche et al., 2017; Pascucci et al., 2019), an important implication of this has not yet been highlighted: if there are two, opposing dependencies induced, then the average results, plotted as a function of the relative difference between inducer and target, will reflect a complex mixture of these dependencies. For an intuitive understanding of this mixture, consider again the derivative of Gaussian model (Figure 1.3B). Two dependencies can be simulated with two derivatives, one whose amplitude is positive (Figure 1.5, first column), and one whose amplitude is negative (Figure 1.5, second column). Plotting a dependence on recently encountered stimuli would show only the summation of these two latent forces (Figure 1.5, third column). Variation in even only one of these latent forces could drastically alter the observed summation, the observed serial dependence. For example, if the continuity field mirrored the tilt aftereffect perfectly (i.e., the derivatives differed only in their amplitudes' sign), then they could combine to produce a flat line – an observed lack of dependence despite robust latent dependencies (Figure 1.5, bottom row). Multiple latent dependencies could therefore account for substantial variability in reported serial dependencies. Considering the many levels of processing, the typical analysis of

[1]As mentioned, the perceptual and post-perceptual accounts are two sides of an ongoing debate. One of the main empirical questions has asked whether participants must make a decision about the inducer to elicit the attractive serial dependence. In Appendix A, I review experiments that have attempted to answer this question, concluding that the evidence remains equivocal.

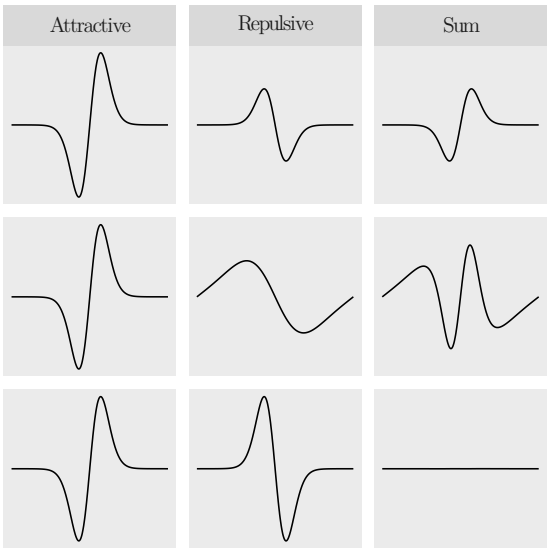


Figure 1.5: Repulsive and attractive dependencies can combine, resulting in many shapes. In a serial dependence experiment, the inducing stimulus may induce many effects simultaneously. The first two columns, an attractive and a repulsive dependence, are summed to give the third. Only the third column would be revealed in a typical analysis of the serial dependence effect. The repulsive effect differs in each row (e.g., experimental forces may conspire to produce different tilt aftereffects), while the attractive effect remains constant. Although the attractive effect never changes, the resulting summation (third column) results in a wide range of effects. Basing inferences on only the summation would obscure the true source of the variability, variation in the repulsive effect.

serial dependence – fitting a single derivative of Gaussian – may not reveal a single dependence, but instead the summation of multiple, latent dependencies.

Variability in serial dependence, even its sign, may therefore not reflect any variability in the magnitude of an attractive dependence, but instead reflect variability in the magnitude of an offsetting, repulsive force. Indeed, this situation is likely, considering that the magnitude of aftereffects (which are repulsive) are known to be variable: e.g., the tilt aftereffect is sensitive to many details of an experiment, including spatial frequency, eccentricity, stimulus size, duration of inducer presentation, duration of target presentation, the relative contrast of the inducer and the target stimuli, and the duration between presentation of the inducer and target (Greenlee &

Magnussen, 1987; Harris & Calvert, 1985, 1989; Magnussen & Johnsen, 1986; Morant & Mikaelian, 1960; Parker, 1972; Wolfe, 1984). Because there is no reason to believe that a serial dependence experiment with oriented stimuli would disable the tilt aftereffect, the magnitude of any observed attractive dependence should depend on at least the same set of experimental factors as the tilt aftereffect. Some of these same factors may also influence attractive forces that contribute to the serial dependence effect (Cicchini et al., 2018; Fischer & Whitney, 2014), but given that they influence the tilt aftereffect, it is all but impossible to determine whether this is the case using the analysis techniques typically found in the serial dependence literature. Indeed, many of the factors known to affect the tilt aftereffect have varied across orientation serial dependence studies (Table 1.1).

If serial dependence reflects the summation of multiple latent forces, it should in principle be decomposable. For instance, consider latent forces occurring at lower levels, which should be more tightly coupled to what an observer senses, versus latent forces occurring at higher levels, which should be more tightly coupled to how that observer responds. Regardless of whether these higher-level, latent forces are perceptual or decisional, they relate to the observer's impression of the stimulus, an impression that is not necessarily accurate. Correspondingly, each inducing trial in a serial dependence experiment entails two covariates: the veridical features of the inducing stimulus and how a participant responded to that stimulus. Both of these may elicit dependencies in errors to the target stimulus. In this chapter, I use these two covariates to disentangle, not specific perceptual and post-perceptual stages, but instead simply the two ends of a perceptual to post-perceptual continuum (Figure 1.4). The empirical questions are 1) whether errors to the target stimulus depend on both the inducing stimulus and responses to the inducer, and, if so, 2) are they attractive or repulsive?

To answer these questions, some trials will be more informative than others. On trials in which participants make minimal errors, their responses will be indistinguishable from the inducing stimulus (see also St John-Saaltink et al., 2016; Fischer & Whitney, 2014, experiment 1c), effectively resulting in only a single covariate. The most important trials will be those in which the stimulus on the previous trial is most dissimilar to the response on the previous trial – those trials in which participants make the largest errors. Therefore, analyzing dependencies following participants' most and least accurate trials (i.e., a split-half analysis) provides a straightforward, if approximate, way to visualize distinct dependencies on the inducing stimulus and their response to that stimulus. More quantitatively, potentially distinct dependencies will be specified by applying a two-factor model to the trial-by-trial data.

To decompose serial dependencies into latent dependencies, I collected new data in a close replication of an orientation estimation experiment, in which the task was designed to be challenging, in order to cause participants to regularly make large errors (Samaha et al., 2019). I used the resulting data to develop two analyses that aim to uncover latent dependencies. Additionally, I reanalyzed four datasets across three published papers (Fischer & Whitney, 2014, experiment 1b; Pascucci et al., 2019, experiments 1 and 2; Samaha et al., 2019), reassessing whether the evidence for serial dependence reflects a new attraction toward recently encountered stimuli.

1.3 Methods

1.3.1 Participants

Power analyses indicated that approximately thirteen participants would be sufficient to both detect a single dependence and reliably distinguish between one versus two dependencies (Appendix B). To allow for the possibility that data from some participants were unusable, sixteen participants were run and given course credit for participating (including the author). All participants had normal or corrected-to-

normal vision, and they all provided usable data. The procedure was approved by the University of Massachusetts Institutional Review Board.

1.3.2 Stimulus Parameters

Stimuli were presented on an LCD monitor (ASUS VG248QE, 1920×1080 cm, 100 Hz refresh rate, 1920×1080 resolution), viewed from approximately 60.96 cm. Stimuli were displayed using the Psychophysics Toolbox (Version 3.0.14; Brainard, 1997; Pelli, 1997) and custom MATLAB code (2015a, MathWorks).

The stimulus parameters and experimental procedure closely replicated the design of (Samaha et al., 2019). Throughout the experiment, participants fixated on a light gray dot (0.08°). Grating stimuli (sine wave with 1.5 cycles per degree and phase 0 subtending a circular region of 2°) were presented on a medium gray background. Mask stimuli consisted of white noise rendered at 100 % contrast. Participants were cued to make orientation reproduction responses with a circle (6° radius) and actuated the response by clicking a mouse near the circle (within 50 pixels of the circle).

The signal-to-noise ratio of the grating was reduced by averaging the grating with white noise (Samaha et al., 2016; Samaha et al., 2019). The contrast of the grating was determined by a pilot study (3 participants, data not shown) with target stimuli created by averaging a 10 % Michelson contrast grating with white noise (100 % contrast), which elicited responses that were within ±25° of the true orientation approximately 80 % of the time.

1.3.3 Procedure

The trial structure is outlined in Figure 1.6. Each trial began with the presentation of a grating, surrounded by a circle. The grating was replaced by a mask after 200 ms, but participants could respond immediately after target onset. A circle surrounded the grating and mask, and participants were responded by using a computer mouse to click on the circle. They were instructed to report the orientation that coincided

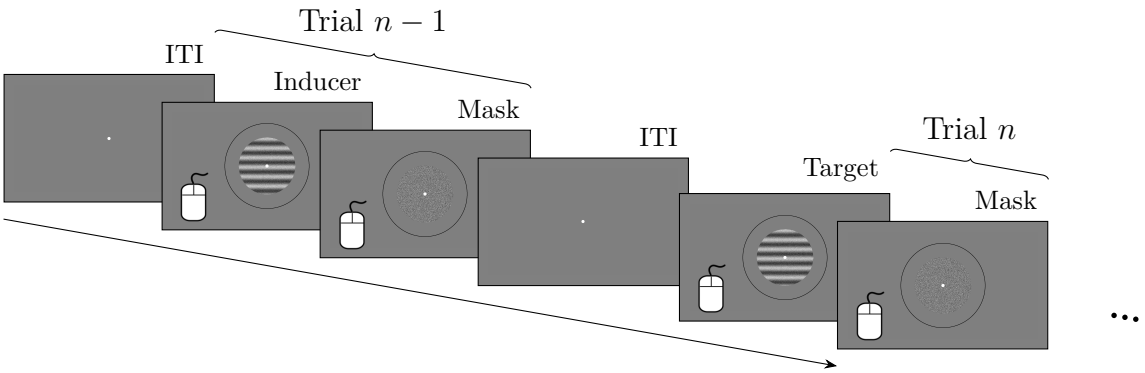


Figure 1.6: Experiment 1 Schematic. On each trial, participants were presented a grating stimulus. The stimulus was masked after 200 ms, but participants could respond immediately after stimulus onset. Participants responded by using a mouse to select an orientation on a circle. To encourage fixation, a dot was visible continuously. The grating and response in one trial (e.g., $n-1$) provided covariates for the next trial (e.g., n), which in turn provided the covariates for the subsequent trial (e.g., $n+1$). Stimuli not presented to scale. ITI: Inter-trial Interval

with a point intersected by a line (imagined) drawn through the center of the grating, both parallel to the dark and light portions of the grating and passing through the fixation dot. They practiced the task for 10 trials with an experimenter present and available to answer questions about the task. Participants could respond on either side of the circle. Trials were separated with a variable fixation period (randomly determined on each trial with a draw from a discrete uniform distribution ranging from 300–500 ms in steps of 20 ms).

The orientation on each trial was drawn at random from a discrete, uniform, distribution, sampling integers between 0–179°. Participants completed 15 blocks, each with 101 trials. The median duration of the experiment was 48 minutes (range: 34–86).

1.3.4 Published Datasets

Four published datasets were reanalyzed (Fischer & Whitney, 2014, experiment 1b; Pascucci et al., 2019, experiments 1 and 2; Samaha et al., 2019). See Table 1.1 for an overview of the methods. Note that, in the original publications, certain

preprocessing were applied that were not required in the analyses developed in this chapter. Hence, in cases where the original authors used similar analyses to what I use (i.e., a moving window, below) figures of the data from published experiments that I present here (produced with minimal preprocessing) will not match the published figures.

1.4 Analyses

The data were analyzed in three complementary ways. The first analysis was a moving window analysis, a relatively non-parametric way to visualize dependencies on either the previous stimulus or response. This is a standard way to analyze serial dependencies, complementary to the derivative of a Gaussian function model. The second analysis was a split-half analysis designed to visualize any differences between serial dependence on the previous response versus the previous orientation. Reliability of any differences was assessed based on standard errors across participants for the two serial dependence functions (i.e., a visual inspection of the average errors following the inducing stimulus and responses). A formal statistical test was performed in the third analysis by comparing different Bayesian models of the data.

1.4.1 Moving Window and Split-Half

To visualize the effects of the previous orientation and response, as might be done in a typical analysis of serial dependence, moving windows were used to smooth participants' errors. The median error for each participant was calculated in sliding windows, centered on equally spaced 200 orientation differences between $-90°$ and $90°$ (width of $\pm 12°$, following Samaha et al., 2019). The orientation differences were either between the inducing and target orientations (i.e., to visualize effects of the inducing stimulus) or between the response to the inducing stimulus and the target orientation (i.e., to visualize the effects of the response to the inducing stimulus). To eliminate

boundary artifacts, windows near ±90° included circularly wrapped copies of the data (i.e., data from stimulus orientations of 78–89° were taken as being adjacent to an orientation of −90°). The median errors in each window were then averaged across participants, and the standard errors of these averages were calculated.

As discussed in the introduction, a plot of errors as a function of just the previous response or the previous orientation will depict relatively entangled effects. For example, if participants respond with perfect accuracy, then the two plots will necessarily be identical. However, when the response to the inducing orientation is far from the inducer's true orientation (i.e., when the previous response was inaccurate), it becomes possible to isolate separate effects of the previous response versus the previous orientation, if such separate effects exist. Hence, two effects could be revealed by first separating trials based on whether participants' response to the inducer was relatively accurate or inaccurate. To conduct this analysis, participants' errors to target stimuli were placed into separate bins based on whether the response on the previous trial was relatively accurate or inaccurate (i.e., a median-split), and these two halves were separately smoothed with a moving window. Critically, this was a within-subjects median-split (i.e., involved splitting each participant's data separately), thus avoiding any confound owing to some subjects being more accurate than others.

Although a median-split provides a straightforward way to visualize separate effects of the previous response versus the previous orientation, the resulting plots will still reflect a mixture of these effects. Even relatively inaccurate trials (i.e., the higher error segment of the median-split) will be somewhat accurate (e.g., perhaps an average absolute error of only 45°), as compared to a stimulus-independent random choice (which would correspond to an average absolute error of 90°). Thus, even for the high-error bin of the split-half analysis, the axes for the two plots will be still be somewhat related, and so the separate effects would still summate or offset each other to some degree. Therefore, statistical analyses were based on a separate, Bayesian

modeling approach to test and quantify the magnitudes of any differences between the two potential kinds of serial dependence.

1.4.2 Bayesian Analyses

Serial dependencies in orientation are typically analyzed with a regression model that uses a modified derivative of Gaussian model (Figure 1.3B). However, the derivative of Gaussian has two limitations. First, the Gaussian's derivative does not capture the circular nature of orientation; when the derivative peaks far from zero, the model is discontinuous – contains a cliff-like drop off – at $\pm 90°$. This can be remedied by replacing the derivative of a Gaussian with the derivative of the density function for a von Mises distribution, which is a circular distribution that otherwise resembles a Gaussian (see Appendix D for a detailed description of the von Mises model). Three versions of the von Mises models were fit to each dataset, where each version included one of the three combinations of effects of the previous orientation and previous response (i.e., two models with only one covariate and one model with both covariates).

The second limitation is that, in some experiments, the direction of serial dependence (and the tilt aftereffect) changes at more extreme values, resulting in so called "peripheral bumps" (Appendix C; Fritsche & de Lange, 2019a; Fritsche et al., 2017; Gibson & Radner, 1937; Samaha et al., 2019). These peripheral bumps cannot be captured by either derivative model, and fitting these models when there are peripheral bumps could be misleading (Appendix C). To account for this second limitation, datasets were also fit with a relatively non-parametric model that included the potential for peripheral bumps: a cyclic spline (Wood, 2017). The cyclic spline model allowed for potentially separate effects of both the previous orientation and response.

All models were constructed and fit using the Stan language (Carpenter et al., 2016), using its interface (RStan, 2.18.2) for the R computing language (R Core Team). Stan draws samples from an approximation of the posterior distribution

using a modified version of the Hybrid (also known as, Hamiltonian) Monte Carlo algorithm (Duane et al., 1987; Hoffman & Gelman, 2014). The validity of this approximation was assessed in two ways. First, chains were monitored for divergences, an indication that the numerical simulation methods in the algorithm are compromised (Betancourt, 2017). Second, the split-$\hat{R}$ ("split r-hat") for each parameter was calculated (Gelman & Rubin, 1992). In all results, there were no divergences and for each parameter the split-$\hat{R}$ was below 1.1 (Gelman et al., 2013, chapter 14).

Inference was based on model comparison, assessing the models' predictive abilities with an approximation to a leave-one-out (LOO) cross-validation score (Appendix E; Vehtari et al., 2017, 2019). Specifically, models were compared based on their Pareto-smoothed importance sampling leave-one-out plus (PSIS-LOO+, Appendix E) cross-validation score, as calculated with the loo software package (Vehtari et al., 2020).

1.5 Results

1.5.1 Experiment 1

Participants viewed a series of differently oriented gratings presented at super-threshold levels and were asked to report each orientation. Across participants, the standard deviation of errors was 15.2°, and the median response time was 931 ms.

A moving window was used to smooth the data (Figure 1.7A). The average data collapsed over all trials, regardless of last trial accuracy, did not reveal consistent dependencies in the data, toward either the inducing orientation or previous response[2]. More specifically, although both dependencies show what appears to be a repulsive effect for large positive relative differences (i.e., values around 45°), neither dependency shows a clear repulsion for values around −45°. However, there appears to

[2]Due to the minimal preprocessing used in this chapter, biases that do not depend on the inducing trial will be visible in the data. For example, errors tended to be negative, potentially indicating that participants consistently reported orientations that were too counterclockwise. Such biases were explicitly modeled in the Bayesian analyses (Appendix D).

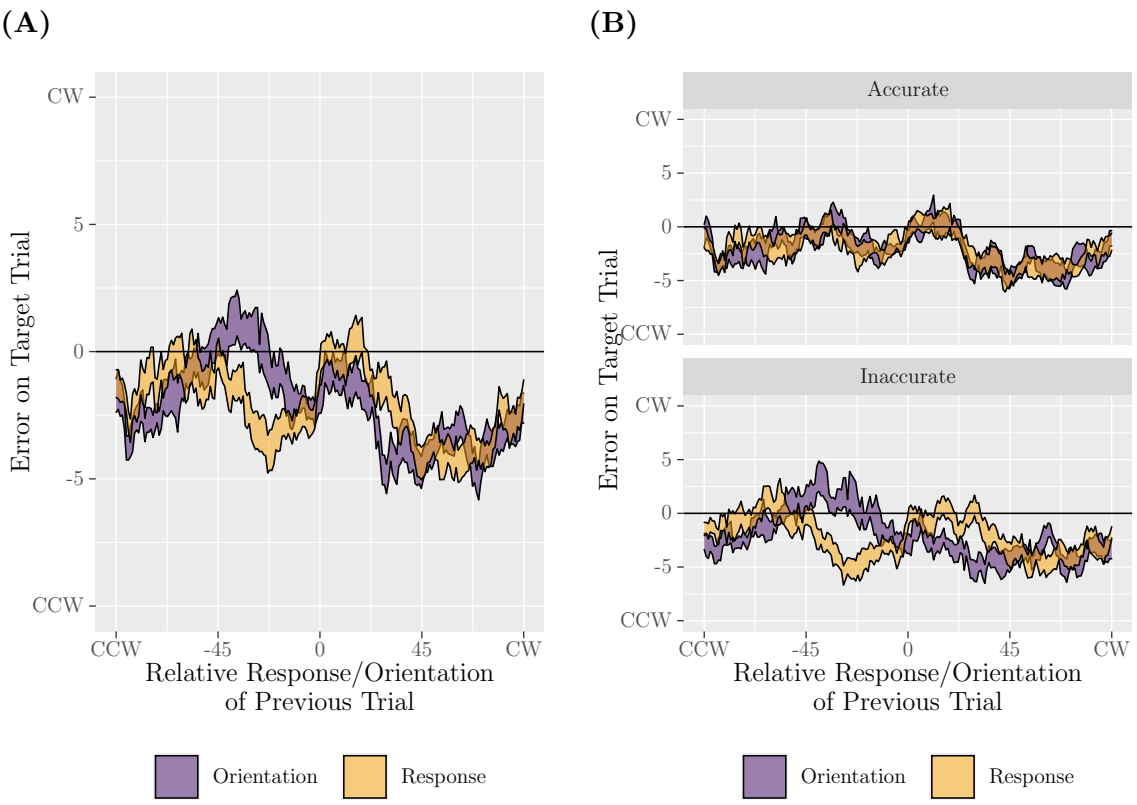


Figure 1.7: Split-half analysis reveals that errors were attracted to the previous response and repelled away from the previous orientation. In all plots, the two ribbons show how errors on the target trial vary as a function of the relative difference between the target orientation and either the inducing orientation or participants' response to that orientation. The ribbons span one standard error of the mean across participants. **(A)** A Standard Analysis, Using All Trials. Errors have been smoothed with a moving, median window. **(B)** Split-Half Analysis. The moving window has been used to separately smooth errors on target trials that followed inducing trials on which participants were the relatively accurate or inaccurate (i.e., a median-split).

be a modest difference between the dependencies for values around $-45°$, with an attractive effect for the response dependency.

Because accurate responses on the last trial confound any separate effects of response versus orientation, the errors *following* the most and least accurate trials were smoothed separately in a split-half analysis (Figure 1.7B). The expected confounding was revealed in errors following relatively accurate trials. Following accurate responses, errors plotted against the previous orientation and against the previous response yielded nearly identical patterns (Figure 1.7B, top). On those trials, the x-

axes of these two plots result in the same sorting of the data. That is, as the previous trial's response becomes more accurate, the previous orientation and the previous response become identical. However, there was a qualitative difference between the two types of dependencies following inaccurate trials, when the previous orientation and response are uncoupled. On these trials, responses were attracted toward the previous response and repelled away from the previous orientation (Figure 1.7B, bottom).

The median split offered a straightforward way to visualize the two effects, but only approximately; even when the previous trial's error was relatively high, the response will still have been related to the inducing stimulus, causing the effects to remain at least partially entangled. To more fully disentangle the two affects and assess them statistically, errors were used to fit two different kinds of Bayesian models. The first kind relied on a relatively non-parametric approach (i.e., cyclic splines). The non-parametric approach modeled the effects flexibly enough to capture the "peripheral bumps" that have occasionally been observed (Appendix C; Fritsche & de Lange, 2019a; Fritsche et al., 2017; Samaha et al., 2019). The posterior distribution from this model contained functions that separately described how errors depended on either the previous orientation or the previous response. These functions exhibited a clear difference, supporting the assertion that there was both an attraction to the previous response and a repulsion from the previous orientation (Figure 1.8, left). Furthermore, these distributions did not provide any evidence of peripheral bumps. Therefore, the dependencies ought to be adequately described with the second kind of Bayesian model, which modeled them with the relatively simple derivative of von Mises. The posterior distributions of the dependencies as modeled with the von Mises model are similar to the non-parametric spline (Figure 1.8, middle), despite that each of the separate dependence functions were less flexible. The posterior predictive distribution of the von Mises model, for the same high error trials of the split-half analysis demonstrate that the von Mises model provides a good fit to the data, further sup-

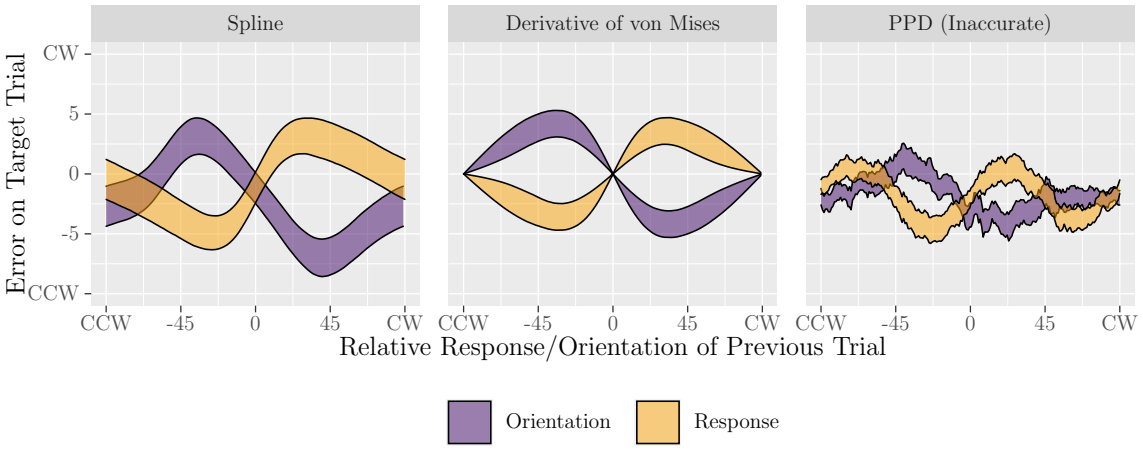


Figure 1.8: Posterior distribution of the effects of the previous orientation and response. **Left**: Posterior predictive distribution of dependencies as modeled with a cyclic spline. **Middle**: The population-level dependencies, as modeled with a derivative of von Mises. **Right**: Posterior predictive distribution of the full von Mises model on target trials that followed inducers to which participants provided inaccurate responses. Compare predictive distribution with bottom row of Figure 1.7B. In all panels, the ribbons give the 95 % highest density interval of either the modeled effects (left, middle) or the posterior predictive distribution (right).

porting the adequacy of the relatively simple von Mises model (compare Figure 1.7B Inaccurate to the right panel of Figure 1.8). Note that in all models, the effect of the previous orientation was not constrained to be repulsive and the effect of the previous response was not constrained to be attractive; all modeled dependencies were allowed to vary in direction (as well as magnitude).

To test for the significance of these effects, model comparison was used (using PSIS-LOO+, Vehtari et al., 2019). Three von Mises models were fit that allowed for either 1) an effect of the previous orientation only, 2) an effect of the previous response only, or 3) an effect of both the previous orientation and the previous response. Model comparison favored the model with both effects as compared to either of the single effect models (ELPD difference ± standard error: full model vs. orientation only: -103.72 ± 16.04; full model vs. response only: -149.99 ± 19.61).

1.5.2 Reanalysis of Published Data

In Experiment 1, two complementary analyses – a median split and Bayesian modeling – revealed that an apparent lack of serial dependence masked two robust but competing dependencies that were akin to a tilt aftereffect and response hysteresis. That is, although there were attractive dependencies of some kind (i.e., towards the previous response), there was no evidence for an attraction to the previous orientation. However, aside from an attraction to the previous response, such as documented in other literatures, it is conceivable that there was some third dependence, an attraction toward the previous orientation that existed in addition to the observed repulsion. Perhaps the balance of these two orientation dependencies was tipped more strongly toward repulsion by the experimental procedure (e.g., something about the presentations may have resulted in a strong tilt aftereffect that completely swamped any attraction towards the previous orientation). To assess whether these results hold more generally, the analyses were repeated on four published datasets, which cover a range of experimental paradigms, including ones that show strong attractive effects to the previous orientation when analyzed with standard methods (Fischer & Whitney, 2014; Pascucci et al., 2019; Samaha et al., 2019)[3].

In brief, all key patterns of Experiment 1 were also present in the published datasets. In the data from Samaha and colleagues, the split-half analysis revealed that the apparent bias toward the previous orientation is likely better explained by the summation of a tilt aftereffect and response hysteresis (Figure 1.9 bottom row, first column). In the remaining datasets, the split-half analyses did not reveal a clear tilt aftereffect (i.e., purple ribbons in the first three rows of the first column of Figure 1.9 straddled 0). It is not clear what to conclude from this apparent lack of difference for some of the split-half results, considering that when there are both attractive

[3]I thank Dr. Fischer and Dr. Samaha for providing me with their data, and Dr. Pascucci and colleagues for uploading their data publicly.

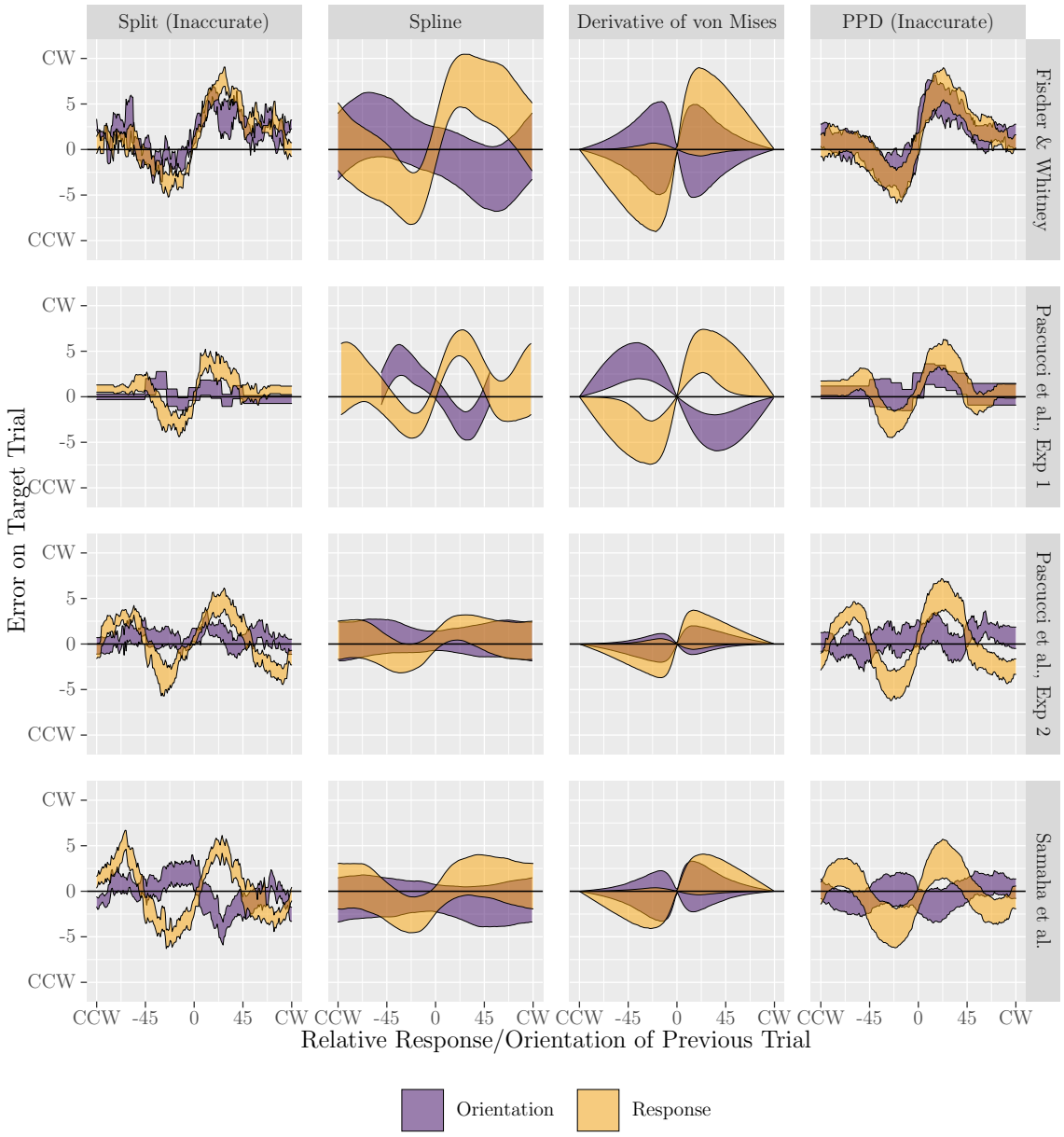


Figure 1.9: A tilt aftereffect and response hysteresis can account for prior studies of serial dependence. Each row gives a different dataset. **First Column**: Split half analysis of serial dependence, errors following inaccurate trials only. **Second Column**: Posterior distributions of the spline models. **Third Column**: Posterior distribution of the dependencies as estimated with the full von Mises model. **Fourth Column**: Posterior predictive distribution of full von Mises model, on trials that followed inaccurate responses (compare with first column).

and repulsive forces the split-half analysis will tend to underestimate any differences. The posterior distributions from the Bayesian models more cleanly uncouple the two effects, making use of all trials to measure response dependencies separate from orientation dependencies. As in Experiment 1, the non-parametric Bayesian model did not produce peripheral bumps for any of the four datasets (i.e., the curves in the second column of Figure 1.9 have only a single peak on each side of 0), indicating that a von Mises derivative model may be sufficient. The full von Mises model (which allowed for dependencies on both the previous orientation and response) suggested that the effect of the previous orientation was never attractive in any of the four datasets (Figure 1.9, third column). This full model fit all datasets well (compare first and fourth columns in Figure 1.9, which show the inaccurate split-half analysis of the empirical data and the posterior predictive distribution of the full von Mises model). Finally, model comparison never supported a model in which serial dependence reflected an attraction to the previous orientation (Figure 1.10).

1.6 Discussion

The context provided by recent stimuli and responses can impact perceptual-decisions. For instance, negative aftereffects demonstrate how recently encountered visual information repels perception, whereas the serial dependence effect has been interpreted as evidence that recent visual information may also attract perception. Nevertheless, the attractive dependence is variable, and its cause has remained unclear (Bae & Luck, 2020; Cicchini et al., 2017; Fischer & Whitney, 2014; Fornaciai & Park, 2018a; Fritsche et al., 2017; Manassi et al., 2018; Pascucci et al., 2019). Viewing a stimulus elicits a cascade of processes, many of which could account for an attractive dependence. Furthermore, effects at every level of this cascade are likely, with behavior reflecting a complex mixture (e.g., attractive serial dependence effects are likely weakened by concomitant tilt aftereffects). In this chapter, I pursued a

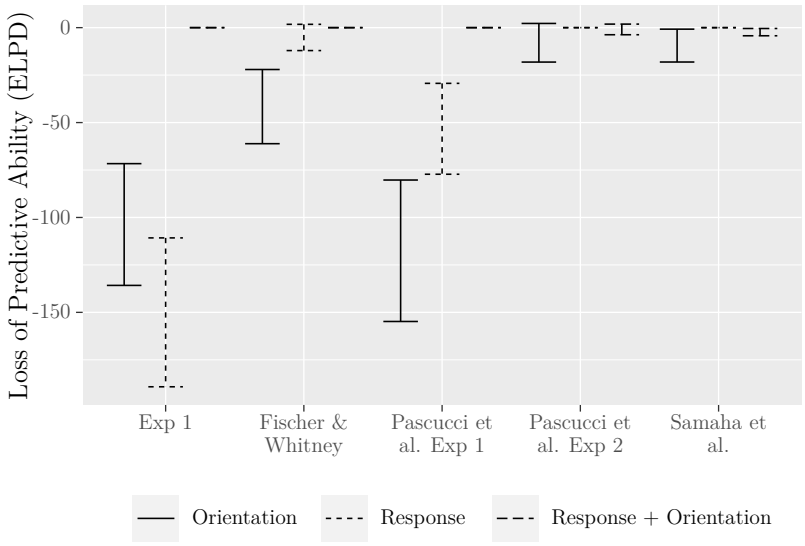


Figure 1.10: A model with an attractive effect of orientation was never the most predictive model. In each dataset, the three versions of the derivative of von Mises model were compared. The x-axis gives the relative expected log predictive density (ELPD, Appendix E) for each model, as compared to the most predictive model (i.e., the most predictive model is always 0). Error bars extend two standard errors of the mean of the difference.

statistical approach that allowed me to estimate effects that were more strongly coupled to either the previous stimulus or the previous response. In a new dataset, I found that an apparent lack of dependence can arise from two robust, latent forces working in opposition. I then applied this new statistical approach to four previously published datasets, two of which were discussed in their respective publications as showing an attraction toward recently encountered stimuli (Fischer & Whitney, 2014; Pascucci et al., 2019; Samaha et al., 2019), primarily because this was the manner in which the data were analyzed (i.e., as a function of previous orientation rather than the previous response). My *de novo* analyses suggest that the perceptual attraction reported in these prior studies was only apparent; there was no evidence of attraction to the previous stimulus after de-confounding the analysis from response dependencies. Instead, the orientation serial dependence effect, traditionally analyzed as a function of the relative difference between the current orientation and the previous

orientation, is a variable mix of repulsion from the previous orientation (i.e., a tilt aftereffect) and an attraction toward the previous response (i.e., response hysteresis), which are separate well-known effects that have been the focus of investigations prior to the development of the serial dependence paradigm.

The results presented in this chapter align with many authors who have argued against a perceptual account of serial dependence, instead proposing that the dependence reflects a mixture of two effects (Bae & Luck, 2020; Fornaciai & Park, 2019, 2020; Fritsche et al., 2017; Pascucci et al., 2019). Some of these authors have sought to falsify a perceptual account of serial dependence through experimental manipulation, such as by contrasting conditions in which participants are exposed to the inducing stimulus but do not respond (i.e., manipulating the presence of a decision, while holding perception constant; Appendix A). The novelty of this chapter lies in the analyses, which demonstrate that analyzing only one dependency at a time may be uninformative. As applied to Experiment 1, these analyses revealed that an apparently weak dependence, perhaps a lack of a dependence, masked two robust but latent dependencies working in opposition. My analyses of existing data from the literature revealed that even when there is an on-average robust attraction effect, that attraction is toward the previous response, which for many trials happens to track the previous orientation. It seems likely that much of the variability in the serial dependence effect may not be due to variability in attractive dependencies, but instead in the relative strengths of response hysteresis versus stimulus repulsion.

The two kinds of analyses developed here complement each other. First, the split-half analysis provided a straightforward, if only approximate, way to visualize whether inducing stimuli and responses to those stimuli may have had different effects on subsequent responses. However, this analysis will tend to underestimate any differences between an attractive and repulsive dependence, particularly in experiments that allow participants to respond with high accuracy. Statistical inferences

were based on a hierarchical Bayesian model fit to trial-by-trial data. Intuitively, it may seem that the parameters of the model are non-identifiable; if an attractive and repulsive effect mirrored each other exactly, then increasing the amplitude of one could be exactly counterbalanced by reducing the amplitude of the other. But this intuition is misleading. To understand why the effects are identifiable, consider that even if the two effects were perfect mirrors (i.e., the widths were the same and the amplitudes were equal in magnitude but opposite in sign), the inputs to these effects, the relative orientation and response of the previous trial, would be the same only if participants were perfectly accurate. But participants are never perfectly accurate, so on each trial the two effects tend to be slightly offset from each other (i.e., the position on the x-axis for one effect on a particular trial will be slightly different than the position on the x-axis for the other effect for that same trial), and this offset varies from trial to trial. This variability in offset is what allows the Bayesian model to estimate the parameters of each effect, using the whole dataset. This offsetting is the same mechanism that allows the split-half analysis to partially untangle two effects after only the relatively inaccurate trials.

One previous study has developed a similar two-process model of serial dependence, incorporating both a repulsive and attractive effect (Pascucci et al., 2019). With their model, Pascucci and colleagues have provided a specific theoretical proposal regarding the psychological processes underlying serial dependencies. In this chapter, I have been more concerned with describing the effect to be explained, that is, with accurately characterizing the trial-wise dependencies that were only latent in the data. I showed that the standard way to analyze the serial dependence effect, a plot of errors based on either the previous orientation or the previous response, will not generally be informative about the latent dependencies affecting participants. The typical analysis strategy can be misleading because different combinations of dependencies could give rise to similar patterns of data (e.g., a lack of dependence

on the previous orientation as revealed by a typical analysis could arise from either a true lack of dependence, or a dependence that is counterbalanced by an attraction toward an accurate response). Pascucci and colleagues have provided a model offering one set of mechanisms whose summation matches the observed data, which is itself a summation of latent attractive and repulsive dependencies. It remains to be seen how well their model can capture the trial-by-trial dependencies that have been isolated in this chapter.

Demonstrating the utility of the purely descriptive two-factor approach developed here, as applied to trial-by-trial data, these results also indicate a novel explanation of the often-observed peripheral bumps. Specifically, the non-parametric spline model, which could have produced peripheral bumps separately for each of its factors, revealed that when the average data actually contain peripheral bumps, this arises from the summation of distinct attractive and repulsive dependencies that do not contain peripheral bumps (see also Alais et al., 2017). This is seen in the bottom row of Figure 1.9 in which the split-half data show peripheral bumps (First column), but neither component of the spline model contained such bumps (adjacent panel). Somewhat surprisingly, the von Mises model captured these peripheral bumps (Fourth column) even though the von Mises cannot possibly contain such bumps (see also the middle row of Figure 1.5). Of note, the inaccurate trials are not uniform in their inaccuracy; some trials are less accurate than others even within this split-half, and, furthermore, the nature of the inaccuracies may be systematic, such as with the "oblique effect" (Appelle, 1972; Jastrow, 1892). The hierarchical Bayesian model included parameters to account for these biases on each target trial, but the model does not attempt to explain the nature of previous trial accuracy, instead treating it as observed data that constrain predictions for subsequent trials. Thus, in generating the on-average next trial predictions (Fourth column), the mixture of previous trial inaccuracies allowed the model to conjure apparent peripheral bumps seemingly out of thin air (i.e.,

compare the sinusoidal pattern in the fourth column to the von Mises functions that produced these predictions, as seen in the third column).

Although the presented analyses revealed a lack of evidence for attraction toward recent stimuli, they do not pinpoint which psychological process produces the attractive dependence. Under the assumptions that there are multiple stages of processing between sensing the stimulus and making a response, and that the order of these processes will determine how closely they are coupled with either the inducing stimulus or the response to that stimulus (Figure 1.4), the attraction seems to be toward a process that is "further along" the cascade those that cause the tilt aftereffect. But I remain agnostic about what stage of processing produces the attraction – agnostic about even whether the attraction is caused by a post-perceptual process. It may be that participants' responses provide a better index of their higher-level stages of perception than do the veridical orientations that they sensed. That is, if participants experienced a perceptual illusion while responding to the inducing orientation (i.e., they "saw" an orientation that was not presented, see also the next paragraph), and then reported that illusory orientation exactly, then the attraction toward their response could reflect an attraction toward their (false) perception of the inducer (see also Fischer & Whitney, 2014, Experiment 1c; St John-Saaltink et al., 2016). Alternatively, participants may have perceived the inducing orientation accurately but responded inaccurately due to noise in mnemonic[4] or motoric processes. In that case, the attraction could reflect an attraction toward their decisions about the inducer. The point of the presented results was not to adjudicate between such accounts. Instead, the point is that further research into such questions should take care to account for at least the two dependencies isolated in this chapter.

[4]Even in Experiment 1, in which participants could initiate their response when the target was first presented, the inducer was almost always replaced by a mask before participants completed their response.

The results presented in this chapter were designed to untangle potential causes of serial dependence, but they are silent about an important effect of that attraction: it remains unclear whether the attraction occurs "at the time of perception" or at some time during post-perceptual processing (Bliss et al., 2017; Cicchini et al., 2017; Fischer & Whitney, 2014; Fritsche et al., 2017; Manassi et al., 2018). Even if one accepts that participants' errors are attracted toward their previous responses, the attractive dependence may cause participants to be more prone to misperceive the target orientation. Put another way, is the attraction a genuine perceptual illusion, or are participants reporting orientations that they did not perceive? Fischer and Whitney (2014) proposed that serial dependence was a genuine illusion. Like research into of the cause of serial dependence, research into the effects of the dependencies have given conflicting results. For example, one operationalization of this question is to ask whether the attraction is present when participants first view the target stimulus (Bliss et al., 2017; Manassi et al., 2018). However, a perceptual decision may not be neatly delineated into sequential phases of perception and decision; a participant can make decisions about their perception looking at the stimulus, and participants can adopt biased response patterns even before seeing the stimulus. Like the analyses explored here, attraction to the previous response may affect both perceptual and post-perceptual processes simultaneously.

Finally, the presented analyses do not necessarily contradict process models that have emerged as accounts of the serial dependence effect, allowing that some of them may only be partial accounts (Alais et al., 2018; Cicchini et al., 2018; Kalm & Norris, 2018; Pascucci et al., 2019; van Bergen & Jehee, 2019). As mentioned above, Pascucci and colleagues presented a model with two competing forces, like those inferred in this chapter. One account proposes that the effect reflects suboptimal behavior – a failure to appropriately adapt to the randomness of the stimuli on each trial (Kalm & Norris, 2018). Others point to functional benefits afforded by serial dependence;

by basing responses to the target at least partially on information remembered about the inducer, participants may be slightly biased, but the added information results in responses that are less variable, and participants could tailor this bias-variance trade-off to the perceived informativeness of the inducer (Alais et al., 2018; Cicchini et al., 2018; van Bergen & Jehee, 2019). Such models may still account for the attractive dependence inferred in this chapter. More importantly, I suggest that, since the serial dependence effect may not be a novel phenomenon, further research into the mechanisms of serial dependencies – both attractive and repulsive – could leverage older literatures on intertrial and interstimulus effects (see also Kiyonaga et al., 2017). This includes research into assimilation versus contrast in psychophysics (e.g., McKenna, 1984), or positive versus negative priming (e.g., Huber et al., 2002; Jacob et al., In Preparation). Knowing that at least two such effects exist, and knowing that different experimental variables may affect each effect separately – producing a wide range of possible results when analyzing average data plotted solely as a function of prior stimulus or response – highlights the need to consider multiple kinds of dependencies that collectively and simultaneously constrain behavior.

1.7 Summary

The serial dependence effect has been interpreted as an attraction to recently encountered stimuli. Such an effect could provide valuable new insight into how perceptual-decisions leverage the recent past, given that it would be exactly the opposite of the tilt aftereffect. However, the amplitude of the effect has been variable. I have pointed out that much of the variability in the serial dependence effect may arise, not necessarily from any differences in attractive dependencies, but instead from the relative strength of attractive and repulsive dependencies that act simultaneously. I have shown that, if there are multiple, latent dependencies present, then neither the magnitude nor the sign of these dependencies will be revealed by looking at only one

dependency at a time (e.g., a plot of the dependence of errors on just the inducing orientation). I developed two analyses that can untangle these latent dependencies. As applied to a new dataset and four published datasets, the analyses revealed a lack of evidence that participants' responses were attracted toward recently encountered stimuli. Instead, the serial dependence effect could be accounted for with a mixture of two known processes: a tilt aftereffect and response hysteresis.

An attraction to the inducer would be interesting, but that attraction would necessarily be latent, operating at the same levels as other kinds of dependencies. To build a more complete picture of how perceptual decisions integrate information across time requires that these latent dependencies be considered explicitly. In this chapter, I have shown how this consideration can be given by explicit modeling, even when the models are descriptive. I complemented the modeling with a visualization – the split-half analysis – to provide a simple justification for the relatively complex model. I will use the same approach of descriptive modeling and visualization in the next chapter, showing how it can be leveraged to learn about neuromodulation in humans.

Table 1.1: Protocols for studies of serial dependence in orientation judgments.

Study	*Experiment*	*Inducer Duration*	*Attention to Inducer*[a]	*Inducer Eccentricity*	*Inducer-Target Position*	*Inducer Contrast*	*Inducer SF*	*Orientation Range*	*Mask Duration*	*Mask-to-Response Delay*	*Response to Inducer*	*ITI Average*[b]	*Trials*	*Amplitude*[c]
Fischer Whitney (2014)	1	500	True	6.5	Same	25	0.33	All	1000	250	True	2000	1040	8.19
Fischer Whitney (2014)	1b[d]	500	True	6.5	Same	25	0.33	All	1000	250	True	2000	3296	4.85
Fischer Whitney (2014)	1c[d]	500	True	6.5	Same	25	0.33	All	1000	250	True	2000	816	3.12
Fischer Whitney (2014)	2	500	True	6.5	Same	25	0.33	All	1000	250	False	2000	416	6.76
Fischer Whitney (2014)	2	500	True	6.5	Same	25	0.33	All	1000	250	True	2000	416	8.75
Fischer Whitney (2014)	4	500	True	9	Same	25	0.33	All	1000	250	True	2700	312	11.00
Fischer Whitney (2014)	4	500	True	9	Different	25	0.33	All	1000	250	True	2000	312	6.00
Fischer Whitney (2014)	4	500	False	9	Different	25	0.33	All	1000	250	False	2000	312	-4.00
Fischer Whitney (2014)	5	500	True	<12.5	Close	25	0.33	All	1000	250	True	2000	200	10.00
Fischer Whitney (2014)	5	500	True	<12.5	Far	25	0.33	All	1000	250	True	2000	200	3.00
Fischer Whitney (2014)	5	500	True	<3	Close	25	0.33	All	1000	250	True	2000	268	8.46
Fischer Whitney (2014)	6		True	6.5/19.5	Same[g]						True		720	7.00
Fischer Whitney (2014)	6		True	6.5	Retinal[g]						True		720	5.00
Fischer Whitney (2014)	6		True	6.5/19.5	Spatial[g]						True		720	4.00
Fischer Whitney (2014)	6		True	6.5/19.5	Different[g]						True		720	3.00
Fischer Whitney (2014)	7[e]	5000	True	6.5	Same	25	0.33	All	1000	0	False	0	1236	-5.00
Liberman Zhang Whitney (2016)	1	570	True	Moving	Congruent	29	4	All	871[h]	0	False	2500	3080	3.00
Liberman Zhang Whitney (2016)	1	570	True	Moving	Incongruent	29	4	All	871[h]	0	False	2500	3080	1.00
Liberman Zhang Whitney (2016)	2	610	True	Periphery[f]	Congruent	29	4	All	885[h]	0	False	2500	3080	0.75
Liberman Zhang Whitney (2016)	2	610	True	Periphery[f]	Incongruent	29	4	All	885[h]	0	False	2500	3080	0.70
Fritsche Mostert De Lange (2017)	1	500	True	11.18	Same	25	0.33	All	1000	250	True	2700	9696	1.15
Fritsche Mostert De Lange (2017)	1	500	True	11.18	Different	25	0.33	All	1000	250	True	2700	9696	1.17
Fritsche Mostert De Lange (2017)	4	250	True	6.5	Same	25	0.33	All	250	50	True	4700	14544	1.08
Fritsche Mostert De Lange (2017)	4	250	True	6.5	Same	25	0.33	All	250	3500	True	1250	14544	1.64
Cicchini Mikellidou Burr (2018)	1	500	True	8.94	Same	25	0.3	Cardinal	1000	0	True	500	1680	0.50[i]
Cicchini Mikellidou Burr (2018)	1	500	True	8.94	Same	25	0.33	Oblique	1000	0	True	500	1680	1.25[i]
Cicchini Mikellidou Burr (2018)	1	500	True	8.94	Same	25	1.2	Cardinal	1000	0	True	500	1680	2.50[i]
Cicchini Mikellidou Burr (2018)	1	500	True	8.94	Same	25	1.2	Oblique	1000	0	True	500	1680	2.00[i]
Cicchini Mikellidou Burr (2018)	2	500	True	8.94	Same	25	0.2	Cardinal	1000	0	True	500	6000	1.50
Cicchini Mikellidou Burr (2018)	2	500	True	8.94	Same	25	0.3	Oblique	1000	0	True	500	6000	2.00
Bae & Luck (2019)[j]	1	200	True	1.09	Same			All	0	1300	True	1500	10240	-3.00
Bae & Luck (2019)[j]	2	200	True	2.17	Both			All	0	1300	True	1500	10240	-1.00
Pascucci et al. (2019)	1	400	True	8.5	Both	50	0.5	All	400	500	True	500	5600	1.32
Pascucci et al. (2019)	2	400	True	0	Same	50	1.2	All	400	500	True	500	4400	1.40
Pascucci et al. (2019)	3	500	True	6.5	Same	50	0.5	All	1000	500	True	500	5040	1.76

continued

Table 1.1 – Continued from previous page

Study	Experiment	Inducer Duration	Attention to Inducer[a]	Inducer Eccentricity	Inducer-Target Position	Inducer Contrast	Inducer SF	Orientation Range	Mask Duration	Mask-to-Response Delay	Response to Inducer	ITI Average[b]	Trials	Amplitude[c]
Pascucci et al. (2019)	3	500	True	6.5	Same	50	0.5	All	1000	500	False	500	3360	-0.82
Pascucci et al. (2019)	6	1000	True	6.5	Same	75	0.5	All	500	500	True	500	3000	2.32
Pascucci et al. (2019)	6	1000	True	6.5	Same	25	0.5	All	500	500	True	500	3000	0.20
Samaha Switzky Postle (2019)	1	33	True	0	Same	8.5/4.25[l]	1.5	All	0	6300	True	700	6000	2.30
Van Bergen Jehee (2019)	1	1500	True	1.5-7.5[k]	Same	10	1	All	0	6500	True	4000	4536[m]	1.50

[a] Attention recorded as False if participants were explicitly instructed to attend to something beside the inducer and otherwise True.

[b] In some experiments 'Average ITI' includes both the ITI and a period of cueing at the start of the target trial. In all experiments excluding van Bergen and Jehee (2019), the reported ITI does not include participants' response time (in the excluded experiment, the duration of response periods were fixed).

[c] Numeric amplitudes were not reported in the following experiments: Fischer Whitney Experiments 4, 5, 6, and 7; Liberman Zhang Whitney experiments 1 and 2; Cicchini et al. Experiments 1 and 2; Bae Luck Experiments 1 and 2. In these experiments, amplitudes were inferred from graphs.

[d] In their paper, 1b and 1c were listed as control experiments. In Experiment 1, the orientations on each trial were explicitly counterbalanced, whereas in 1b they were randomized. In 1c, on some trials participants were additionally asked whether the inducing orientation was more clockwise or counterclockwise than the target.

[e] Each trial contained both the inducing and target stimuli, with only a 1000 ms mask in between. Trials were separated by 2000 ms.

[f] Numeric value not reported.

[g] Fixation points varied. Targets could be in the same spatial position as the inducer, the same retinal position, both, or neither.

[h] Mask duration refers to the duration for which the inducer was occluded. The target was additionally masked for 1000 ms.

[i] These values are only partially analogous to other reported amplitudes. These authors used only a restricted range of orientation differences. The reported amplitudes are an average of the two differences that elicited the largest biases (one CW and one CCW, always symmetric).

[j] Stimuli were neither gratings nor Gabors, but instead black, 'teardrop' shapes.

[k] Values describe an annulus stimuli.

[l] In this experiment, the gratings were convolved with white noise. The reported amplitude averages across two conditions. The conditions kept the signal-to-noise ratio of the stimuli equal, but altered the luminance amplitudes. The luminance was calibrated to each participant. 8.5 % reflects the average, calibrated contrast. These gratings were convolved with noise rendered a 1000 % contrast. In the other condition, the contrasts of both signal and noise were halved.

[m] Participants completed different numbers of trials. This count reflects the average of the maximum and minimum number of trials across participant.

Note: Eccentricity: distance in degrees visual angle from fixation to the center of the stimuli; Contrast: Percent Michelson; SF: Spatial Frequency, cycles per degree; ITI: Inter-trial Interval; Empty cells imply either that the value was not reported or not applicable. All timing is reported in milliseconds.

CHAPTER 2

INFERRING MODULATIONS TO NEURAL TUNING FUNCTIONS FROM VOXEL TUNING FUNCTIONS

2.1 Abstract

In this chapter, I present a method for exploring the neural underpinnings of the kinds of sensory and perceptual phenomena studied by cognitive psychologists and neuroscientists. To illustrate the need and motivation for this new method, I begin by describing currently unresolved questions about the neural mechanisms that underlie attractive serial dependence and show how these could be answered with the new method. These mechanisms rely on changes to either the tuning properties of neurons or the readout of these tuning functions. However, using current techniques, such changes cannot be identified with the neuroimaging methods available in humans. To overcome this problem, I present a novel analysis technique for decomposing functional magnetic resonance imaging (fMRI) data into component tuning functions, characterizing the nature of neuromodulation. I end the chapter by revisiting how the technique could be used to uncover the neural mechanisms underlying serial dependence.[1].

[1]Sections of this chapter, particularly the Introduction and Discussion, overlap with two papers that are already in preparation. One paper, (Sadil, Huber, & Cowell in prep) presents the novel, Bayesian analysis of voxel tuning functions. The other paper (Cowell, Sadil & Huber, in prep) discusses specific inadequacies of one currently available analysis technique, the inverted encoding model.

2.2 Neural Accounts of Serial Dependence

Many neurons in mammalian cortex respond selectively to certain features of stimuli (e.g., orientation, color, pitch), exhibiting a pattern of activity that is referred to as a "tuning function" (e.g., Hsiao et al., 2002; Hubel & Wiesel, 1959; Maunsell & Van Essen, 1983). A neuron with a tuning function will be said to be "tuned" to that feature, "preferring" whichever feature value elicits the most activity (e.g., a horizontal orientation). These tuning functions are enticing since they may be a basis for some neural codes (e.g., Gold & Shadlen, 2002; Wenderoth & Johnstone, 1987).

Populations of tuning functions have been used to model behavior (e.g., Grill-Spector et al., 2006; Schwartz et al., 2007; Webster, 2015). These models have two stages (Figure 2.1). The first stages assumes a group of neurons that are each tuned to the same stimulus feature (e.g., orientation, color, pitch). The tuning functions tend to be bell-shaped. The second stage characterizes how the output of these tuning functions are combined into a single decision. Decisions are modeled by treating each tuning function as a source of evidence for a single feature value, as though each neuron "votes" for, e.g., a particular orientation. In Figure 2.1A, neurons vote in proportion to their level of activity, although those votes can also be weighted by a separate function. In the simplest case, neurons only vote for their preferred orientations. The final response is given by tallying the votes, subject to some response variability (e.g., Gaussian noise around the decided upon orientation)[2]. The population depicted in Figure 2.1A would tend to vote for a stimulus value very close to whatever stimulus value is presented, i.e., it would accurately represent the sensory input.

[2]More complex versions of these models may incorporate more sources of variability, such as variability in the stimulus or variability in the tuning function (e.g., Dosher & Lu, 1998; Wei & Stocker, 2015).

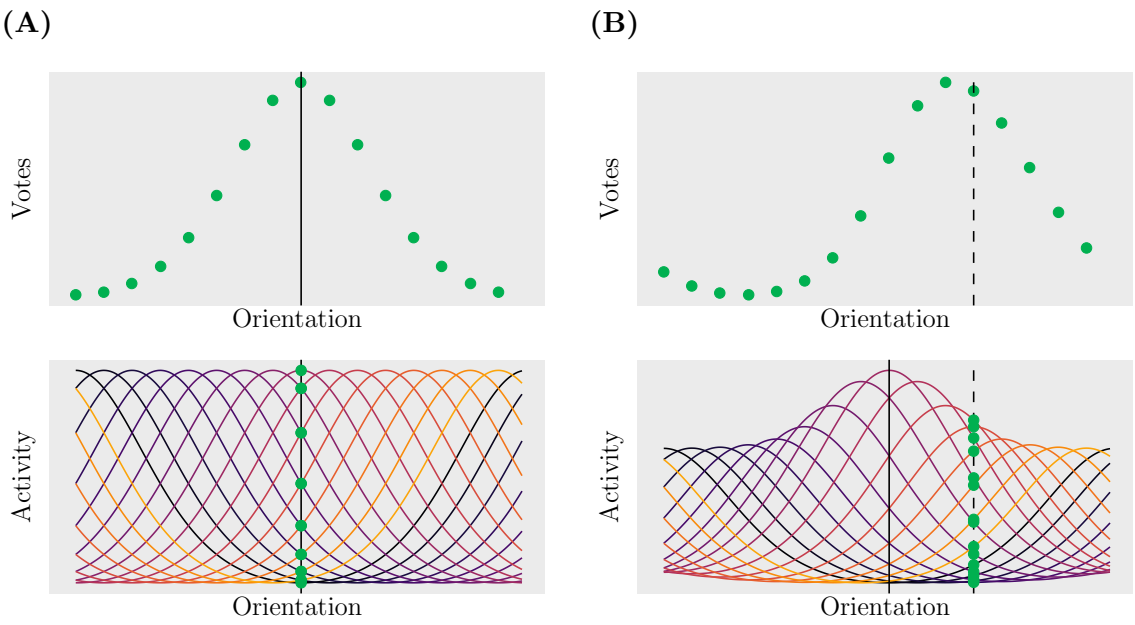


Figure 2.1: Population Coding Model. The bottom panels represent the sensory representation of the stimulus and the top panels represent how this representation is transformed into a decision. In the bottom panels, each colored line represents a neural tuning function. The vertical lines represent stimuli (solid: inducing stimulus, dashed: target). A response is modeled by allowing for each neuron to vote for some orientation. In these cases, the strength of the vote is proportional to the neuron's level of activity (green dots), and the neurons vote for their preferred orientations (c.f., Figure 2.2B). The top panels depict these votes as a function of each orientation. **(A)** Unmodulated, unbiased responding. **(B)** "Gain" mechanism, resulting in an attractive dependence (Fischer & Whitney, 2014). Other forms of neuromodulation have also been proposed to account for attractive dependencies (Figure 2.2).

Many behavioral effects, such as serial dependence, have been accounted for with modulations to the population of tuned neurons (e.g., Fischer & Whitney, 2014; Papadimitriou et al., 2017; Pascucci et al., 2019). For example, perhaps the first stimulus increases the sensitivity of neurons tuned to nearby orientations, such that they are excessively active on the target trial (Figure 2.1B). Under this "gain" hypothesis, neurons that prefer the inducing orientation will respond with greater magnitude than normal when the target stimulus is presented (Fischer & Whitney, 2014). Depending on the magnitude of this gain modulation, the widths of the tuning functions, and the difference between the inducing and target orientations, the increased gain causes the most responsive neuron to be the one that prefers an orientation intermediate

between the inducing and target stimulus, biasing the votes in favor of an orientation intermediate between the inducer and target. Note that this kind of neural model differs from the descriptive models of the previous chapter (e.g., the derivative of von Mises models), in that the descriptive model aims to account for the pattern of errors to be explained, whereas these neural models aim to offer an explanation with neural mechanisms; these neural-level models are successful only if they can produce response patterns that resemble the models from the previous chapter.

In addition to the gain hypothesis, serial dependence has been explained with three other forms of neuromodulation (Fischer & Whitney, 2014; Papadimitriou et al., 2017; Pascucci et al., 2019). The efforts to tell a neural-level story of serial dependence underscores that it can be difficult to tell a neural-level story of a behavioral effect without neural data. To clarify this difficulty, I will next describe the other three mechanisms (Figure 2.2). The mechanisms include neuromodulation in the tuning functions as well as modulation to how the votes are aggregated across the population (i.e., the gray shadows in Figure 2.2)[3].

Fischer and Whitney (2014) proposed two mechanisms that could elicit attractive serial dependence. For both mechanisms, the evidence is weighted equally across tuning functions (e.g., a flat gray shadow indicates equal weighting, resulting in an unbiased integration of evidence). The gain hypothesis was discussed above. The second proposed mechanism relies on channels shifting their preferred orientations (Figure 2.2A). Under a "repulsive shift" hypothesis, the tuning functions that peak around the inducing orientation are repelled away from that orientation. After shifting, the function that peaks at the target orientation will be one that previously peaked at an orientation which was intermediate to the inducer and target. Although the tuning functions shift, neurons still vote for whichever orientation they originally

[3]The models additionally include some process that causes the modulation to decay across trials, such that, e.g., the modulation induced by trial n has only a minimal effect on trial $n + 10$.

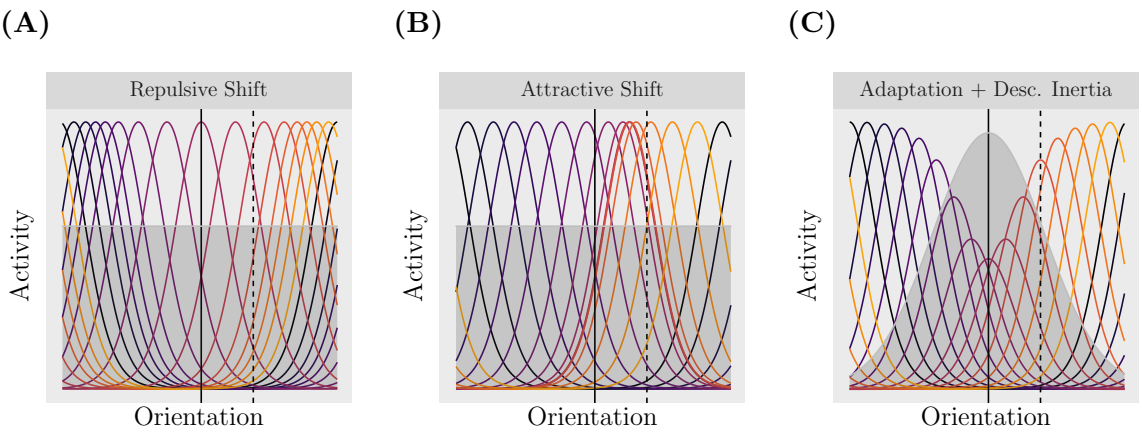


Figure 2.2: Serial dependence could arise from many kinds of neuromodulation (Fischer & Whitney, 2014; Papadimitriou et al., 2017; Pascucci et al., 2019). Subfigures show different kinds of neuromodulation (see also Figure 2.1B). Tuning is depicted as in Figure 2.1, but in some mechanisms the votes are additionally weighted, with the weights depicted by gray shadows. The tuning and the weighting functions may be modulated by the inducing orientation (solid line). With each modulation, the average response to the target orientation (dashed line) will be intermediate to the inducer and target, resulting in an attractive dependence. See main text for a description of each kind of neuromodulation.

preferred, again resulting in the participant endorsing an orientation intermediate to the inducing and target stimuli.

Whereas the models presented by Fischer and Whitney (2014) incorporated just a single mechanism, Pascucci et al. (2019) proposed that serial dependence arises from both altered neural tuning functions and an altered readout process (Figure 2.2C). Their model was designed to account for the two kinds of dependencies that their experimental manipulations revealed observed, an attraction and a repulsion. The model captures repulsive biases by decreasing the gain on the tuning functions centered on the inducing orientation, resulting in a tilt aftereffect (c.f., Magnussen & Kurtenbach, 1980). Their model captures attractive biases by increasing the weight given to neurons that prefer feature values closest to the inducing stimulus, an increase they call decisional inertia (Akaishi et al., 2014). The inertia allows the evidence provided by neurons whose gain has been reduced to still strongly influence the decision, attracting responses to the response made about the inducing orientation.

Although neurally inspired (Dragoi et al., 2000; Maunsell & Treue, 2006; Patterson et al., 2013; Wissig & Kohn, 2012), the models of Fischer and Whitney and Pascucci and colleagues have not yet been constrained by neural data. One electrophysiological study linked attractive serial dependence with modulated tuning functions (Papadimitriou et al., 2017), but the modulation did not match any of the three mechanisms outlined above. In that study, neurons in the frontal eye fields of macaques were recorded while they performed a memory-guided saccade task. On each trial, a location was identified with a dot, and the macaque was required to remember that location throughout a short retention interval (during which the dot was no longer visible). After retention, the macaque needed to saccade to the remembered location. Their saccades exhibited an attractive serial dependence, such that the monkeys tended to saccade to locations between the last two remembered locations (i.e., between the locations that were cued on the current and previous trial). The tuning functions – spatial receptive fields – of neurons in the frontal eye fields shifted their preferred locations in a manner that could be related to the serial dependence, but the mechanism was more complex than the shift proposed by Fischer and Whitney. More specifically, rather than a repulsive shift (as seen in Figure 2.2A), Papadimitriou et al. (2017) showed how a relatively strong attractive shift toward the target stimulus (Figure 2.2B), paired with a lingering shift toward the inducing stimulus, could lead to attractive dependence (Zirnsak et al., 2014). Unlike in the repulsive shift mechanism proposed by Fischer and Whitney (2014), Papadimitriou and colleagues proposed that, in addition to a shift in tuning, neurons also changed how they voted. In essence, this can be thought of as over-representing recently viewed stimuli, with the net effect of an attractive shift and a voter preference shift, being something akin to gain increases for recently viewed stimuli. Such a model reconciles the observed frontal eye field attractive shifts with the observed serial dependence response attraction.

The results from Papadimitriou et al. (2017) provide a neural account of attractive serial dependence in judgments of spatial position. But it does not yet provide a neural basis for the repulsive dependencies in judgments of spatial position that have also been observed (Bliss et al., 2017), nor is it clear that the same mechanisms will apply to judgments of orientation. Given the data so far, it appears that several different mechanisms can plausibly account for serial dependence in orientation judgments. Furthermore, given that viewing orientations can modulated orientation tuning in many ways (Bharmauria et al., 2019; Dragoi et al., 2000; Patterson et al., 2013; Wissig & Kohn, 2012), neural data may prove necessary to adjudicate between these mechanisms. But although additional, invasive single-cell recordings might differentiate between these models, is it also not clear whether the results from monkey studies would generalize to humans, considering that monkeys typically receive extensive training prior to recording. To study these mechanisms in humans, functional magnetic resonance imaging (fMRI) can be used, but existing fMRI analysis techniques cannot determine the form of neuromodulation, considering that fMRI measurements reflect the metabolic activity of many thousands of neurons rather than the electrical response of individual neurons. In this chapter, I address this final issue, uncovering the form of neuromodulation given fMRI measurements. I develop a novel analysis technique that aims to link neural models of serial dependence (and indeed neural models of many other perceptual or cognitive phenomena) with neural data in humans.

2.3 A Technique for Determining Neuromodulation of Tuning Functions from fMRI Data

Although the spatial resolution of fMRI is relatively high among non-invasive neuroimaging techniques, the resolution is still orders of magnitude beyond that of a single neuron: a patch of striate cortex the size of a typical voxel ($2\,\mathrm{mm}^3$) reflects

the activity of approximately 300,000 – 500,000 neurons (Goense & Logothetis, 2008; Leuba & Garey, 1989). Nonetheless, promising studies have revealed feature-selective tuning in voxels, derived via both univariate and multivariate analyses of human fMRI data (Dumoulin & Wandell, 2008; Engel et al., 1997; Kamitani & Tong, 2005; Serences et al., 2009). Voxel tuning functions for features such as orientation are presumed to derive at least indirectly from the tuning of individual neurons: different neurons have different preferences, and the distribution of preferences of the neurons contributing to a voxel is not uniform, resulting in voxel preferences (Boynton, 2005). Like neural tuning functions, voxel tuning functions are modulated by manipulations of perceptual or cognitive state (Saproo & Serences, 2010; Scolari & Serences, 2010; Serences & Saproo, 2010; Serences et al., 2009; Zhang et al., 2010), implying the same promise for empirical investigations of the neural mechanisms of perception and cognition.

Indeed, the superficial resemblance of voxel tuning functions to neural tuning functions makes it tempting to draw a direct analogy, inferring properties of neural-level tuning from voxel-level observations. But, as several researchers who exploit voxel tuning to examine the brain's perceptual representations have acknowledged, we cannot (Çukur et al., 2013; Nevado et al., 2004; Serences et al., 2009; Sprague et al., 2018). The relationship between voxel tuning functions and neural tuning functions presents an ill-posed inverse problem: a voxel's tuning function is determined by multiple unknown factors, including the shape of the underlying neurons' tuning functions and the distribution of neurons across different preferred stimulus values (e.g., the number of neurons within a voxel that prefer a vertical orientation, as opposed to horizontal or oblique angles). Because there are many possible combinations of these unknown factors that could give rise to the same observed voxel tuning function (a many-to-one mapping), any tuning observed in the fMRI data, whether derived from

univariate or multivariate analyses, cannot necessarily be taken as a proxy for tuning at the neural-level (Sprague et al., 2018; c.f., Keliris et al., 2019; Nevado et al., 2004).

Providing a partial solution to this inverse problem, I developed a novel modeling procedure that can draw qualitative conclusions about neural-level response properties from the voxel-level, Blood Oxygenation-Level Dependent (BOLD) signal (Ogawa et al., 1992; Ogawa & Lee, 1990). My goal was not to derive quantitative parameterizations of neural tuning functions (e.g., characterizing the exact shape of tuning functions for individual neurons or neural-subpopulations) but to characterize modulations of tuning functions by manipulations of cognitive or perceptual state. In short, given fMRI data from two experimental conditions across which neural tuning curves are modulated (e.g., for orientation tuning, low versus high-contrast visual stimuli), the procedure uses model comparison to infer the most likely form of modulation at the neural level.

I developed a hierarchical Bayesian modeling procedure that uses only BOLD-level data to infer the form of neuromodulation that accompanies a change in perceptual state (e.g., modulations in orientation tuning that are produced by an increase in stimulus contrast). Similar to previous encoding models that exploit tuning (Brouwer & Heeger, 2009; Kay et al., 2008; Mitchell et al., 2008), the procedure assumes that each voxel tuning function emerges from a linear combination of neural tuning functions (Figure 2.3). Given this assumption, I derived a mathematical relationship between the voxel and neural tuning functions. I used this relationship to construct models that each provide alternative accounts of how neuromodulation could lead to changes to the BOLD data across a change in perceptual state (Figure 2.4). Next, I used model comparison to adjudicate between the alternative accounts of neuromodulation. This procedure demonstrates that, even though the voxel tuning function reflects the aggregated activity of hundreds of thousands of neurons, measurable changes in

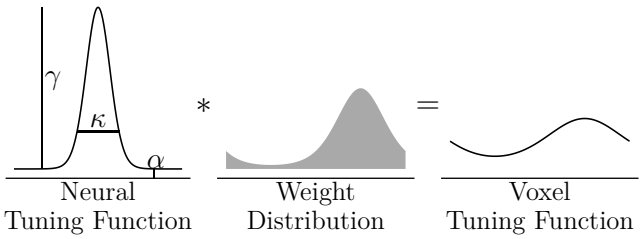


Figure 2.3: Convolution ($*$) of neural tuning functions with weight distribution leads to voxel tuning functions. The proposed method uses the observed voxel tuning functions to estimate parameters related to neural tuning functions and the distribution of those tuning functions within each voxel.

voxel tuning functions can be used to uniquely specify the form of neuromodulation underlying those changes.

Providing a test case where "ground truth" was known from single-cell recording studies, I examined fMRI orientation tuning functions in visual cortex under conditions of high and low stimulus contrast. This verification of the technique was a necessary first step before it can be applied to situations where the form of neuromodulation is unknown, such as with the various possible tuning function changes that might underlie serial dependence effects. It is known from electrophysiology that changes in visual contrast induce multiplicative scaling in neural tuning functions (Alitto & Usrey, 2004; Sclar & Freeman, 1982; Skottun et al., 1987), and the question asked was whether this ground truth could be recovered. I used the new hierarchical Bayesian modeling procedure, along with model comparison, to assay how well different forms of neuromodulation (e.g., multiplicative scaling and additive shift) could account for the effect of stimulus contrast on the BOLD data. Model comparison revealed that, in line with electrophysiological data, multiplicative scaling of the underlying neural tuning functions best accounted for the changes in voxel-tuning measured with fMRI. This recovery of ground truth occurred even though the empirical result for the majority of the individual voxels was an additive shift.

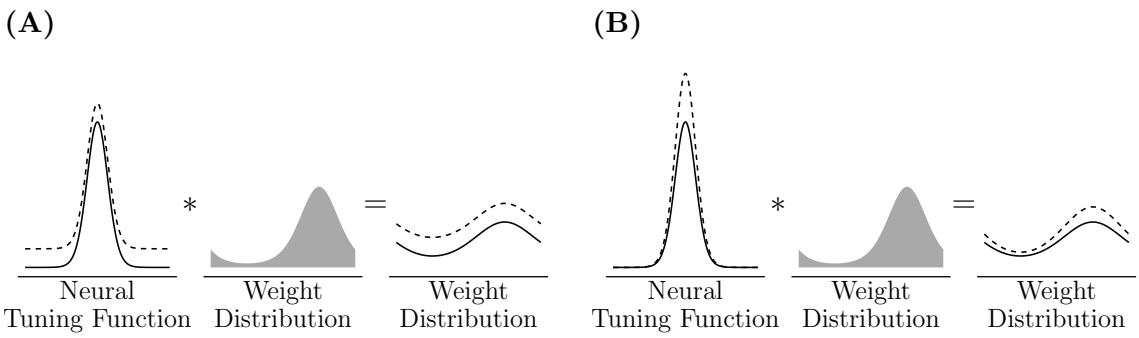


Figure 2.4: Modulations to the neural tuning functions produce distinct modulations to the voxel tuning functions. **(A)** Multiplicative gain in the neural tuning function causes multiplicative gain in the voxel tuning function. **(B)** If the neural tuning function undergoes an additive shift, so will the voxel tuning function. Compare with Figure 2.3 and Equation 2.4.

2.4 Methods

2.4.1 Participants

Seven participants (22 – 31 years old; 3 females, 2 did not report) each completed three sessions and received monetary compensation ($50 per 2-hour session). All participants had normal or corrected-to-normal vision. One additional participant completed a single session but exhibited substantial motion; their data were excluded from analyses. The procedure was approved by the University of Massachusetts Institutional Review Board.

2.4.2 Behavioral Stimulation and Recording

Behavioral Stimulation and Recording Stimuli were presented to participants with a gamma corrected 32″ LCD monitor at 120 Hz refresh rate (Cambridge Research Systems). The experiment was designed using the Psychophysics toolbox (Version 3.0.14; Brainard, 1997; Pelli, 1997) and custom MATLAB code (2018b, MathWorks). Behavioral responses were collected with a button box (Current Design). Eyetracking data was recorded at a rate of 1000 Hz with the Eyelink 1000 Plus system on a long range mount (SR Research), and the eyetracker was controlled using the Eyelink

Toolbox extension to Psychtoolbox (Cornelissen et al., 2002). Due to a technical error, the responses of one participant were not recorded.

2.4.2.1 Orientation Scans

Participants completed 18 functional runs across the three sessions, to collect the primary experimental data (Figure 2.5), in addition to some other functional runs that will be described below. During each primary experimental run, eight oriented grayscale gratings were presented twice at each of two levels of contrast (50 % or 100 % Michelson contrast in six out of seven participants, 20 % or 80 % Michelson contrast in one participant). Grating parameters replicated those of Rademaker et al. (2019). Gratings (spatial frequency of 2 cycles per degree) were masked with annuli (1.2° inner and 7° outer radii). The annuli edges were smoothed with an isotropic 2D Gaussian kernel (1° kernel, 0.5° standard deviation). Throughout each run, a magenta fixation dot was presented in the center of the screen (0.2°, RGB: 0.7843, 0, 0.8886).

A trial consisted of the presentation of a counterphasing (5 Hz) grating for five seconds. In the middle three seconds of each trial, the spatial frequency of the grating either increased or decreased (1 cycle per degree) for 200 ms. Participants were instructed to indicate via a button press the direction of change as soon as they noticed the change. Per run, gratings were presented at multiple orientations, twice at each combination of orientation and contrast. In most runs, there were eight orientations, but in one session of one participant (totaling six runs), only seven orientations were presented. A five-second fixation period preceded the first trial, and a fifteen-second fixation period succeeded the final trial. The total run time was 490 seconds.

2.4.2.2 Population Receptive Field Mapping Scans

For the current application, pRF mapping was used to limit analyses to voxels that were responding to the grating stimulus, but neither the edges of the stimulus nor more peripheral region of the display. pRF mapping scans closely followed the

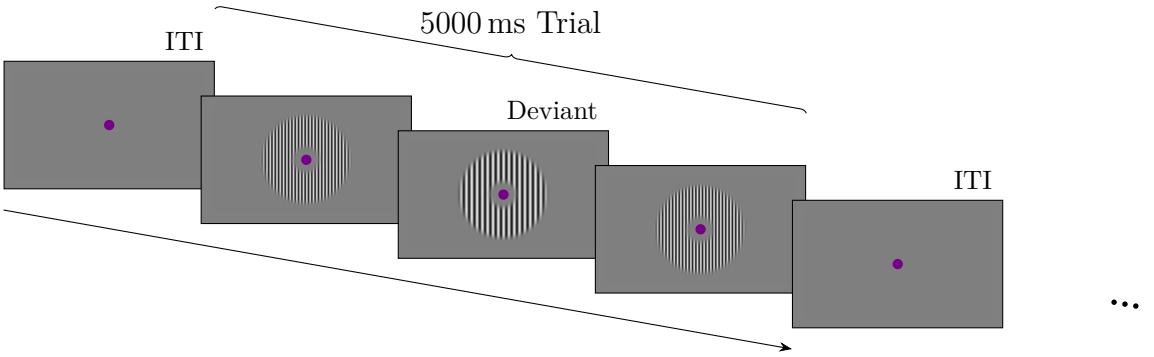


Figure 2.5: Outline of task. Counterphasing (5 Hz), oriented gratings were presented for 5 seconds, with an 8-12 second inter-stimulus interval (ISI). During one, 0.2-second flash of the grating, the spatial frequency of the grating either increased (depicted) or decreased. Participants were instructed to indicate the direction of the change during each trial.

protocol specified by Benson et al. (2018). Briefly, natural images (from Kriegeskorte et al., 2008) were overlaid on pink noise and viewed through a series of apertures. Within one run per session, the apertures enabled view of either a moving bar or rotating wedges and rings that expanded and contracted. In the bar runs, a bar traversed the central region in cycles. During each cycle, the bar was visible for 28 seconds, followed by a 4-second blank period. The bar moved in one of eight directions (east, north, west, south, northeast, northwest, southwest, or southeast, in that order). A 16-second blank period preceded the first cycle, a 12-second blank period followed the fourth cycle, and there was a 16-second blank period at the end of all cycles (300 seconds in total).

In the second pRF scan, the apertures were either wedges that rotated clockwise or counterclockwise, or they were rings that expanded or contracted. These runs started with a 16-second blank period, followed by two, 32-second cycles of a counterclockwise rotating wedge, two 28-seconds of expanding rings (each followed by a 4-second blank period), two 32-second clockwise wedge rotations, and two 28-second cycles of contracting rings (followed by 4 and 26 seconds of blank, respectively). The total run time was again 300 seconds.

Throughout the pRF scans, the color of a central fixation dot (0.2°) changed between black, white, and red. Participants were instructed to monitor the color of the fixation dot and press a button when the dot turned red. To help participants maintain fixation, a circular fixation grid was presented throughout the scan.

2.4.3 fMRI Data Acquisition

MRI data were collected on a 3 T Siemens Skyra scanner with a 64-channel head coil. In each of the three sessions and in addition to field-mapping and functional scans, a T1-weighted anatomical scan was collected (MPRAGE, FOV 256×256, 1 mm isotropic, TE 2.13 ms, Flip Angle 9°). The anatomical scan was used to align field-mapping and functional images parallel to the calcarine sulcus. Gradient recall echo scans were used to estimate the magnetic field. The pRF and the primary functional data were collected with the same scan parameters (TR 1000 ms, TE 31 ms, flip angle 64°, FOV 94×94, 2.2 mm isotropic, 2.2 mm slice gap, Multiband Acceleration Factor 4). To aid alignment of functional and anatomical images, single-band reference images were collected prior to each functional run for all but three participants (TR 8000 ms, TE 65.4 ms, flip angle 90°, FOV 94×94, 2.2 mm isotropic, 2.2 mm slice gap).

Preprocessing of these images was performed with fMRIPrep 1.4.0 (Esteban, Blair, et al., 2018; Esteban, Markiewicz, et al., 2018)[4], which relies on Nipype 1.2.0 (Gorgolewski et al., 2011; Gorgolewski et al., 2018) and Nilearn 0.5.2 (RRID:SCR_001362 Abraham et al., 2014).

[4]The fMRIPrep package uses software from many sources and integrates them into a standardized analysis pipeline. Given the standardization, the package authors provide a boilerplate description of the preprocessing, text which they have released under a CC0 license. The following two sections contain text that is a nearly direct copy of the boilerplate.

2.4.4 Anatomical Data Preprocessing

The T1-weighted (T1w) images were corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al., 2010), distributed with ANTs 2.2.0 (Avants et al., 2008). The T1w images were then skull-stripped with a Nipype implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. A T1w-reference map was computed after registration of the individual T1w images (after INU-correction) using `mri_robust_template` (FreeSurfer 6.0.1, RRID:SCR_001847, Reuter et al., 2010). Brain surfaces were reconstructed using `recon-all` (Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al., 2017). Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` (FSL 5.0.9, RRID:SCR_002823, Zhang et al., 2001).

2.4.5 Functional Data Preprocessing

For each of the functional runs, the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. A deformation field to correct for susceptibility distortions was estimated based on a field map that was co-registered to the BOLD reference, using a custom workflow of fMRIPrep derived from D. Greve's `epidewarp.fsl` script (www.nmr.mgh.harvard.edu/~greve/fbirn/b0/epidewarp.fsl) and further improvements of Human Connectome Project Pipelines (Glasser et al., 2013). Based on the estimated susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which

implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using `mcflirt` (FSL 5.0.9, Jenkinson et al., 2002). The BOLD time-series were resampled to surfaces on the following spaces: fsaverage and fsnative (FreeSufer). The BOLD time-series were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as "preprocessed BOLD". A reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. A set of physiological regressors were extracted to allow for component-based noise correction (Behzadi et al., 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128 s cut-off) for the anatomical CompCor (aCompCor). The time-series entering the CompCor analyses are derived from a mask at the intersection of subcortical regions with the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

A general linear model (GLM) was fit using SPM12 (version 7487, RRID: SCR_007037) to the time-series of each voxel during each orientation run to estimate voxel-wise responses to each grating[5]. Prior to fitting the GLM, each voxel's timeseries was con-

[5]Fitting the GLM can be viewed as a preprocessing step to reduce the dimensionality of the data; the method presented here could be configured to run on the raw timeseries, but working with

verted into a percent signal change, relative to the average signal within a run (across voxels). Design matrices were convolved with the canonical hemodynamic response function, parameterized with the SPM12 defaults (Friston et al., 1998), and additionally contained both six motion (three translation and three rotation) and multiple aCompCor regressors. For each run, the number of components was determined by the broken-stick method (Jackson, 1993).

2.4.6 Population Receptive Field Mapping

To mitigate any effect of stimulus vignetting – whereby the orientation information differs between the edge and central portions of a grating (Carlson, 2014; Roth et al., 2018; c.f., Wardle et al., 2017) – I restricted the analyses to only voxels whose population receptive fields (pRFs) did not overlap with the edges of the stimulus (Appendix F). I estimated the pRFs of each voxel with standard methods. First, the preprocessed functional data for the pRF scans were converted into percent mean signal change, relative to the average signal within a run (across voxels). The compressive spatial summation model was fit to each voxel using analyzePRF_HCP7TRET (osf.io/bw9ec; Benson et al., 2018; Kay et al., 2013). Following Benson et al. (2018), the compressive exponent of this model was set to a fixed value (0.05). The resulting pRF parameters were then combined with an anatomical prior for a Bayesian estimation of the parameters (Benson & Winawer, 2018, neuropythy 0.94). Only the parameters of voxels for which the pRF explained more than 10 % of the variance of the functional run were used as empirical parameters for the Bayesian estimation; the remaining voxels' posterior pRF parameters were determined entirely by the prior.

The resulting pRF parameters determined whether a voxel would be analyzed. The pRF resembles an isotropic, bivariate Gaussian. Each of the three pRF sessions

---

the beta weights of a GLM rather than the raw timeseries drastically reduced the computational requirements.

were analyzed separately, resulting in three sets of pRF parameters per voxel. Within a set of parameters, two indicate the center of the pRF, and the third determines the size of the pRF – the standard deviation of the Gaussian. A voxel was analyzed only if a circle centered on its pRF and whose radius was equal to two standard deviations was entirely contained by the grating stimuli, and only if this threshold was met in all three sessions.

2.4.7 Neural Tuning Function Model

Similar to approaches like forward encoding models (Brouwer & Heeger, 2009; Kay et al., 2008), the new technique assumes that the voxel tuning functions arise from a linear combination of the underlying tuning functions, but the technique goes beyond the forward encoding model in two important ways. First, the technique does not assume a fixed shape for the tuning functions but instead estimates it. Second, the technique relaxes the assumption that only a discrete number of tuning functions contribute to the activity of each voxel by modeling a continuum of tuning functions. That is, I assume that the tuning function for each voxel arises from a continuous distribution over all possible orientation preferences (i.e., infinitely many tuning functions, each contributing to the voxel to a different degree). The technique estimates the density of orientation preferences within each voxel, a probability distribution describing the relative weight with which each section of the continuum contributes to a voxel's activity.

To constrain this distribution of orientation preferences, I assume that it follows a unimodal, roughly bell-shaped curve and justify this assumption though simulation. I simulated the composition of a voxel by considering the partitioning of orientation-tuned neurons in cortex into voxels. To do so, I first used a model of the arrangement of orientation preferences in striate cortex to create an orientation preference map (Afgoustidis, 2015). I then partitioned the map into voxels whose size matched the

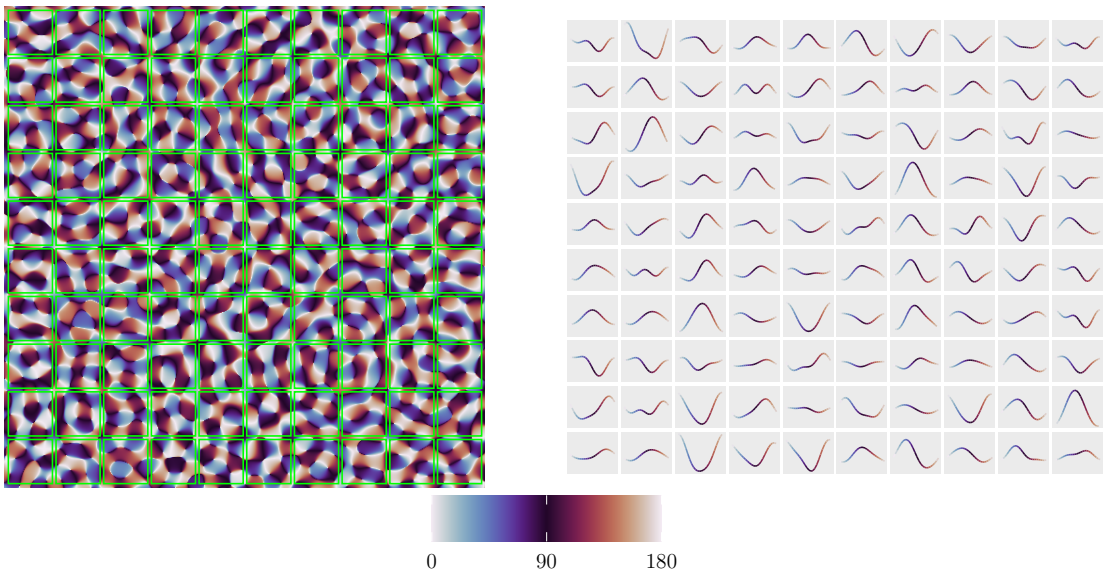


Figure 2.6: Simulation of continuous orientation preferences. **Left:** A simulated patch of cortex using a model of orientation preference maps (Afgoustidis, 2015). The map was parceled into 100 voxels whose size matched the voxels in this study (green lines). **Right:** Weight Distributions. Kernel density smoothing was used to estimate the proportion of neurons tuned to each orientation, within each simulated voxel. Most voxels that exhibit substantial tuning (relatively peaked distributions) have a unimodal orientation preference, with the remaining voxels exhibiting relatively flat tuning. In both plots, each color is assigned to a single orientation.

voxels from the study (Figure 2.6, Left). Within each voxel, I then calculated the proportion of that voxel preferring each orientation and smoothed the resulting distributions using a circular kernel density smoother. Across almost all voxels, the distribution of orientation preferences resembled a unimodal, bell-shaped function (Figure 2.6, Right).

The above simulation envisions voxels' activity as arising from a contiguous patch of cortex. However, voxels likely act like a more complex, spatiotemporal filter of neural activity (Kriegeskorte et al., 2010). That is, the activity of a voxel may reflect the neural activity from many, disparate regions of cortex, and the location of those regions may change through time. The exact function relating neural activity to voxel activity will depend on the underlying vasculature and hemodynamics. However, vas-

culature itself may be functionally organized, such that arteries feed disparate cortical columns tuned to the same orientation (Gardner, 2010). A voxel may therefore be driven by the metabolic activity of disparate but similarly tuned cortical columns. This more complex arrangement may therefore still cause voxels to reflect the activity of a population of neurons whose orientation preference is, at least roughly, unimodal and bell-shaped.

I instantiate the assumed distribution of orientation preferences within each voxel with the von Mises distribution. The von Mises distribution resembles a Gaussian distribution, but it is periodic, thereby accounting for the circular nature of orientations. Within each voxel (but not across voxels), the channels are assumed to have the same neural tuning functions. The assumed neural tuning functions are modified versions of a von Mises density function. These two assumptions, a von Mises orientation preferences distribution and a von Mises tuning function, can each be described by a simple formula. When convolved, they produce a function relating neural responses to voxel tuning functions; I will derive and describe the results of that convolution (Equation 2.3, below), but first I separately describe the orientation preferences distribution (Equation 2.1) and the von Mises tuning function (Equation 2.2).

For an orientation, r, the density function of the von Mises distribution takes the following form

$$f(r|\phi, \kappa_1) = \frac{\exp\left(\kappa_1 \cos\left(r - \phi\right)\right)}{2\pi I_0(\kappa_1)} \tag{2.1}$$

In Equation 2.1, κ_1 is a concentration parameter, ϕ is the orientation at which the function takes the largest value, and $I_0(\cdot)$ is the modified Bessel function of the first kind, of order 0. The numerator of Equation 2.1 creates a periodic[6] Gaussian

[6]The von Mises density function has a period of 2ϕ, and orientations have a period of π; a grating that is oriented to the upper vertical meridian is the same as a grating oriented to the lower vertical meridian. To account for this difference, orientations are rescaled (doubled) during model fitting.

shape peaking at ϕ. The concentration parameter determines the "sharpness" of the function, and the denominator keeps the integral of the function at one for all concentration parameters. In this formulation, Equation 2.1 describes the relative contribution to a voxel's activity of all orientation-tuned neurons within that voxel, and so I refer to it as a "weight distribution". A voxel's weight distribution models the tuning preferences for all neurons within that voxel (e.g., Figure 2.6, Right). A flat weight distribution (low κ_1) translates to the voxel sampling equally from neurons tuned to all orientations. A peaked weight distribution (high κ_1) translates to the voxel sampling mostly from neurons tuned to orientations around ϕ.

The density function of a von Mises distribution can be modified to produce a neural tuning function. Any tuning function must capture at least three features: 1) sharpness of tuning, 2) peak activity, and 3) baseline activity. Sharpness can be captured with the concentration parameter. Peak activity can be captured by scaling the density function with a gain parameter, γ. Baseline activity can be captured with an additive offset, α. The resulting neural tuning function is given by Equation 2.2.

$$NTF(r|\kappa_2, \gamma, \alpha) = \alpha + \gamma f(r|0, \kappa_2) \tag{2.2}$$

In Equation 2.2, the location parameter of the density function is set to zero. This is because the neural tuning function describes the activity of a neuron relative to its preferred orientation. The different orientation preferences are captured with the weight distribution.

The overall voxel tuning function results from convolving Equations 2.1 and 2.2. The convolution of two von Mises distributions is well approximated by another von Mises distribution (Mardia & Jupp, 1999, p. 44). Given this approximation, and by the associativity and distributivity with respect to scalar multiplication of convolutions, the voxel tuning function is therefore given by Equation 2.3.

$$VTF(r|\kappa_v, \gamma_v, \alpha_v, \phi_v) = \alpha_v + \gamma_v f(r|\phi_v, \kappa_v) \tag{2.3}$$

The subscript v has been added to the parameters of Equation 2.3 to emphasize that, in this model, these parameters are estimated separately for each voxel. Note that the concentration parameter of the voxel tuning function, κ_v, is equal to neither κ_1 nor κ_2, but is instead a function of the two (Mardia & Jupp, 1999, p. 44). The model does not distinguish between voxels that reflect the activity of only a narrow range of orientation-tuned neurons (high κ_1) each of which with wide tuning (low κ_1), versus voxels that reflect the activity of a wide range of orientation-tuned neurons each with narrow tuning (low κ_1 and high κ_2). The fact that these concentration parameters are unidentifiable in each voxel can be understood by noting that convolution is commutative[7]. However, estimating each of κ_1 and κ_2 individually was not necessary for estimating the neuromodulations that were the target of this chapter (additive offset and multiplicative gain).

Neuromodulations were instantiated by allowing the parameters of the voxel tuning function (Equation 2.3) to vary with different experimental conditions (Figure 2.4). In the collected data, the orientation tuning functions have been modulated by the contrast of the stimulus. It is known from electrophysiological studies that changes in contrast induce multiplicative gain in neural tuning functions (Alitto & Usrey, 2004; Sclar & Freeman, 1982; Skottun et al., 1987). To allow for multiplicative modulation, I included another voxel-specific gain parameter, g_v, that modulates the neural gain, producing the high-contrast voxel tuning function shown in Equation 2.4a (below). Even when the neural tuning functions undergo a multiplicative gain change, changes to voxel tuning function might be equally well described by an additive offset if the voxel tuning function arises from neurons whose preferences are

[7]In this model, the individual concentration parameters would be identifiable if an experimental condition modulated just one of them. For example, perceptual learning might alter the tuning bandwidth and yet leave the weight distribution intact (Yang & Maunsell, 2004).

uniformly distributed (i.e., voxels with very "flat" tuning functions). The goal of the technique is to distinguish between these two candidate mechanisms for modulation of the underlying neural tuning functions, using only the BOLD data[8]. To ask whether the technique can identify the correct form of neuromodulation, I also fit a second model that included an alternative, plausible (but, according to electrophysiological data, incorrect) kind of modulation: an additive offset (a_v, Equation 2.4b, below). These equations (2.4a and 2.4b) define two models: one in which only a multiplicative gain change was allowed and one in which only an additive offset was allowed. Because higher levels of contrast tend to increase neural activity, the additive modulation parameter is constrained to be positive and the multiplicative modulation parameter is constrained to be greater than one (c.f., Shmuel et al., 2002). In summary, Equation 2.3 captures the situation for low contrast whereas either Equation 2.4a or 2.4b captures the situation for high contrast.

$$\text{Multiplicative} \qquad \alpha_v + g_v \gamma_v f(r|\phi_v, \kappa_v) \tag{2.4a}$$

$$\text{Additive} \qquad a_v + \alpha_v + \gamma_v f(r|\phi_v, \kappa_v) \tag{2.4b}$$

Assuming a linear relationship between neural activity and the statistical beta values that result from applying a standard GLM to the BOLD data, Equations 2.3, 2.4a, and 2.4b specify predicted beta values for each voxel under each tested orientation; i.e., the estimated average response of a voxel to orientations at different levels of contrast. However, there is variability in how voxels respond across runs.

[8]Other forms of neuromodulation could be built into the model. For example, a shift in orientation preference could be modeled by allowing ϕ_v to change across conditions, and a change in concentration could be modeled by allowing κ_v to change (assuming that the condition does not affect the weight distribution). However, both of these forms of neuromodulation could be rejected for the current dataset even without formal model comparison; neither of them allow for contrast to alter the average activity of a voxel across orientations, which was clearly present in the data.

This variability was modeled with a Gaussian distribution and was also allowed to vary across voxels (i.e., an assumption that some voxels may be noisier than others). Each of the voxel-specific parameters (i.e., parameters controlling the tuning function, modulations to those tuning functions, and variability in the output of the tuning function) were estimated with a hierarchical Bayesian model.

I have developed two versions of the technique, which differ in the number of hierarchical layers. Both versions estimate all voxel-specific parameters, but the second version additionally estimates systematic variability across participants. A schematic of the first version, which collapses across participants, is given by Figure 2.7. As seen in Figure 2.7, every voxel was allowed its own value for preferred orientation (ϕ_v), concentration parameter (κ_v), additive constant (a_v), multiplicative constant (g_v), and noise parameter (σ_v). In addition, for the high-contrast presentation (Equations 2.4a and 2.4b), rather than a low-contrast presentation (Equation 2.3), each voxel had its own neuromodulation magnitude parameter, whether that occurred through an additive shift (a_v) or multiplicative gain (g_v). The second, more complex version could leverage, for example, differences in how participants respond to the contrast manipulation: perhaps one participant did not attend to the stimuli and so the influence of contrast was subject to floor effects. In this case, the participant's voxels would have smaller values for the neuromodulation parameters (a_v and g_v), relative to a more vigilant participant. Participant-level parameters in this second version were themselves analyzed hierarchically (i.e., there would be an additional plate in Figure 2.7, covering participants). Both versions of the model led to the same conclusions in the collected dataset.

The Bayesian model was estimated with the R computing language (R Core Team) and its interface (RStan, 2.18.2) to the Stan language (Carpenter et al., 2016; Stan Development Team, 2019). In all fits, four chains were initialized with random starting values for the parameters, and the sampling algorithm was given 1000 draws to

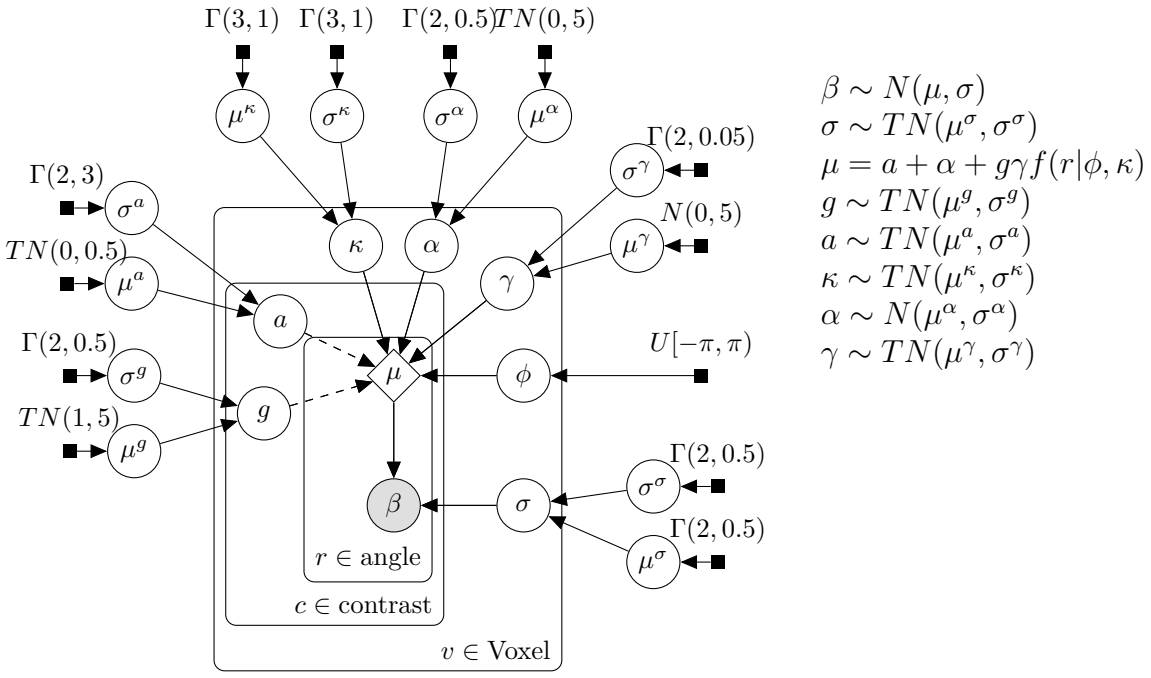


Figure 2.7: Schematic of model, collapsing across participants. Filled square nodes indicate priors, open circles are estimated parameters, the shaded circle is the observed data, and the open diamond is the result of a deterministic function of the parameters (Equations 2.3 and 2.4). Nodes are grouped with the square "plates", indicating over which subsets of the data the node is replicated. The distribution assigned to each node is listed to the right of the diagram. $N(\mu, \sigma)$ is a normal with location μ and scale σ, and $TN(\mu, \sigma)$ is a normal with the same parameters, truncated below at μ. $\Gamma(\zeta, \tau)$ is a gamma distribution with shape ζ and rate τ. Parameters γ, κ, a, and g follow truncated normal distributions, and α follows a normal distribution. Both kinds of modulation (a and g) are depicted in this single diagram, but only one modulation was allowed during model fitting. In another version of the model, each of the μ^x parameters are themselves estimated hierarchically across participants.

adapt. After adaptation, each chain was used to draw 2000 samples from the posterior distribution. The Stan language provides a robust variation of Hamiltonian Monte Carlo (Duane et al., 1987; Hoffman & Gelman, 2014), which efficiently and accurately approximates Bayesian posterior distributions (Betancourt, 2017). Two diagnostics were used to assess the accuracy of the estimation. First, I calculated the split-$\hat{R}$ ("split r-hat") for each parameter in the model (Gelman & Rubin, 1992). This value is analogous to an F-score in an analysis of variance, and values close to 1 imply that

the different Markov chains are sampling from the same distribution. The split-$\hat{R}$ was below 1.1 for all parameters (Gelman et al., 2013, chpater 14). Second, chains were monitored for divergences, whose presence would indicate that the samples, even if they are from a common distribution, likely misrepresent the true posterior distribution (Betancourt, 2017). There were no divergences in the fits of the multiplicative and additive models to real data. All model comparison was done with Pareto-smooth importance sampling, leave-one-out cross-validation plus (PSIS-LOO+, Appendix E; Vehtari et al., 2017).

2.4.8 Model Recovery

The technique relies on model comparison. Each model allows for a single kind of neuromodulation (i.e., multiplicative or additive), and if a model provides a substantially better fit to the empirical data, then this will be taken as evidence that the kind of neuromodulation allowed by the model is the one that actually transpired. To demonstrate that the proposed modeling technique can determine which of the two assayed kinds of neuromodulation describes the data reliably, I conducted two model recovery analyses: data informed and data uninformed (Jang et al., 2012; Wagenmakers et al., 2004). Although such analyses were originally designed for frequentist modeling, they are also applicable to Bayesian modeling.

Data uninformed model recovery attempts to answer whether any of the models are overly flexible given prior assumptions about possible parameter values. Synthetic data were generated using a "best guess" set of priors for how empirical data collected in the real world would be distributed (Figure 2.7). The models were then fit to these synthetic data, keeping in mind the "ground truth" about which model generated the data. A model is considered overly flexible if it always fits the data best, even when the data were generated from a different model. For example, the multiplicative model might be considered too flexible if it always appeared to fit better than the

additive model, even when the additive model was used to generate the data. If that scenario were to happen, an outcome in which the multiplicative model provides the best fit to the empirical data could not be taken at face value; since the multiplicative model fits better when even the data were generated with additive neuromodulation, a better fit to real data would not indicate that the multiplicative model is more likely to be the correct model. Hence, data uninformed model recovery assesses whether model comparison is likely to be fair on datasets consistent with the assumed priors.

For data-uninformed model recovery, 100 datasets were simulated from the priors of both models[9]. The datasets contained one participant contributing 100 voxels measured in 18 runs at 8 levels of orientation. To each of these simulated datasets, both the additive and multiplicative models were fit. This procedure resulted in 200 simulated datasets and 400 model fits.

The second kind of model recovery, data informed model recovery, repeats the model recovery simulations with the posterior distributions of each model. While data-uninformed model recovery assesses the fairness of model comparison across a wide range of datasets that are plausible given the prior distribution, model informed recovery assesses fairness on datasets that are plausible while taking into account what has been learned about the parameters of a model by observing data. Although the two models may make clearly differentiable predictions under the prior distributions (resulting in simulated datasets that easily identify the generating models), the posterior distributions of the models may generate essentially identical datasets if the data reveal that the effect of contrast tends to be minuscule in all voxels[10].

---

[9]Note that all model recovery simulations were done with the more complex version of the technique, in which participant-level effects are estimated hierarchically. Additionally, in the recovery simulations only, the voxel-specific noise parameter was fixed to a single value across participants. Fixing the voxel-specific noise parameter did not alter the main conclusions of the model comparison when it was applied to real data; the multiplicative model was still preferred.

[10]Formally, data informed model recovery approximates a quantity that is very similar to the quantity approximated by PSIS-LOO+ (see also Wagenmakers et al., 2004). PSIS-LOO+ approximates the expected log pointwise predictive density (Appendix E). This data informed model re-

Data informed model recovery was implemented analogously to the data-uninformed model recovery. One hundred datasets were generated from the posterior of the multiplicative and additive models. The datasets contained six participants, each contributing 200 voxels measured in 18 runs at 8 levels of orientation. The multiplicative and additive models where then fit to these simulated datasets and model comparison was conducted as described above.

2.4.9 Rank-ordered Visualization Technique

The formal decision between different forms of neuromodulation is made with model comparison. However, it is useful to visualize the specific features of the data that drive model comparison, features that are only captured by one but not the other model. For this, a new visualization can be used. The visualization technique relies on how certain forms of neuromodulation may be visible in the voxel tuning functions, given certain assumptions. Ideally, one could visualize those neuromodulations by simply constructing the empirical voxel tuning function of each voxel at the two levels of contrast. However, individual voxel responses are noisy, and so the betas must be collapsed across voxels, allowing for the possibility that voxels are differently affected by contrast.

A key difference between the additive and multiplicative models is that the multiplicative model allows the effect of contrast to vary by orientation whereas the additive model does not. That is, for the multiplicative neuromodulation contrast has a greater magnifying effect for the peak than the tails of the voxel tuning function, whereas for additive neuromodulation contrast has the same effect on the peak and the tails. This distinction can be depicted graphically (Figure 2.8). If the additive model were true, then the beta values for a single voxel, at high contrast should

---

covery approximates a value that adds an outer expectation, e.g., what could be called something like the expected log pointwise predictive density that would be expected across replications of the experiment.

be greater than their low-contrast counterparts to the same extent at all orientations (Figure 2.8, top left). If the multiplicative model were true, then the betas that are highest at low-contrast – those betas produced by orientations that are closest to the voxel's preferred orientation – will be more affected by increasing contrast than betas for other orientations (Figure 2.8 bottom left). This implies that the models can be differentiated by plotting the high-contrast betas against the low-contrast betas and calculating the slope of the best fitting line between the resulting points (Figure 2.8, right). A slope of one implies an additive model, but a slope greater than one implies a multiplicative model.

To aggregate across voxels, beta values were sorted and grouped by a rank-ordering of their magnitudes across orientations. That is, the average beta value for each voxel was first calculated at each orientation (i.e., averaged across runs and levels of contrast). Then, the resulting average values at each orientation were ranked according to magnitude within each voxel (e.g., ranked from 1-8, for each of the 8 presented orientations). Importantly, this scheme ranks the orientations without regard to contrast, forcing the ranking of orientations within each voxel to be the same at high and low-contrast. Next, the original beta values, kept separate by contrast but grouped together to determine rank, were averaged over voxels and runs. For example, if the stimulus orientation of 0° was rank-1 in one voxel, the orientation of 45° was rank-1 in another voxel, and the orientation of 90° was rank-1 in a third voxel (etc.) the beta values at all of these rank-1 orientations were averaged, disregarding orientation but keeping the average values separate for the high and low-contrast conditions. This produced a beta value for rank 1 at high contrast and a yoked beta value for rank 1 at low contrast, without regard to the value of the presented stimulus orientation.

Note that many voxels were only weakly responsive to stimulation or were only weakly affected by contrast. Voxels that are unaffected by contrast will necessarily

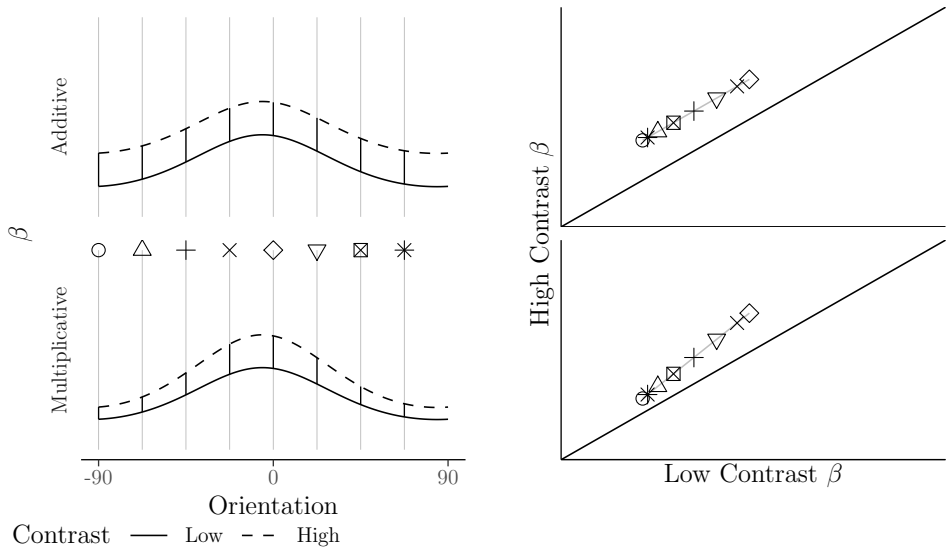


Figure 2.8: Plotting beta weights at high versus low-contrast uncovers neuromodulation. **Left:** Simulated voxel tuning functions in which higher levels of contrast induce either an additive (top) or multiplicative (bottom) neuromodulation. The eight vertical lines are eight hypothetical orientations at which these voxel tuning functions might be probed, which would produce eight beta values per level of contrast, High vs. Low. **Right:** The two kinds of neuromodulation reveal different signatures when the high contrast beta values are plotted against the low-contrast betas. The diagonal line corresponds to no effect of contrast. A line drawn through the points produced by the additive model necessarily has a slope equal to 1; under this neuromodulation the effect of contrast does not depend on the orientation. A line drawn through the points produced by the multiplicative model necessarily has a slope greater than 1; under this neuromodulation the effect of contrast is largest at those orientations which are closest to the voxel's preferred orientation.

have a slope near one. Including these voxels in the visualization would produce a plot that necessarily appears to favor the additive model, even if multiplicative gain was the true form of neuromodulation. Therefore, this analysis was repeated four times using an increasingly stringent criterion to exclude voxels that were weakly affected by contrast. The criterion was based on the average effect of contrast for each voxel, without regard to orientation (i.e., average across runs and orientations). Voxels were included only if they surpassed a specified quantile. The four analyses employed an exclusion criterion at the quantiles of: 0 (i.e., include all voxels), 0.5, 0.75, and 0.9.

2.5 Results

2.5.1 Behavior

Participants were able to perform the 2AFC spatial frequency change detection task. Average accuracy (across types of spatial frequency change) was 72 % and 74 % for low and high-contrast gratings, respectively ($p = 0.13$).

Eyetracking data were collected as a record that participants maintained adequate fixation during each of the runs. Across runs, over 90 % of all participants' fixations ended within 2° of the location of the run's average fixation (range: 93–99 %). No further analyses were performed on these data.

2.5.2 Voxel Counts

To mitigate stimulus vignetting, analyses were restricted to voxels with population receptive fields that overlapped with the stimulus (Carlson, 2014; Roth et al., 2018; c.f. Wardle et al., 2017). This provided 1,010 voxels from V1 in total for analysis.

2.5.3 Data Uninformed Model Recovery

To evaluate the proposed technique's ability to recover both kinds of modulation, I conducted data-uninformed model recovery by simulating datasets from the prior distribution, fitting two models to these datasets that allowed for different kinds of modulation, and comparing the resulting fits of the model using cross validation (Vehtari et al., 2017; Wagenmakers et al., 2004). This revealed that the models did not exhibit strong mimicry; model comparison picked the true data-generating model in all 200 simulated datasets. These simulations provide strong evidence that the technique can recover the appropriate class of modulation when it is present in the data.

2.5.4 Application to Observed Data

Given that the technique recovers the correct type of modulation in simulated datasets, I next asked whether the technique can recover from fMRI data the correct kind of neuromodulation undergone by neural tuning functions. Specifically, it is known from electrophysiology that the tuning functions of orientation-sensitive neurons undergo multiplicative neuromodulation (Alitto & Usrey, 2004; Sclar & Freeman, 1982; Skottun et al., 1987). Given that the neural tuning functions undergo multiplicative gain, model comparison on the collected data ought to prefer the multiplicative model.

I fit two versions of the model to the real data. Both versions of the model estimated the voxel-specific parameters hierarchically. But in one version, an additional, participant level hierarchy was included in the model. In both models, the multiplicative model provided the best fit to the data (a difference of 13.2 standard errors for the simpler version, and 14.5 for the more complex version).

2.5.5 Data Informed Model Recovery

To probe the validity of model comparison for these particular data, as opposed to data generated from the prior distribution, data-informed model recovery was performed (Wagenmakers et al., 2004). Datasets were generated from the posterior distribution of the multiplicative and additive models. Next, the multiplicative and additive models were refit to each of these simulated datasets. As with the uninformed model recovery simulations, all 200 datasets were fit best by the model whose posterior was used to generate the data.

2.5.6 Rank-ordered Visualization of Data

The model recovery results suggest that the method is capable of selecting the appropriate kind of neuromodulation, and the applying it to real data suggested that contrast induces a multiplicative gain, in accord with electrophysiology. But what

features of the data allow the multiplicative model to outperform the additive model? One diagnostic feature is the slope of the high contrast beta values plotted against the low contrast betas (Figure 2.8). Additive neuromodulation necessarily produces points that lie along a line with a slope of one, while multiplicative gain can produce a line whose slope is greater than one. Beta values simulated from the two models reflect this (Figure 2.9A). Critically, in the observed empirical data, this slope is also greater than one. This is most strongly the case when analyzing only those voxels that were most affected by the contrast manipulation (Figure 2.9B, left-most versus right-most panel), as would be expected under the multiplicative model. It is this feature of the data that favors the multiplicative model over the additive model.

This visualization highlights that additive neuromodulation can mimic multiplicative gain when considering the results across voxels. That is, without considering how the results for one stimulus orientation relate to the results for a different stimulus orientation for the each voxel, both models produce the same general pattern of results. In Figure 2.9A, the lines connect different stimulus orientations separately for each voxel to provide a within-voxel analysis. Within each voxel, the slope of the betas predicted by the additive model is necessarily one (Figure 2.9A, left). However, the additive model is hierarchical, and so allows that voxels may be differently affected by contrast and so shifted away from the diagonal by different amounts. Those voxels that were the most responsive at low contrast appeared to be the most responsive at high contrast, meaning that the slope of a line fit to all predicted betas, collapsed across voxels, would be larger than one.

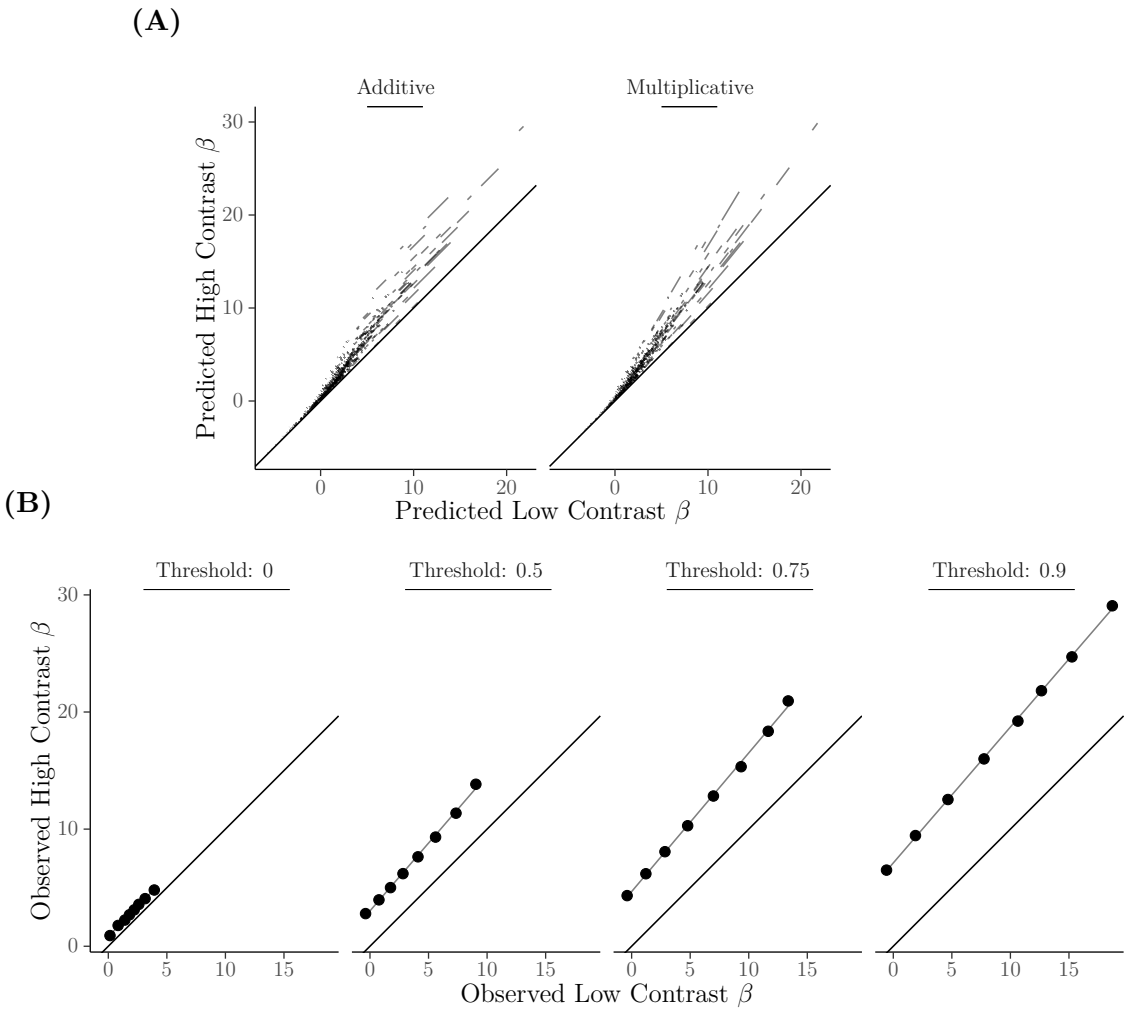


Figure 2.9: Multiplicative model wins because it alone can capture the interaction between orientation and contrast within voxels. Compare to Figure 2.8. **(A)** Posterior distribution of the predicted betas from the two models, plotted as separate lines for each voxel that connect the average predicted betas across each of the stimulus orientations. Most voxels have low betas at each level of contrast and are only weakly influenced by the experimental manipulation (i.e., lines are near the diagonal for small betas). Across voxels (i.e., across lines), both models can capture how contrast has a larger effect on those voxels with higher betas; betas lying further rightward along the x-axis are located further away from the diagonal. However, only the multiplicative model can capture the interaction between orientation and contrast within voxels. That is, some voxels have slope larger than one in the multiplicative but not in the additive model). **(B)** Empirical betas, plotted as in (A), but aggregating across voxels according to rank ordered orientation preferences (see main text). Panels include different numbers of voxels, filtered based on the difference in their average response at low and high contrast (e.g., 0 includes all voxels, 0.9 includes only the top 10 % of voxels). With filtering, it is apparent that in the observed betas, increasing contrast increased the response of voxels most strongly at their preferred orientations (i.e., the slope is larger than one).

Even though the additive model captures the same general pattern of results across voxels, it produces a different set of predictions within each voxel. As seen in Figure 2.9A, the posterior predictive distribution for the multiplicative model shows slopes greater than one separately for each voxel (i.e., the individual line segments have slope greater than one), and the voxels with larger beta values tended to be voxels that had slopes that were more clearly greater than one. This is precisely the same pattern seen in the real data as seen in Figure 2.9B. When all voxels are included in the rank-ordered visualization technique (threshold 0), the results show a slope near one, which is to be expected because the majority of voxels produced very low beta values and, correspondingly, relatively less sensitivity to the contrast manipulation. However, when only considering the top 10 % of voxels in terms of the contrast manipulation (threshold .9), the beta values are in general larger, and the slope is clearly greater than one, as predicted by the multiplicative (but not additive) model.

2.6 Discussion

Many cognitive and perceptual phenomena, including serial dependence, have been accounted for by theories that make predictions about the nature of neural tuning and changes to that tuning. Yet these predictions have not been tested in humans because there are no non-invasive techniques for recording from individual neurons. Instead, researchers rely on methods like fMRI to infer these properties based on voxels that capture the hemodynamic response arising from hundreds of thousands of neurons. Voxel tuning functions have been interpreted as revealing population-level properties of stimulus representations in the brain, but it is widely acknowledged that they cannot be taken to reflect the response properties of the neurons that underlie the BOLD signal (Çukur et al., 2013; Nevado et al., 2004; Serences et al., 2009; Sprague et al., 2018). To address this limitation, I developed a procedure that models the

relationship between neural-level responses and the BOLD signal, in order to allow inferences about modulations in neural tuning functions directly from fMRI data. The procedure employs a novel, hierarchical Bayesian model of BOLD data, combined with model comparison. Applying the procedure to a simple test case, in which the neural-level "ground truth" modulation in tuning is known from electrophysiology (Alitto & Usrey, 2004; Sclar & Freeman, 1982; Skottun et al., 1987) recovered the correct form of modulation in neural-level tuning curves. In addition to formal assessment with model comparison, I also developed a technique for qualitatively visualizing the form of voxel tuning modulation in empirical data based on a rank ordering across stimuli separately for each voxel. This produced within-voxel slopes greater than one when plotting one condition (low-contrast) against the other condition (high-contrast), which is the unique signature of multiplicative gain.

A simple kind of neuromodulation – modulation of orientation tuning by stimulus contrast – was used to test the method, but the same method is applicable to a broad range of domains and manipulations. Now that the technique has been verified with a known form of modulation and with model recovery studies, it can be used to study modulations of neural tuning functions where electrophysiology has not yet already provided an answer and for complex tasks that cannot be taught to animals. The main requirement for using the technique is a voxel tuning function created by testing the BOLD response at different levels of a single stimulus dimension (e.g., direction of motion, isoluminant hue, pitch, etc.), that is affected by cognitive manipulations of interest (e.g., with and without attention, before and after perceptual learning, during encoding versus during retrieval). For instance, this technique may help solve currently unanswered questions about the neural underpinnings of serial dependencies. There are currently many possible models of serial dependence demonstrating how it could be elicited by different forms of neuromodulation (Fischer & Whitney, 2014; Papadimitriou et al., 2017; Pascucci et al., 2019), but it has proven difficult to adju-

dicate between these forms of neuromodulation with only behavioral data (Fischer & Whitney, 2014).

Previously developed voxel tuning methods are often based on a forward encoding model – a generative model that postulates the form of stimulus representations employed by the brain in order to predict BOLD data for a given stimulus input (Brouwer & Heeger, 2009, 2011; Scolari et al., 2012; Sprague & Serences, 2013). The structure of these encoding models is similar to the technique presented here, characterizing the voxel-level tuning function as arising from tuning functions belonging to some population of neurons that exists at a spatial scale below that of voxels. Although it is usually assumed that these lower level tuning functions are not directly related to neural tuning functions, referred to instead as "channel" tuning functions, these channel functions play the same role in the proposed method as the tuning functions that I have labeled as "neural" tuning functions. Moreover, one frequent application of such encoding models – sometimes called the Inverse Encoding Model (IEM) – claims to uncover modulations to those channel tuning functions (Brouwer & Heeger, 2011; Sprague et al., 2015).

Some authors have explicitly claimed that the inverted encoding model provides information about only a population-level activity and not the neural-level (Sprague et al., 2018), whereas I have claimed that methods presented in this chapter can uncover neural-level modulations. Applications of the IEM typically assume a discrete number of channels and assume specific parameter values for the channel function, with these simplifying assumptions enabling the use of multivariate regression to identify channel responses. In my approach, I have instead attempted to justify the neural-level label by allowing the data to determine the shape of neural tuning function and by assuming a continuous distribution across neurons preferring different stimulus values. These differences alone partly explain why the new technique affords inferences that more accurately track neuromodulation. However, given that the channel tuning functions

of the IEM play a similar role as the tuning functions that I have labeled as neural in the present approach, and given that the IEM can uncover modulations to channel tuning functions, it may appear that the method presented in this chapter does not offer a qualitatively new kind of inference. Hence, I will clarify three key differences between the techniques that account for why the new method offers inferences that are qualitatively different than those of the inverted encoding model.

The first and simplest difference is that the presented technique models variability in tuning functions across voxels, whereas inverted encoding models typically assume that all voxels have exactly the same channel response function. Accounting for the variability proved critical for capturing the manner in which different voxels were differently affected by contrast. Some voxels in the dataset were only weakly responsive to visual stimulation and, among those that were responsive, even fewer were strongly affected by contrast. Although I restricted the analyses to only those voxels whose population receptive fields overlapped with the stimulus (which of course tends to select for voxels that were more responsive to stimulation), the population receptive field mapping was likely imperfect, resulting in the inclusion of some voxels whose receptive fields may have included regions beyond the stimulus. Additionally, inadequacies in how the beta values were estimated likely vary across voxels (e.g., the assumed hemodynamic response function might have matched the true function of some voxels better than others, Handwerker et al., 2004). These and other sources of variability (variability in the success of alignment of voxels across sessions, partial voluming, the responsiveness of neurons within each voxel, proximity of each voxel to blood vessels, etc) will be present in any fMRI experiment. In general, a failure to consider these sources of variability may incorrectly favor the additive gain model, considering that additive gain is a simple shift of the mean values without any concomitant changes to the shape of voxel tuning functions (see for instance the difference between including all voxels, as shown in the left panel of Figure 2.9B, versus only

considering the higher signal voxels, as shown in the right panel of Figure 2.9B). This variability was handled in the present model by the hierarchical nature of the technique, in which the large number of voxel parameters are constrained by assuming that the parameters are drawn from global-level distributions. In this manner, the technique uses all of the data to measure the relatively subtle distinctions between the different forms of modulation (i.e., the model comparison is properly constrained by those elements of the data that are most diagnostic).

A second and more important difference between my proposed method and the IEM relates to a currently contentious question about whether the technique even allows for valid inferences about modulations to the so-called channel tuning functions (Gardner & Liu, 2019; Liu et al., 2018; Sprague et al., 2018; Sprague et al., 2019). Standard practice is to fit the inverted encoding model in a two-step procedure. The first step assumes a set of channel tuning functions and uses a "training" set of BOLD data to estimate the weights of the hypothetical channels for each voxel, via least-squares linear regression. Then, in a second step, the IEM method takes an independent "test" dataset and inverts the estimated weight matrix, to produce an estimate of the channels' responses (across all voxels) in the test data. A key advantage of this approach is that it can be used to estimate modulations in the channel responses by some experimental manipulation (e.g., Brouwer & Heeger, 2011; Ho et al., 2012; Saproo & Serences, 2014; Scolari et al., 2012; Sprague & Serences, 2013). That is, whenever a test dataset contains two conditions (e.g., produced by some manipulation), if we infer the IEM weights from a common, independent training dataset, we can compare the channel outputs produced separately by data from the two test conditions to examine how the brain's stimulus representations are modulated across the conditions.

Critics of the IEM have argued that the estimated channel responses produced in the second step are determined entirely by the channel response functions that are

assumed in the first (Gardner & Liu, 2019). Proponents of the IEM have countered that although the form of the IEM output is determined by its assumptions, this is true of most model-based analyses, and as long as the assumptions are sensible, inspired by neural evidence, and explicitly held in mind, the results can be meaningfully interpreted (Sprague et al., 2019). Importantly, Sprague and colleagues further claim that modulations in tuning revealed by the IEM (i.e., any difference in tuning across conditions) are invariant to certain assumptions. In particular, they show that inferred modulations may be invariant to a linear transformation of the tuning functions. That is, they present a simulation in which two IEMs, whose channel functions are related by a linear transformation, are fit to a common dataset. Despite the different tuning functions, the two models uncovered the same kind of modulation.

This debate about the validity of the inverted encoding model highlights the second key advantage of the technique presented in this chapter: the user does not need to base their conclusions on a fixed set of assumed tuning functions. The simulation presented by Sprague and colleagues may apply generally; it may be that all modulations recoverable by the IEM are invariant to any linear transformation. But even that conclusion would not be a helpful guarantee. Non-linear transformations arise from seemingly innocuous choices in the application of an IEM. For example, using the inverted model requires choosing a set number of channels. This choice is often not based on theoretical considerations but instead on both the tested number of stimulus values and the assumed tuning functions. In models of orientation tuning, the chosen number is often small, e.g., six (Brouwer & Heeger, 2011), ten (Scolari et al., 2012), nine (Rademaker et al., 2019). Common choices in tuning function shape often also result in non-linear transformations (e.g., sine functions raise to different powers). Depending on the chosen number and shape of the tuning functions, the data generally provide enough constraint to fit IEMs with a range of channel numbers. But the

results from applying an IEM with six channels could be qualitatively different than the results from applying an IEM with ten channels to the same dataset.

For the IEM, there does not appear to be a general way to determine, *a priori*, the shapes and number of channels; instead, researchers must check *post-hoc* whether the IEM produces the same results across some sufficiently wide range of values for the shape and number of channels. But, what range is sufficient? How could uncertainty in the results be quantified if the conclusions held for fifteen but not sixteen channels? Given the already problematically high degrees of freedom allowed to fMRI researchers (Carp, 2012), it is critical to be able to show which conclusions depend on which structural assumptions (Draper, 1995). This, currently informal, practice of checking and re-justifying tuning functions is formalized and made transparent in the proposed approach. In the proposed technique, neural subpopulations within the voxel are modeled with a continuum of orientation preferences, and the shape of the tuning functions is determined separately for each voxel by parameters whose values are estimated with the data[11].

The third key difference between the IEM and the presented technique centers on how these two approaches allow researchers to make a decision about modulation. The output of the IEM is a set of estimated channel functions across conditions, but researchers must decide with each experiment how to determine whether any observed difference in the estimated channel functions are meaningful. In practice, this means using the inferred channel responses as data with which to fit *another* model. These extra models range from mixed-effects linear models (i.e., fitting a generalized, linear model to the channels' infered responses at each of the tested orientations; Scolari et al., 2012) to a descriptive model of the functional form the channel functions (e.g., fitting a von Mises function to the channel responses; Liu

[11]This is not to say that the current technique is assumption-free: assumptions are built into the current technique which will influence its validity (see, e.g., Limitation and Extensions, below).

et al., 2018) to domain-specific models (e.g., the normalization model; Brouwer & Heeger, 2011). Null-hypothesis significance testing or model comparison is based on that final model. But each layer of modeling discards information about how well the earlier models fit their data or how well the model parameters were constrained, and this lost information will be inaccessible to the final statistical test. To be clear, there is nothing inherently wrong with fitting a model to estimated parameters (both the presented technique and the IEM are not fit to the raw BOLD signal but instead to the estimated response of each voxel, as given by an initial general linear model). However, particularly given the sensitivity of the estimated channel responses to relatively arbitrary modeling decisions (discussed above), that the IEM requires a supplementary model makes it even more challenging to evaluate results. In contrast, the presented technique attempts to explicitly identify the form of neuromodulation through a single step of model fitting and comparison. Given the relatively close link between model comparison and the data afford by the new technique, using the new technique facilitates transparency in how features of the data drive the quantitative statistical conclusions.

To summarize: the new method offers three key advantages: 1) it explicitly accounts for variability across voxels through hierarchical modeling, 2) it seeks to eliminate the use of arbitrary assumptions about the neural subpopulation, instead using a Bayesian model to determine the most likely values of unknown model parameters, and, perhaps most importantly 3) it explicitly attempts to identify the correct form of modulation in at the neural level, by assaying different models of modulation, then employing model comparison and model recovery to determine the most likely model and verify the reliability of that conclusion.

2.6.1 Limitations and Extensions

As implemented, the method assumes that the distribution of neurons tuned to the feature under examination (i.e., orientation), within each voxel, is unimodal. The unimodal assumption provided a convenient formula for a voxel tuning function given the parameters of a neural tuning function, and simulated maps of orientation preference justified this assumption for the collected dataset (Figure 2.6). However, these simulations may not be relevant for all stimulus features, nor all datasets. For example, the distributions of orientation preferences may be multimodal for smaller voxels (e.g., as determined by scanner parameters) or, conversely, as applied to cortical regions that are more spatially heterogeneous in their selectivity (e.g., greater dissimilarity from one millimeter to an adjacent millimeter).

Although I have shown that the technique can uncover certain kinds of modulations to neural tuning functions (i.e., multiplicative gain versus additive offset), the technique may not be able to uncover every form of neuromodulation. The technique still requires that the forms of neuromodulation in question affect the voxel tuning function distinctly. That distinctness may not always hold, particularly when an experimental manipulation induces multiple kinds of neuromodulation. For example, although it is known that contrast selectively alters the gain of orientation tuning functions, modulations following adaptation are substantially more complex (Dragoi et al., 2000; Patterson et al., 2013; Wissig & Kohn, 2012). In one case, following long adaptation to oriented stimuli, orientation tuning functions can both widen and shift their preferred orientation (Dragoi et al., 2000). In principle, the proposed technique could handle these additional modulations by allowing more parameters to vary between each model, since in the model both the width and preferred direction map on to individual parameters. But in applying this technique to explore neuromodulations

beyond a multiplicative gain and additive offset, further model recovery simulations are advised[12].

Although the new technique provides a way to determine the qualitative type of neuromodulation undergone by tuning functions, the technique does not estimate the quantitative magnitude of the neuromodulation. That is, it would be a mistake to infer, e.g., the magnitude of the difference in neural firing rate at high versus low-contrast from the posterior distribution of the multiplicative gain parameter. This limitation is caused by simplifying assumptions in the technique. One such simplification involved modeling the voxel tuning function as arising from a linear combination of the activity of the underlying neural tuning functions; in truth, the relationship between neural firing rate and BOLD signal is non-linear (Bao et al., 2015). Another simplification is that the technique supposes that all neurons within each voxel have tuning functions that share a single shape (though it allows these functions to be different for different voxels). Considering the relatively wide tuning of inhibitory interneurons (Webb, 2005), this assumption is also almost certainly incorrect.

These simplifying assumptions are not strictly necessary, and it will be beneficial to research how to avoid them and thereby also achieve quantitative estimation of neuromodulations. Similar research into neuronal and voxel tuning has demonstrated that quantitative inferences about neural tuning functions are feasible with only voxel-level data (Keliris et al., 2019). Like the technique presented here, their method relies on explicitly modeling how the activity of a continuous population of neurons combines to give rise to the BOLD signal. Some of their success may be due

[12]This warning also suggests that there are circumstances where certain forms of modulation can be ruled out even without estimating a model. For example, given that the BOLD signal in the data was generally higher following high as compared to low contrast stimuli (i.e., the data in Figure 2.9B are mostly above the diagonal), a model in which contrast altered only the sharpness of the neural tuning functions (κ_v in Equation 2.3) would be unable to account for the observed data.

to the particular tuning that was their focus: receptive field size. Receptive field sizes vary relatively slowly across cortex (e.g., the receptive field sizes of neurons within one voxel is predictive of the receptive field sizes in neighboring voxels, see also Appendix F), whereas features like orientation vary on a submillimeter scale (e.g., 45° might be represented by multiple, noncontiguous subpopulations of neurons within each voxel, Figure 2.6). It is encouraging that voxel-level data can support quantitatively accurate estimates of neural tuning, but further research will be required to determine how well that accuracy generalizes to tuning for other features.

CHAPTER 3

GENERAL SUMMARY

In this dissertation I developed new techniques for identifying the latent factors that contribute to overt behavioral and neural responses and I applied these techniques to the study of the processes underlying visual orientation perception.

In the first chapter, I focused on behavioral changes caused by repeatedly viewing and judging orientations. Classic work on the tilt aftereffect (Gibson & Radner, 1937) demonstrates a repulsive effect of passively viewing an orientation. However, recent work on serial dependence suggests an effect that is attractive rather than repulsive (Fischer & Whitney, 2014). I performed a meta-analysis of the serial dependence literature in judgments of orientation, finding substantial variability in the magnitude of the serial dependence effect, with no single experimental factor explaining this variability. Making sense of this situation, I found that every trial reflects a combination of both repulsion from the previously viewed stimulus and attraction to the previously given response, with these factors canceling each other out to differing degrees across different trials and different experiments. When the response on the previous trial is accurate, previous response and previous orientation take on the same value, and thus the separate effects of prior response versus prior stimulus are entirely confounded. However, on the trial following one in which a participant is less accurate, it is possible to untangle these separate effects. I developed two techniques that use naturally occurring previous-trial errors to separate these effects: a simple split-half technique and hierarchical Bayesian model of the trial-by-trial data. My analyses of a new dataset and several existing datasets indicate that the attrac-

tive factor underlying serial dependence is not a novel perceptual effect, but rather a form of response hysteresis. Across studies of serial dependence, I claimed that this previously-documented, attractive factor (response hysteresis) has combined with a previously-documented, repulsive factor (the tilt aftereffect), producing a wide range of outcomes from overall repulsion to overall attraction.

Similar to the multiple factors that can underlie behavior, aggregate neural behavior, such as measured by an fMRI voxel, will reflect the activity of many neural subpopulations. Therefore, fMRI data cannot be used to directly measure neural tuning functions, such as the response of a simple cell in primary visual cortex across the full range of stimulus orientations. In the second chapter I again used a hierarchical Bayesian model and model comparison to address this inverse problem, but applied to neural rather than behavioral data. Although it does not provide quantitative measurements of the underlying neural tuning functions, my technique identifies the type of modulation that occurs with experimental manipulations (e.g., with versus without attention). I validated this technique with a new fMRI experiment that manipulated the visual contrast of orientated gratings. For a manipulation of visual contrast, prior electrophysiological work indicates that neural tuning functions undergo multiplicative gain, and my technique recovered this "ground truth" from the fMRI data, ruling out an alternative explanation in terms of additive gain, despite the fact that the average fMRI data appeared additive. Now that the technique has been validated, it can be applied to situations (e.g., serial dependence) where the form of neural modulation is currently unknown.

These two chapters highlight two closely related issues in the study of brain and behavior: 1) most measurements, such as selection of an orientation or the response of an fMRI voxel, reflect a combination of multiple latent factors; and 2) analyses that consider only the average measurement will be insufficient for identifying these latent factors. Cognitive science is a discipline dedicated to latent processes – the

unobservable steps that translate a stimulus into a response – and so there is a critical need to identify these latent factors. In my work, rigorous statistical modeling of the full dataset (trial-by-trial and voxel-by-voxel) was used to identify and differentiate the latent factors that gave rise to average responses. Such latent forces cannot simply be marginalized away, but explicit modeling can isolate them. In both chapters, application of hierarchical Bayesian modeling to the full dataset made sense of the variance and covariance contained within the data, and the results from these analyses were more readily linked to the literature and previously established findings. In Chapter 1, I showed that the serial dependence effect, which was previously attributed to a novel perceptual process, was instead a combination of two well-established effects: the tilt aftereffect and response hysteresis. In Chapter 2, I showed how the population-level activity of fMRI data, which showed an on-average additive effect of visual contrast, was instead the result of multiplicative gain, such as previously identified by single cell electrophysiology.

APPENDIX A

THE EVIDENCE BETWEEN PERCEPTUAL AND POST-PERCEPTUAL ACCOUNTS OF SERIAL DEPENDENCE REMAINS EQUIVOCAL

It remains unclear what causes serial dependence. As discussed in the introduction of Chapter 1, the inducer initiates many psychological processes, and each of these may give rise to dependencies. One way to narrow the range of possible causes would be to determine whether there are attractive biases only after participants make a decision about an inducing stimulus (Bae & Luck, 2020; Cicchini et al., 2017; Fischer & Whitney, 2014; Fornaciai & Park, 2018a; Fritsche et al., 2017; Manassi et al., 2018; Pascucci et al., 2019; Suárez-Pinilla et al., 2018). Researchers have manipulated whether participants make decisions, but conditions designed to prevent decisions have resulted in both attractive and repulsive dependencies on the inducing stimulus. These conflicting results may be due to differences in how decisions have been manipulated (Bae & Luck, 2020; Pascucci et al., 2019). One manipulation involves equating a decision with a response (i.e., supposing that participants make decisions if and only if they make a motoric response) and then asking participants to only respond on a subset of trials; the empirical question becomes whether responses are attracted toward stimuli to which no response was made (Fischer & Whitney, 2014; Fornaciai & Park, 2018a, 2018b; Fritsche & de Lange, 2019b; Manassi et al., 2018; Suárez-Pinilla et al., 2018). But although the manipulation is straightforward, post-perceptual processing can occur despite the lack of a response – e.g., participants can make decisions implicitly (Fornaciai & Park, 2018a; Pascucci et al., 2019).

Pascucci et al. (2019) attempted to rule out implicit decisions by instructing participants to actively ignore some inducing stimuli. They observed that dependencies on the inducing stimuli, when unattended, were repulsive (i.e., the dependency that remained was more in line with a tilt aftereffect). However, two complications prevent a straightforward interpretation of these results. First, instructions to ignore the inducing stimuli have also resulted in attractive serial dependence (Fornaciai & Park, 2018a). Second, the strength of the serial dependence effect is modulated by attention (Fischer & Whitney, 2014; Fornaciai & Park, 2018b; Fritsche & de Lange, 2019b), suggesting that by instructing participants to ignore (i.e., not attend) the inducing stimulus Pascucci and colleagues may have eliminated the attractive dependence. Therefore, their results do not conclusively demonstrate that attractive dependencies require participants to make a decision about the inducer.

Bae and Luck (2020) attempted to uncouple effects of the inducing stimuli from post-perceptual processing by using stimuli that varied on two feature dimensions, asking participants to report only one feature on each trial. Importantly, participants were not aware about which feature they would need to report, and so the task required participants to attend to both features. They observed attractive dependencies between trials in which participants reported the same feature and repulsive dependencies between trials in which participants reported different features (i.e., the unreported feature on trial $n-1$ repelled the report of that feature on trial n), concluding that encoding the inducing stimulus was insufficient to cause an attractive dependence. But a perceptual account could still escape falsification by claiming that encoding the inducer is sufficient to cause attractive dependence, yet the attraction may be reduced by other processes. That is, the effects (e.g., an attraction toward both features of the inducer) of one event (e.g., encoding the inducer's features) can be weakened or reversed by a second event (e.g., reporting on only one feature). When the first event is always followed by the second, then the effects of the first will be

unobservable. There is some support for this interpretation in the results presented by Bae and Luck. In their first two experiments, participants always reported one feature, direction of motion in a random dot kinematogram, by making a continuous response. Participants reported the second feature, the color of the dots, with a two-alternative forced-choice. It was in these two experiments that, when participants reported the color of the inducer, reports about the target's direction were repelled away from the inducer's direction. But switching between the two response modalities may have affected that dependence; when participants reported the inducer's color by selecting a value from a continuous and circular color wheel, reports about the target's direction were instead always attracted to the inducer's direction (see also the attractive dependencies reported in a similar experiment by Fischer et al., 2020). Hence, it remains unclear whether an attractive dependence requires participants to make a decision about the inducing stimulus.

APPENDIX B

POWER ANALYSIS OF SERIAL DEPENDENCE AND BAYESIAN MODEL RECOVERY

A frequentist power analysis was conducted to guide how many participants were recruited for Experiment 1 in Chapter 1, and a related model recovery simulation was performed to validate the Bayesian model. The primary inferential questions of Chapter 1 hinge on estimated amplitudes of the derivative of von Mises, an amplitude that might indicate either an attractive or repulsive effect of either the previous orientation or the previous response. These checks were designed to 1) roughly calculate how many participants would be required to reliably reject a null model, and 2) explore the ability of the Bayesian model to adjudicate between one versus two dependencies. Both the power analysis and the model recovery simulations provide information about the former, but only the model recovery provides information about the latter.

Statistical significance in studies of serial dependence are usually reported based on a permutation test for the estimated amplitude of a derivative of a Gaussian function, collapsing across participants. Unfortunately, both trial-to-trial variability in orientation responses and variability in the estimated amplitude across participants are not often reported, complicating a power analysis for a permutation test. Moreover, the statistical tests reported in this chapter were not based on permutation tests, but instead Bayesian methods. As a rough approximation, an analytic power calculation was instead conducted based on a mixed-effects linear model. Estimates of the amplitude of the derivative of Gaussian vary, but peak around a few degrees (Figure 1.3C). I used the conservative estimate of 1°. The one study that reported

variability in the amplitude across participants estimated a standard deviation of 0.91° (Fischer & Whitney, 2014, based on four participants). Assuming a standard deviation of 1° both across participants and across conditions, the sampling distribution for this average will have standard error of $\sqrt{3/2n}$, where n is the number of participants (Wood, 2017, section 2.4). With these approximations, power to detect a single amplitude with an alpha of 0.05 would exceed 0.8 only if there are at least 13 participants.

To assess whether the Bayesian analyses can distinguish between a single versus two dependencies, a model recovery simulation was conducted. Datasets for a single participant were simulated using simplified versions of the derivative of von Mises model (Appendix D)[1]. To analyze the simulated data, I made two simplifications. First, the simplified derivative of von Mises model excluded biases due to oblique effects and static clockwise/counterclockwise biases. Second, the simplified model was non-hierarchical, appropriate for one participant. There were two versions of this simplified model. One version, a full model, included two dependencies, whereas the other included only a single dependency.

These two versions of the von Mises models were used to simulate datasets. To simulate a dataset, parameters were first drawn from the prior distribution of the model. Since the priors do not constrain the sign of the amplitude parameters, this meant that a dataset simulated from the full model might include two attractive dependencies, two negative dependencies, or one attractive and one negative dependence. Using the sampled parameters, the responses of a single participant were simulated (per dataset, either 5,000; 10,000; or 15,000 trials). When generating data, the inputs to the von Mises were calculated before calculating the errors. For ex-

[1]Only a single participant was used in these recovery simulations given A) the lack of reliable estimates of the amplitude across participants, and B) multiple participants were not required to answer whether the analyses were equipped to distinguish between one versus two dependencies.

ample, in the reduced model, every trial was first associated with a random value between $[-\pi, \pi)$, which was used as the input to a derivative of von Mises (whose parameters were given by the draw from the prior). The output of the von Mises was then summed with a Gaussian noise term, producing the error for a trial. In the full model, the procedure was the same, except every trial was associated with two independent inputs to two independent von Mises (analogous to the inducing orientation and participants' response to that orientation). Both the (simplified) reduced and full model were then fit to each dataset. The models were then compared with PSIS-LOO. This procedure was repeated 100 times for each trial count.

The generating model was often the preferred model (Figure B.1). The procedure was somewhat asymmetric, with the reduced model being chosen correctly at a higher rate than the full model. However, after approximately 10,000 trials, the full model was chosen correctly at least 80 % of the time.

These analyses were designed to show two things. First, that Experiment 1 of Chapter 1 would provide enough data to make reliable conclusions about serial dependence, and second that one versus two dependencies were identifiable with the Bayesian analyses. The 80 % cutoff highlighted in this appendix is arbitrary and mentioned here out of convention. In the experiment, sixteen participants contributed usable data. If a frequentist test were performed on estimates of the amplitude of dependencies in those participants (sixteen), the analytic power calculation suggests that the test would be likely to have enough observations to detect a true effect. Each participant contributed 1,515 trials, for a total of 24,240 trials. The model recovery shows that, with this amount of data, the Bayesian analysis is likely able to identify one versus two dependencies.

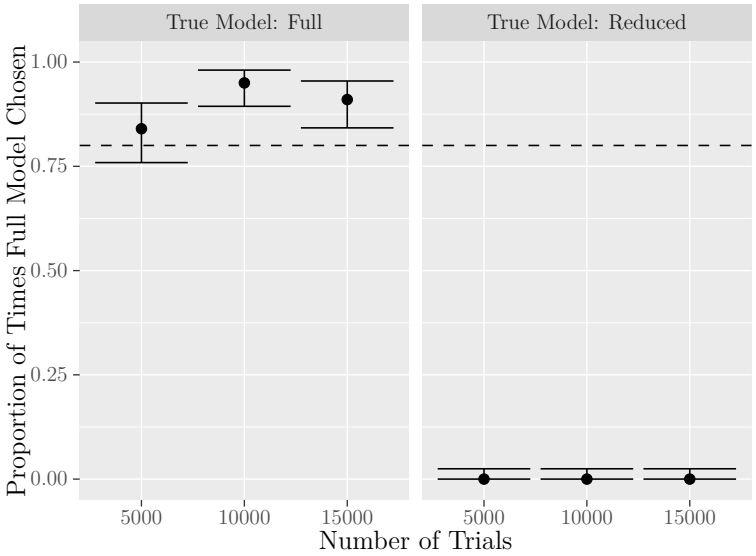


Figure B.1: One versus two dependencies are often recoverable with sufficient trials. The plot shows the proportion of times the Bayesian analysis as applied to simulated data chose the full model, out of 100 simulated datasets. Dots mark the average proportion, and error bars encompass the 95 % highest density interval of the posterior distribution for the proportion parameter, given a beta prior and binomial likelihood (Jeffreys' prior). The two panels indicate whether the simulated datasets were generated with the full (i.e., two effects) or reduced (i.e., one effect) model. Dashed line marks 0.8.

APPENDIX C

THE DERIVATIVE OF GAUSSIAN WILL LEAD TO QUALITATIVELY INCORRECT CONCLUSIONS WHEN THERE ARE PERIPHERAL BUMPS

Although convenient, modeling serial dependence with the derivative of Gaussian (or von Mises) would be inappropriate if the dependence exhibited "peripheral bumps". They can be seen in Figure C.1 by the presence of two peaks on either side of zero in the dashed line. These peripheral bumps label how the sign of the dependence between trials in can be both attractive and repulsive. For example, although the direct tilt aftereffect is repulsive, when the inducer and test stimuli differ by more than $\sim 45^{\circ}$ then perception is often attractive toward the stimulus (Clifford et al., 2000; Gibson & Radner, 1937). If the peripheral bumps are large enough, the best fitting derivative may ignore the central peaks and only account for the bumps, reversing the sign of the derivative (Figure C.1).

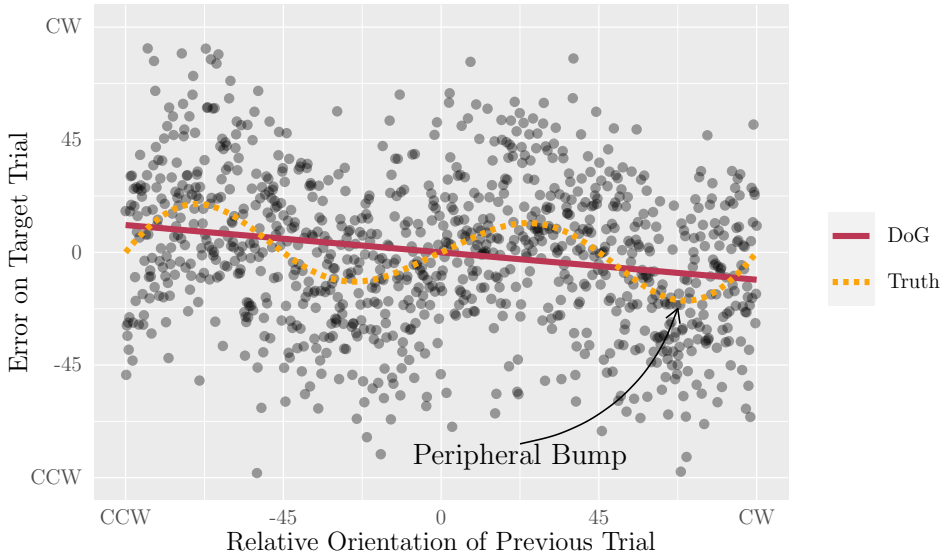


Figure C.1: Misfits of the Gaussian's derivative. The dots are simulated trials, simulated with a function (dashed line) that exhibited peripheral bumps. The data were fit with a derivative of Gaussian function, and the best-fitting derivative is shown with a solid line. The derivative does not match the data-generating function, and its sign mischaracterizes the sign of the generating function at low orientation differences (i.e., at values close to 0 along the x-axis). It is at these low values where both the serial dependence effect and the tilt aftereffect tend to peak.

APPENDIX D

BAYESIAN MODELS OF SERIAL DEPENDENCE

To account for the circular nature of orientation, the commonly used derivative of Gaussian model was replaced with a derivative of the density function of a von Mises distribution. The two derivatives look similar and have the same number of parameters (with the same interpretation) but working with the von Mises ensures that function is continuous at $\pm 90°$.[1]

The density function of a von Mises distribution with a width parameter, w, for an orientation, x (ignoring for a moment, whether the orientation is reported or presented), centered on orientation 0 is given by

$$f(x|w) = \frac{\exp\left(w\cos\left(x\right)\right)}{I_0(w)} \tag{D.1}$$

In that density function, $I_0(\cdot)$, is the modified Bessel function of the first kind of order 0. It normalizes the density function so that its integral is 1 between $[-\pi, \pi)$. The derivative of this density function with respect to x is equal to the following.

$$f'(x|w) = \frac{w\sin\left(x\right)\exp\left(w\cos\left(x\right)\right)}{I_0(w)} \tag{D.2}$$

This derivative has only a single parameter, which governs both the width and height. To serve as a model for serial dependence, this derivative must be rescaled. The resulting function (Equation D.3) gives the component of the error on each trial

[1]The density function for a von Mises is periodic every 2π, but gratings are period every π. To account for this, orientations were always doubled during model fitting.

that is due to a dependence. The cause of this dependence, x, may be the difference between the target orientation and either the inducing orientation or the response to that orientation, depending on which dependencies are included in a model.

$$f(x|a,w) = \frac{aw\sin(x)\exp(w\cos(x))}{I_0(w)f'\left(2\arctan\left(\sqrt{\sqrt{4w^2+1}-2w}\right)\middle|w\right)} \tag{D.3}$$

The value $2\arctan\left(\sqrt{\sqrt{4w^2+1}-2w}\right)$ is the orientation at which the derivative peaks, itself a function of w. So, the rescaling divides the derivative by its maximum (forcing the maximum to 1) and then multiplies by a. With this rescaling, the parameter a can be interpreted as the maximum error, which is equivalent to the amplitude of the derivative of Gaussian[2].

Two additional sources of bias were incorporated into all models (i.e,. all versions of the von Mises models and the spline model). First, participants may exhibit a general clockwise or counterclockwise bias, an offset that affects all orientations equally (all prior studies analyzing serial dependence that account for this bias do so in a preprocessing step, removing the average orientation from each participant prior to model fitting). Second, when reporting orientations, humans tend to be more erroneous on those orientations which are intermediate to the cardinal and oblique axes (Appelle, 1972; Jastrow, 1892; Wei & Stocker, 2015). This anisotropy was also present in Experiment 1 (Figure D.1). Some serial dependence studies have accounted for this anisotropy through preprocessing (Pascucci et al., 2019), but not all. I accounted for it in the Bayesian model by including a sinusoidal term that cycled twice (e.g., contained two peaks between 0 and 180°). A schematic of the hierarchical Bayesian model is given in Figure D.2. This diagram is designed to give a high-level overview of the relationships among parameters, and how the model relates these

[2]Note that this rescaled derivative can be simplified further – optimized for computation during model fitting. It is presented in this relatively raw form since the simplification does not lend further extra insight into the function.

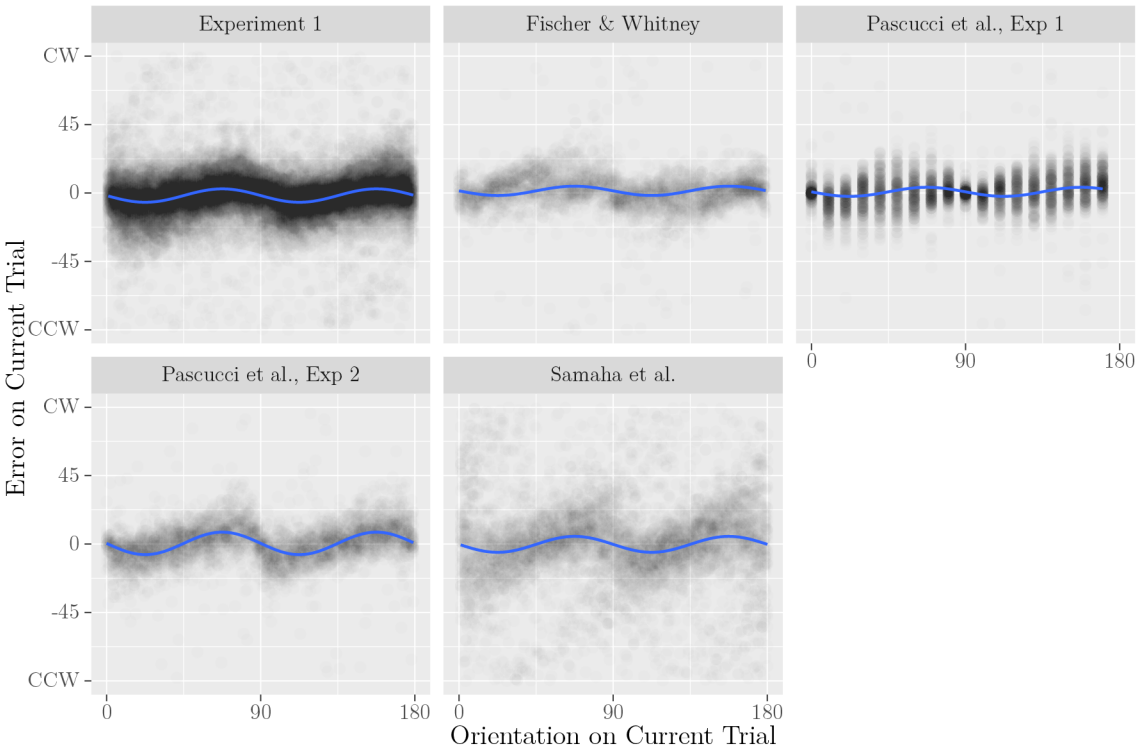


Figure D.1: Errors were largest on orientations intermediate to the cardinal and oblique axes. Each point corresponds to the error on a single trial. The x-axis follows the convention that 0° and 180° are horizontal and increasingly positive orientations are more counterclockwise. Blue lines indicate best fitting sinusoids (minimum squared error).

parameters to the data. The distributions assigned to each parameter that is not a prior is listed to the right of the diagram, and the priors are listed with the square notes.

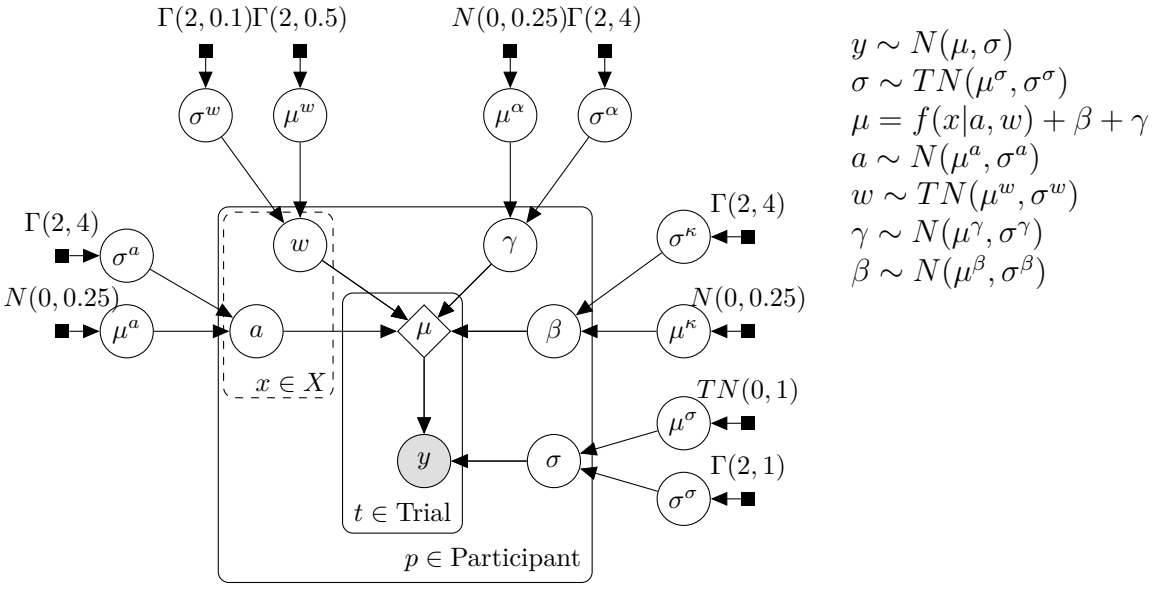


Figure D.2: Hierarchical Bayesian Model of Serial Dependence. Observed data, y, are indicated with a shaded node. They were modeled with a normal distribution with standard deviation σ and location μ. The parameter μ is the output of a deterministic function, the summation of biases due to the oblique effect, clockwise/counterclockwise biases, and serial dependencies caused by either the previous response, the previous orientation, or both. Nodes are grouped with the square "plates", indicating over which subsets of the data the node is replicated. The dashed plate around a and w is to indicate where the three von Mises models differed; in two models, there was a single x, but in the full model there were two kinds x. The magnitudes of the summands are given by β, γ, and a. The parameter a is the amplitude of the rescaled derivative of von Mises with width w (Equation D.3). These four parameters – β, γ, a, and w – were estimated for each participant, hierarchically. These hierarchies were modeled with a normal distribution for each of β, γ, and a, and a half-normal distribution for w. The location and scale of these normal (or truncated-normal, truncated at 0) distributions are given by a μ and σ in the diagram, respectively. The priors on the population-level effects are given by the filled square nodes. $N(\mu, \sigma)$: Normal with location μ and scale σ; TN: truncated-normal location μ and scale σ; $\Gamma(\zeta, \tau)$: Gamma with shape ζ and rate τ.

APPENDIX E

GLOSSARY OF STATISTICAL TERMINOLOGY

Bayes' Rule gives that $p(\theta|y) \propto p(\theta)p(y|\theta)$, that the posterior distribution, $p(\theta|y)$ of the parameters, θ, is proportional to a prior distribution on those parameters, $p(\theta)$, multiplied by the likelihood of the data, $p(y|\theta)$. The rule specifies a procedure for using observed data, y, to learn about the parameters, but it does not give a procedure for making decisions based on those parameters. In this dissertation, the statistical decisions are based primarily on the ability of different models to predict new data, $\tilde{y}$, as approximated by PSIS-LOO+[1]. However, this approach is relatively rare in cognitive psychology. For this reason, this appendix provides a brief glossary. The notation closely follows that of Vehtari et al. (2017).

Posterior Predictive Distribution:A distribution indicating what new data are likely under a posterior distribution. $p(\tilde{y}_i|y) = \int p(\tilde{y}_i|\theta)p(\theta|y)d\theta$. Note that the distribution is for a single new observation, $\tilde{y}_i$.

ELPD: Expected log, pointwise predictive density. This is an expectation of the log score of the posterior predictive distribution for each observation, summed across all observations in a dataset: $\sum_{i=1}^{n} \int p_t(\tilde{y}_i) \log(\tilde{y}_i|y)\, d\tilde{y}_i$. The distribution $p_t(\tilde{y}_i)$ is the true (unknown) data-generating distribution for observation $\tilde{y}_i$ (Geisser & Eddy, 1979; Gneiting & Raftery, 2007). Roughly, the ELPD measures the predictive ability of a model by first scoring how likely a model "thinks" a particular observation will

[1]The more familiar Akaike Information Criterion (AIC) approximates a similar quantity as does PSIS-LOO+, but the AIC is both calculated slightly differently and carries much stronger assumptions about the underlying model (Akaike, 1974; Vehtari & Ojanen, 2012).

be (through the log of the posterior predictive distribution), then weighting that score by the probability of observing that observation (through the true, data generating distribution), and finally combining these weighted scores across all possible values of each observation and all observations in a dataset (the integration and summation).

Bayesian LOO CV: Bayesian Leave-one-out Cross-Validation. Can be used to approximate the ELPD. $\sum_{i=1}^{n} \log \int p(y_i|\theta)p(\theta|y_{i-1})d\theta$, where y_{-i} is a dataset with an observation, i, excluded. The left-out observation is thought of as a sample of the data-generating distribution, $p_t(\tilde{y}_i)$, allowing the ELPD to be approximated despite not knowing the form of that distribution.

PSIS-LOO: Pareto-Smoothed Importance Sampling version of LOO CV, a computationally efficient method of approximating the Bayesian LOO CV, which does not require recalculating a posterior distribution n times.

$\hat{\mathbf{k}}$: k-hat statistic. The statistic measures of the pointwise accuracy of PSIS-LOO and is used to assess the validity of the cross-validation approximation for each observation. Following the recommendation of Vehtari et al. (2017), the approximation was marked as invalid when the statistic exceeded 0.7.

PSIS-LOO+: Another approximation to the ELPD, in which the PSIS-LOO approximation for those observations that have problematic $\hat{k}$ have been replaced by direct calculations of the Bayesian LOO CV.

APPENDIX F

POPULATION RECEPTIVE FIELD MAPPING

Many neurons in visual cortex, especially in earlier visual cortex, are said to be "tuned" to spatial location. Spatial tuning refers to the propensity of a neuron to become more or less active as some visual object is moved across a subjects' field of view. Such tuning could be measured at the single-cell level by recording from a neuron in a subject's visual cortex while they fixate on the center of a computer monitor displaying a moving bar of light. If the neuron were tuned, it would be most active when the bar is in a certain region of the subject's visual field, less active when the bar is near that region, and progressively less active as the bar moves farther away. The function relating the neuron's activity to the position of the stimulus in the subject's visual field is the tuning function, also referred to as the neuron's receptive field in this case of tuning for spatial location. The existence of spatial tuning implies that neurons in certain regions of visual cortex will only be driven by stimulation in certain parts of a subject's visual field. Hence, knowing the spatial tuning properties of neurons facilitates studying the responses of visual cortex.

But as is the topic of Chapter 2, the tuning properties of individual neurons cannot be measured with fMRI. However, the goal of figuring out which regions of cortex are stimulated by which parts of the visual field does not require precise information about the constituent neurons within a voxel; knowing that even the population is stimulated necessarily implies that at least some neurons within that population are stimulated. The process of uncovering the spatial tuning of voxels in visual cortex is called population receptive field mapping (Dumoulin & Wandell,

2008). The receptive fields can be mapped by recording the activity of each voxel while a participant is shown some visually salient sequence of images, and then fitting a model of the population receptive field (e.g., a bivariate Gaussian). The optimal images for the experiment will depend on which part of visual cortex is the focus of the experiment. For example, a counter-phasing black and white checkerboard might be close to optimal for neurons in primary visual cortex, but the checkerboard would only weakly stimulate neurons in higher level visual regions. The stimuli used in Chapter 2 have been designed to stimulate both low and high-level visual regions (Benson et al., 2018; Kriegeskorte et al., 2008). The exact model for population-level receptive field is currently a matter of debate (Kay et al., 2013; Keliris et al., 2019; Merkel et al., 2018; Zeidman et al., 2017; Zuiderbaan et al., 2012). The compressive spatial summation model was used in Chapter 2 because this model has been shown to adequately fit fMRI data generated with the same stimuli, in a large dataset from the Human Connectome Project (Benson et al., 2018; Kay et al., 2013).

Population receptive field mapping can additionally take advantage of the retinotopic arrangement of spatial tuning. A retinotopic arrangement means that neighboring neurons are tuned to neighboring locations in the visual field. Therefore, the activity of a voxel will tend to reflect the activity of neurons tuned to a similar region of the visual field. This arrangement produces something similar to the assumption of unimodal weight distributions (Figure 2.6). But spatial tuning is even simpler than orientation tuning. While each orientation is preferred by many groups of spatially disparate neurons within a single visual region (e.g., within V1), each part of the visual field is only preferred by neurons in a small, spatially contiguous region of cortex. Moreover, spatial tuning changes only gradually between voxels, so the tuning of one voxel is highly informative about the spatial tuning of neighboring voxels. This kind of spatial autocorrelation might partially account for why even just a prior on spatial

tuning functions can predict so well a voxel's receptive field (Benson & Winawer, 2018).

BIBLIOGRAPHY

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00014

Afgoustidis, A. (2015). Orientation maps in V1 and non-Euclidean geometry. *The Journal of Mathematical Neuroscience*, *5*(1), 12. https://doi.org/10.1186/s13408-015-0024-7

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Akaishi, R., Umeda, K., Nagase, A., & Sakai, K. (2014). Autonomous mechanism of internal choice estimate underlies decision inertia. *Neuron*, *81*(1), 195–206. https://doi.org/10.1016/j.neuron.2013.10.018

Alais, D., Kong, G., Palmer, C., & Clifford, C. (2018). Eye gaze direction shows a positive serial dependency. *Journal of Vision*, *18*(4), 11. https://doi.org/10.1167/18.4.11

Alais, D., Leung, J., & Van der Burg, E. (2017). Linear summation of repulsive and attractive serial dependencies: Orientation and motion dependencies sum in motion perception. *The Journal of Neuroscience*, *37*(16), 4381–4390. https://doi.org/10.1523/JNEUROSCI.4601-15.2017

Alitto, H. J., & Usrey, W. M. (2004). Influence of contrast on orientation and temporal frequency tuning in ferret primary visual cortex. *Journal of Neurophysiology*, *91*(6), 2797–2808. https://doi.org/10.1152/jn.00943.2003

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin*, *78*(4), 266–278. https://doi.org/10.1037/h0033117

Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Bae, G.-Y., & Luck, S. J. (2020). Serial dependence in vision: Merely encoding the previous-trial target is not enough. *Psychonomic Bulletin & Review*, *27*(2), 293–300. https://doi.org/10.3758/s13423-019-01678-7

Bao, P., Purington, C. J., & Tjan, B. S. (2015). Using an achiasmic human visual system to quantify the relationship between the fMRI BOLD signal and neural response. *eLife*, *4*, 1–21. https://doi.org/10.7554/elife.09600

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Benson, N. C., Jamison, K. W., Arcaro, M. J., Vu, A. T., Glasser, M. F., Coalson, T. S., Van Essen, D. C., Yacoub, E., Ugurbil, K., Winawer, J., & Kay, K. (2018). The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. *Journal of Vision*, *18*(13), 23. https://doi.org/10.1167/18.13.23

Benson, N. C., & Winawer, J. (2018). Bayesian analysis of retinotopic maps. *eLife*, *7*. https://doi.org/10.7554/eLife.40224

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo.

Bharmauria, V., Bachatene, L., & Molotchnikoff, S. (2019). The speed of neuronal adaptation: A perspective through the visual cortex. *European Journal of Neuroscience*, ejn.14393. https://doi.org/10.1111/ejn.14393

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

Boynton, G. M. (2005). Imaging orientation selectivity: Decoding conscious perception in v1. *Nature Neuroscience*, *8*(5), 541–542. https://doi.org/10.1038/nn0505-541

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *Journal of Neuroscience*, *29*(44), 13992–14003.

Brouwer, G. J., & Heeger, D. J. (2011). Cross-orientation suppression in human visual cortex. *Journal of Neurophysiology*, *106*(5), 2108–2119.

Carlson, T. A. (2014). Orientation decoding in human visual cortex: New insights from an unbiased perspective. *Journal of Neuroscience*, *34*(24), 8373–8383. https://doi.org/10.1523/JNEUROSCI.0548-14.2014

Carp, J. (2012). On the plurality of (methodological) worlds: Estimating the analytic flexibility of fMRI experiments. *Frontiers in Neuroscience*, *6*, 1–13. https://doi.org/10.3389/fnins.2012.00149

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2016). Stan : A probabilistic programming language. *Journal of Statistical Software*, *10*(2), 1–43. https://doi.org/10.18637/jss.v076.i01

Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences*, *111*(21), 7867–7872. https://doi.org/10.1073/pnas.1402785111

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

Cicchini, G. M., Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1890). https://doi.org/10.1098/rspb.2018.1722

Clifford, C. W. G., Wenderoth, P., & Spehar, B. (2000). A functional angle on some after-effects in cortical vision. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *267*(1454), 1705–1710. https://doi.org/10.1098/rspb.2000.1198

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, and Computers*, *34*(4), 613–617. https://doi.org/10.3758/BF03195489

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature neuroscience*, *16*(6), 763–70. https://doi.org/10.1038/nn.3381

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis I. segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Dosher, B. A., & Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the*

National Academy of Sciences, *95*(23), 13988–13993. https://doi.org/10.1073/pnas.95.23.13988

Dragoi, V., Sharma, J., & Sur, M. (2000). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron*, *28*(1), 287–298. https://doi.org/10.1016/S0896-6273(00)00103-3

Draper, D. (1995). Assessment an propogation of model uncertainty. *Journal of the Royal Statistical Society*, *57*(1), 45–97. https://doi.org/10.2307/2346087

Duane, S., Kennedy, A., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, *195*(2), 216–222. https://doi.org/10.1016/0370-2693(87)91197-X

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. https://doi.org/10.1016/j.neuroimage.2007.09.034

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*(2), 181–192. https://doi.org/10.1093/cercor/7.2.181

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent James D. andGoncalves, M., DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIPrep. *Software*. https://doi.org/10.5281/zenodo.852659

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., & Gorgolewski, K. J. (2018). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*. https://doi.org/10.1038/s41592-018-0235-4

Fischer, C., Czoschke, S., Peters, B., Rahm, B., Kaiser, J., & Bledowski, C. (2020). Context information supports serial dependence of multiple visual objects across memory episodes. *Nature Communications*, *11*(1), 1932. https://doi.org/10.1038/s41467-020-15874-w

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. (2018a). 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. (2018b). Serial dependence in numerosity perception. *Journal of Vision*, *18*(9), 15. https://doi.org/10.1167/18.9.15

Fornaciai, M., & Park, J. (2019). Spontaneous repulsive adaptation in the absence of attractive serial dependence. *Journal of Vision*, *19*(5), 21. https://doi.org/10.1167/19.5.21

Fornaciai, M., & Park, J. (2020). Attractive serial dependence between memorized stimuli. *Cognition*, *200*, 104250. https://doi.org/10.1016/j.cognition.2020.104250

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage*, *7*(1), 30–40. https://doi.org/10.1006/nimg.1997.0306

Fritsche, M., & de Lange, F. P. (2019a). Reference repulsion is not a perceptual illusion. *Cognition*, *184*(December 2018), 107–118. https://doi.org/10.1016/j.cognition.2018.12.010

Fritsche, M., & de Lange, F. P. (2019b). The role of feature-based attention in visual serial dependence. *Journal of Vision*, *19*(13), 21. https://doi.org/10.1167/19.13.21

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology*, *27*(4), 590–595. https://doi.org/10.1016/j.cub.2017.01.006

Gardner, J. L. (2010). Is cortical vasculature functionally organized? *NeuroImage*, *49*(3), 1953–1956. https://doi.org/10.1016/j.neuroimage.2009.07.004

Gardner, J. L., & Liu, T. (2019). Inverted encoding models reconstruct an arbitrary model response, not the stimulus. *eNeuro*, *6*(2), ENEURO.0363–18.2019. https://doi.org/10.1523/ENEURO.0363-18.2019

Garner, W. R. (1953). An informational analysis of absolute judgements of loudness. *Journal of Experimental Psychology*, *46*(5), 8.

Geisser, F., & Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, *74*(365), 153–160.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian data analysis* (Third Edit). Chapman and Hall/CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. I. quantitative studies. *Journal of Experimental Psychology*, *20*(5), 453–467. https://doi.org/10.1037/h0059826

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review*, *108*(1), 33–56. https://doi.org/10.1037/0033-295X.108.1.33

Gilden, D., Thornton, T., & Mallon, M. (1995). 1/f noise in human cognition. *Science*, *267*(5205), 1837–1839. https://doi.org/10.1126/science.7892611

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., & Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Goense, J. B., & Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology*, *18*(9), 631–640. https://doi.org/10.1016/j.cub.2008.03.054

Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, *36*(2), 299–308. https://doi.org/10.1016/S0896-6273(02)00971-6

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, *5*, 13. https://doi.org/10.3389/fninf.2011.00013

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D. D., Wong, J., Loney, F., … Ghosh, S. (2018). Nipype. https://doi.org/10.5281/zenodo.2685428

Greenlee, M. W., & Magnussen, S. (1987). Saturation of the tilt aftereffect. *Vision Research*, *27*(6), 1041–1043. https://doi.org/10.1016/0042-6989(87)90017-4

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. https://doi.org/10.1016/j.tics.2005.11.006

Gutnisky, D. A., & Dragoi, V. (2008). Adaptive coding of visual information in neural populations. *Nature*, *452*(7184), 220–224. https://doi.org/10.1038/nature06563

Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, *21*(4), 1639–1651. https://doi.org/10.1016/j.neuroimage.2003.11.029

Harris, J., & Calvert, J. (1985). The tilt after-effect: Changes with stimulus size and eccentricity. *Spatial Vision*, *1*(2), 113–129. https://doi.org/10.1163/156856885X00134

Harris, J., & Calvert, J. (1989). Contrast, spatial frequency and test duration effects on the tilt aftereffect: Implications for underlying mechanisms. *Vision Research*, *29*(1), 129–135. https://doi.org/10.1016/0042-6989(89)90179-X

Ho, T., Brown, S., van Maanen, L., Forstmann, B. U., Wagenmakers, E.-J., & Serences, J. T. (2012). The optimality of sensory processing during the speed-accuracy tradeoff. *The Journal of Neuroscience*, *32*(23), 7992–8003. https://doi.org/10.1523/jneurosci0340-12.2012

Hoffman, M., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 30.

Hsiao, S. S., Lane, J., & Fitzgerald, P. (2002). Representation of orientation in the somatosensory system. *Behavioural Brain Research*, *135*(1), 93–103. https://doi.org/10.1016/S0166-4328(02)00160-2

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, *148*, 574–91. https://doi.org/10.1113/jphysiol.1959.sp006308

Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Quach, R. (2002). Mechanisms of source confusion and discounting in short-term priming 2: Effects of prime similarity and target duration. *Journal of Experimental Psychology: Learning, Memory,*

and Cognition, *28*(6), 1120–1136. https://doi.org/10.1037//0278-7393.28.6.1120

Jackson, D. A. (1993). Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology*, *74*(8), 2204–2214. https://doi.org/10.2307/1939574

Jacob, L. P. L., Potter, K. W., & Huber, D. E. (In Preparation). A neural habituation account of the negative compatibility effect.

Jang, Y., Wallsten, T. S., & Huber, D. E. (2012). A stochastic detection and retrieval model for the study of metacognition. *Psychological review*, *119*(1), 186–200. https://doi.org/10.1037/a0025960

Jastrow, J. (1892). Studies from the university of wisconsin: On the judgment of angles and positions of lines. *The American Journal of Psychology*, *5*(2), 214. https://doi.org/10.2307/1410867

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jones, M., Curran, T., Mozer, M. C., & Wilder, M. H. (2013). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review*, *120*(3), 628–666. https://doi.org/10.1037/a0033180

Kalm, K., & Norris, D. (2018). Visual recency bias is explained by a mixture model of internal representations. *Journal of Vision*, *18*(7), 1. https://doi.org/10.1167/18.7.1

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. https://doi.org/10.1038/nn1444

Kanai, R., Tsuchiya, N., & Verstraten, F. A. (2006). The scope and limits of top-down attention in unconscious visual processing. *Current Biology*, *16*(23), 2332–2336. https://doi.org/10.1016/j.cub.2006.10.001

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355. https://doi.org/10.1038/nature06713

Kay, K. N., Winawer, J., Mezer, A., & Wandell, B. A. (2013). Compressive spatial summation in human visual cortex. *Journal of Neurophysiology*, *110*(2), 481–494. https://doi.org/10.1152/jn.00105.2013

Keliris, G. A., Li, Q., Papanikolaou, A., Logothetis, N. K., & Smirnakis, S. M. (2019). Estimating average single-neuron visual receptive field sizes by fMRI. *Proceedings of the National Academy of Sciences*, *116*(13), 201809612. https://doi.org/10.1073/pnas.1809612116

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

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Kriegeskorte, N., Cusack, R., & Bandettini, P. (2010). How does an fMRI voxel sample the neuronal activity pattern: Compact-kernel or complex spatiotemporal filter? *NeuroImage*, *49*(3), 1965–1976. https://doi.org/10.1016/j.neuroimage.2009.09.059

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–85. https://doi.org/10.1137/0701007

Leuba, G., & Garey, L. (1989). Comparison of neuronal and glial numerical density in primary and secondary visual cortex of man. *Experimental Brain Research*, *77*(1). https://doi.org/10.1007/BF00250564

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

Liberman, A., Zhang, K., & Whitney, D. (2016). Serial dependence promotes object stability during occlusion. *Journal of Vision*, *16*(15), 16. https://doi.org/10.1167/16.15.16

Liu, T., Cable, D., & Gardner, J. L. (2018). Inverted encoding models of human population response conflate noise and neural tuning width. *The Journal of Neuroscience*, *38*(2), 398–408. https://doi.org/10.1523/JNEUROSCI.2453-17.2017

Magnussen, S., & Kurtenbach, W. (1980). Adapting to two orientations: Disinhibition in a visual aftereffect. *Science*, *207*(4433), 908–909. https://doi.org/10.1126/science.7355271

Magnussen, S., & Johnsen, T. (1986). Temporal aspects of spatial adaptation. a study of the tilt aftereffect. *Vision Research*, *26*(4), 661–672. https://doi.org/10.1016/0042-6989(86)90014-3

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

Mardia, K. V., & Jupp, P. E. (1999). *Directional statistics*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316979

Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, *49*(5), 1127–1147. https://doi.org/10.1152/jn.1983.49.5.1127

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

McKenna, F. P. (1984). Assimilation and contrast in perceptual judgments. *The Quarterly Journal of Experimental Psychology Section A*, *36*(3), 531–548. https://doi.org/10.1080/14640748408402176

Merkel, C., Hopf, J. M., & Schoenfeld, M. A. (2018). Spatial elongation of population receptive field profiles revealed by model-free fMRI back-projection. *Human Brain Mapping*, *39*(6), 2472–2481. https://doi.org/10.1002/hbm.24015

Mitchell, T., Shinkareva, S., & Carlson, A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191–1195. https://doi.org/10.1126/science.1152876

Morant, R. B., & Mikaelian, H. H. (1960). Inter-field tilt after-effects. *Perceptual and Motor Skills*, *10*(2), 95–98.

Nevado, A., Young, M. P., & Panzeri, S. (2004). Functional imaging and neural information coding. *NeuroImage*, *21*(3), 1083–1095. https://doi.org/10.1016/j.neuroimage.2003.10.043

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation:

Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, *89*(13), 5951–5955. https://doi.org/10.1073/pnas.89.13.5951

Ogawa, S., & Lee, T.-M. (1990). Magnetic resonance imaging of blood vessels at high fields:in vivo andin vitro measurements and image simulation. *Magnetic Resonance in Medicine*, *16*(1), 9–18. https://doi.org/10.1002/mrm.1910160103

Papadimitriou, C., White, R. L., & Snyder, L. H. (2017). Ghosts in the machine II: Neural correlates of memory interference from the previous trial. *Cerebral Cortex*, *27*(4), 2513–2527. https://doi.org/10.1093/cercor/bhw106

Parker, D. M. (1972). Contrast and size variables and the tilt after-effect. *Quarterly Journal of Experimental Psychology*, *24*(1), 1–7. https://doi.org/10.1080/14640747208400260

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

Patterson, C. A., Wissig, S. C., & Kohn, A. (2013). Distinct effects of brief and prolonged adaptation on orientation tuning in primary visual cortex. *Journal of Neuroscience*, *33*(2), 532–543. https://doi.org/10.1523/JNEUROSCI.3345-12.2013

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, *22*(8), 1336–1344. https://doi.org/10.1038/s41593-019-0428-x

Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, *53*(4), 1181–1196. https://doi.org/10.1016/j.neuroimage.2010.07.020

Roth, Z. N., Heeger, D. J., & Merriam, E. P. (2018). Stimulus vignetting and orientation selectivity in human visual cortex. *eLife*, *7*, 1–25. https://doi.org/10.7554/eLife.37241

Samaha, J., Barrett, J. J., Sheldon, A. D., LaRocque, J. J., & Postle, B. R. (2016). Dissociating perceptual confidence from discrimination accuracy reveals no

influence of metacognitive awareness on working memory. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00851

Samaha, J., Switzky, M., & Postle, B. R. (2019). Confidence boosts serial dependence in orientation estimation. *Journal of Vision*, *19*(4), 25. https://doi.org/10.1167/19.4.25

Saproo, S., & Serences, J. T. (2014). Attention improves transfer of motion information between V1 and MT. *Journal of Neuroscience*, *34*(10), 3586–3596. https://doi.org/10.1523/JNEUROSCI.3484-13.2014

Saproo, S., & Serences, J. T. (2010). Spatial attention improves the quality of population codes in human visual cortex. *Journal of Neurophysiology*, *104*(2), 885–895. https://doi.org/10.1152/jn.00369.2010

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews Neuroscience*, *8*(7), 522–535. https://doi.org/10.1038/nrn2155

Sclar, G., & Freeman, R. D. (1982). Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Experimental Brain Research*, *46*(3), 457–461.

Scolari, M., & Serences, J. T. (2010). Basing perceptual decisions on the most informative sensory neurons. *Journal of Neurophysiology*, *104*(4), 2266–2273. https://doi.org/10.1152/jn.00273.2010

Scolari, M., Byers, A., & Serences, J. T. (2012). Optimal deployment of attentional gain during fine discriminations. *Journal of Neuroscience*, *32*(22), 7723–7733. https://doi.org/10.1523/JNEUROSCI.5558-11.2012

Sekuler, R., & Littlejohn, J. (1974). Tilt aftereffect following very brief exposures. *Vision Research*, *14*(1), 151–152. https://doi.org/10.1016/0042-6989(74)90133-3

Serences, J. T., & Saproo, S. (2010). Population response profiles in early visual cortex are biased in favor of more valuable stimuli. *Journal of Neurophysiology*, *104*(1), 76–87. https://doi.org/10.1152/jn.01090.2009

Serences, J. T., Saproo, S., Scolari, M., Ho, T., & Muftuler, L. T. (2009). Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *NeuroImage*, *44*(1), 223–231.

Shmuel, A., Yacoub, E., Pfeuffer, J., Van de Moortele, P.-F., Adriany, G., Hu, X., & Ugurbil, K. (2002). Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human

brain. *Neuron*, *36*(6), 1195–1210. https://doi.org/10.1016/S0896-6273(02)01061-9

Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., & Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: A comparison of single cells and behavior. *Journal of neurophysiology*, *57*(3), 773–786.

Sprague, T. C., Adam, K. C. S., Foster, J. J., Rahmati, M., Sutterer, D. W., & Vo, V. A. (2018). Inverted encoding models assay population-level stimulus representations, not single-unit neural tuning. *Eneuro*, *5*(3), ENEURO.0098–18.2018. https://doi.org/10.1523/ENEURO.0098-18.2018

Sprague, T. C., Boynton, G. M., & Serences, J. T. (2019). The importance of considering model choices when interpreting results in computational neuroimaging. *eneuro*, *6*(6), ENEURO.0196–19.2019. https://doi.org/10.1523/ENEURO.0196-19.2019

Sprague, T. C., Saproo, S., & Serences, J. T. (2015). Visual attention mitigates information loss in small- and large-scale neural codes. *Trends in Cognitive Sciences*, *19*(4), 215–226. https://doi.org/10.1016/j.tics.2015.02.005

Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience*, *16*(12), 1879–1887. https://doi.org/10.1038/nn.3574

St John-Saaltink, E., Kok, P., Lau, H. C., & de Lange, F. P. (2016). Serial dependence in perceptual decisions is reflected in activity patterns in primary visual cortex. *The Journal of Neuroscience*, *36*(23), 6186–92. https://doi.org/10.1523/JNEUROSCI.4390-15.2016

Stan Development Team. (2019). RStan: The R interface to Stan.

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

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, *6*(1), 32239. https://doi.org/10.1038/srep32239

Taubert, J., Van der Burg, E., & Alais, D. (2016). Love at second sight: Sequential dependence of facial attractiveness in an on-line dating paradigm. *Scientific Reports*, *6*(1), 22740. https://doi.org/10.1038/srep22740

Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology*, *19*(1), 11–14. https://doi.org/10.1016/j.cub.2008.10.014

Tustison, N. J., Avants, B. B., Cook, P. A., Yuanjie Zheng, Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

van Bergen, R. S., & Jehee, J. F. (2019). Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. *The Journal of Neuroscience*, *39*(41), 8164–8176. https://doi.org/10.1523/JNEUROSCI.3212-18.2019

Vehtari, A., Gabry, J., Yao, Y., & Gelman, A. (2020). Loo: Efficient leave-one-out cross-validation and WAIC for bayesian models.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vehtari, A., Gelman, A., & Gabry, J. (2019). Pareto smoothed importance sampling. *arXiv*.

Vehtari, A., & Ojanen, J. (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, *6*(1), 142–228. https://doi.org/10.1214/12-SS102

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology*, *48*(1), 28–50. https://doi.org/10.1016/j.jmp.2003.11.004

Wardle, S. G., Ritchie, J. B., Seymour, K., & Carlson, T. A. (2017). Edge-related activity is not necessary to explain orientation decoding in human visual cortex. *The Journal of Neuroscience*, *37*(5), 1187–1196. https://doi.org/10.1523/JNEUROSCI.2690-16.2016

Webb, B. S. (2005). Early and late mechanisms of surround suppression in striate cortex of macaque. *Journal of Neuroscience*, *25*(50), 11666–11675. https://doi.org/10.1523/JNEUROSCI.3414-05.2005

Webster, M. A. (2015). Visual adaptation. *Annual Review of Vision Science*, *1*(1), 547–567. https://doi.org/10.1146/annurev-vision-082114-035509

Wei, X.-X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience*, *18*(10), 1509–1517. https://doi.org/10.1038/nn.4105

Wenderoth, P., & Johnstone, S. (1987). Possible neural substrates for orientation analysis and perception. *Perception*, *16*(6), 693–709. https://doi.org/10.1068/p160693

Wissig, S. C., & Kohn, A. (2012). The influence of surround suppression on adaptation effects in primary visual cortex. *Journal of Neurophysiology*, *107*(12), 3370–3384. https://doi.org/10.1152/jn.00739.2011

Wolfe, J. M. (1984). Short test flashes produce large tilt aftereffects. *Vision Research*, *24*(12), 1959–1964. https://doi.org/10.1016/0042-6989(84)90030-0

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (Second edition). CRC Press/Taylor & Francis Group.

Xia, Y., Leib, A. Y., & Whitney, D. (2016). Serial dependence in the perception of attractiveness. *Journal of Vision*, *16*(15), 28. https://doi.org/10.1167/16.15.28

Yang, T., & Maunsell, J. H. R. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *J. Neurosci.*, *24*(7), 1617–1626. https://doi.org/10.1523/JNEUROSCI.4442-03.2004

Zeidman, P., Silson, E. H., Schwarzkopf, D. S., Baker, C. I., & Penny, W. (2017). Bayesian population receptive field modelling. *NeuroImage*, *180*, 173–187. https://doi.org/10.1016/j.neuroimage.2017.09.008

Zhang, J., Meeson, A., Welchman, A. E., & Kourtzi, Z. (2010). Learning alters the tuning of functional magnetic resonance imaging patterns for visual forms. *Journal of Neuroscience*, *30*(42), 14127–14133. https://doi.org/10.1523/JNEUROSCI.2204-10.2010

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Zirnsak, M., Steinmetz, N. A., Noudoost, B., Xu, K. Z., & Moore, T. (2014). Visual space is compressed in prefrontal cortex before eye movements. *Nature*, *507*(7493), 504–507. https://doi.org/10.1038/nature13149

Zuiderbaan, W., Harvey, B. M., & Dumoulin, S. O. (2012). Modeling center-surround configurations in population receptive fields using fMRI. *Journal of Vision*, *12*(3), 10–10. https://doi.org/10.1167/12.3.10