Perceptual Learning And Visual Short-Term Memory: The Limitations And Mechanisms Of Interacting Processes

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of the Ohio State University

By

Nicholas Michael Van Horn, B.A., M.A.
Graduate Program in Psychology

The Ohio State University

2014

Dissertation Committee:
Alexander A. Petrov, Advisor
James T. Todd
Julie D. Golomb
Abstract

Visual short-term memory (VSTM) is commonly perceived as a temporary buffer into which information is moved for retention across relatively short intervals. Guided by work on the “standard model” of working memory, these buffers are assumed to reside in modality-specific areas of the prefrontal cortex (PFC). Recent converging evidence has begun to call these conventional views into question, with a growing consensus that working memory is the emergent result of attention-guided activity across many areas of the brain. This “emergent-property” view suggests that working memory is sustained in part by the very same sensory areas involved in encoding external input. If visual short-term memory is represented and maintained on the same neural populations involved in common perceptual tasks as evidence suggests, we might expect possible interactions. The current thesis describes a series of experiments designed to build a detailed profile of these interactions. In Chapter 2 we establish the utility of using visual aftereffects as investigative tools in two behavioral experiments. In the first experiment we measure the magnitude of static and dynamic motion aftereffects before and after perceptual learning on a motion discrimination task using identical stimuli. Our results indicate that learning does not affect the duration of the aftereffects, strongly suggesting that improvements in discrimination are not mediated by changes in the underlying perceptual representations of the stimuli. Rather, our evidence supports the view that practice changes the relative contributions of perceptual outputs to decision-making areas. A second experiment rules
out a competing explanation in which the locus of learning does not share enough overlap with the neural populations responsible for the observed aftereffects. Next, in Chapter 3 we demonstrate the efficacy of analyzing multivariate response data over traditional reliance on accuracy or response time alone. We apply the diffusion model (DM) to accuracy and response time distributions on data from two perceptual learning experiments. Results reveal that practice-induced improvements in motion direction discrimination arise from the combination of stimulus-specific improvements in the quality of sensory information entering decision areas, and non-specific gains in the timing of decision-making processes. Following this, in Chapter 4 we leverage the methodological gains of the previous results to test the hypothesis that the contents of VSTM can influence the perception of stimuli during the memory retention interval. We use a dual-task design to test the effect of VSTM on perception. Results suggest that orientation information in memory has a repulsive effect on the perception of subsequently presented orientations, not unlike tilt aftereffects. Further DM analysis confirms that our results are perceptual in nature and do not arise from artifacts in response time shifts. Finally, in Chapter 5 we describe an experiment designed to test if the interactions observed in Chapter 4 share the same neural mechanism as memory masking, and what, if any, effect practice has on these two forms of interference. An asymmetrical effect due to memory-stimulus similarity, as well as contrasting effects due to training on the two sources of interference strongly indicate the involvement of two separate processes. Taken together, this collection of studies highlights the existence of another form of VSTM-perception entanglement that is complementary to memory masking, but behaves much differently, particularly under the influence of practice. The results indicate that VSTM and perception interact in systematic ways that are critical to understand, most notably in laboratory settings where the effects can disrupt the subtle differences upon which observers must make
decisions. A novel model is proposed that parsimoniously reconciles observed patterns of perceptual learning, as well as the apparently disparate forms of interactions measured throughout. Collectively, the present work supports the emergent-property hypothesis of working memory, as well as a revised version of the standard model.
To Tara, Addie, and Evie.
Acknowledgments

Alex Petrov has been a tremendous advisor throughout my time as a student. It was a lecture by him that originally inspired me to enter the field of cognitive science, and it has been an honor to work alongside him in that pursuit. I also owe a debt of gratitude to my wife Tara. Without her, I may never have walked down the path to academia, and it was her continued presence that made the whole journey possible. Finally, I have benefited in innumerable ways from my incredible relationship with my friend Aaron. Our voluminous conversations over email and our late night discussions could fill libraries. This work has indirectly benefited in incalculable ways from my friendship with him.
Vita

2007 .................................  B.A. Japanese, The Ohio State University

2011 .................................  M.A. Cognitive Psychology, The Ohio State University

2008-Present .........................  Graduate Teaching Associate, Department of Psychology, The Ohio State University

Publications


Fields of Study

Major Field: Psychology

Specialization: Cognitive Psychology
# Table of Contents

Abstract  ................................................................. ii
Dedication ................................................................. iv
Acknowledgments ......................................................... vi
Vita ................................................................. vii
List of Tables .......................................................... xi
List of Figures .......................................................... xiii

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>2.1</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>17</td>
</tr>
<tr>
<td>2.2.1</td>
<td>18</td>
</tr>
<tr>
<td>2.2.2</td>
<td>26</td>
</tr>
<tr>
<td>2.3</td>
<td>31</td>
</tr>
<tr>
<td>2.3.1</td>
<td>31</td>
</tr>
<tr>
<td>2.3.2</td>
<td>32</td>
</tr>
<tr>
<td>2.4</td>
<td>34</td>
</tr>
<tr>
<td>2.5</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
</tr>
<tr>
<td>3.1</td>
<td>45</td>
</tr>
<tr>
<td>3.2</td>
<td>50</td>
</tr>
<tr>
<td>3.2.1</td>
<td>50</td>
</tr>
<tr>
<td>3.2.2</td>
<td>50</td>
</tr>
<tr>
<td>3.2.3</td>
<td>50</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.2.4</td>
<td>3.2.4 Data analysis</td>
</tr>
<tr>
<td>3.3</td>
<td>3.3 Results</td>
</tr>
<tr>
<td>3.4</td>
<td>3.4 Discussion</td>
</tr>
<tr>
<td>4</td>
<td>4 The contents of visual working memory systematically bias perception</td>
</tr>
<tr>
<td>4.1</td>
<td>4.1 Introduction</td>
</tr>
<tr>
<td>4.2</td>
<td>4.2 Experiment</td>
</tr>
<tr>
<td>4.2.1</td>
<td>4.2.1 Method</td>
</tr>
<tr>
<td>4.3</td>
<td>4.3 Results</td>
</tr>
<tr>
<td>4.3.1</td>
<td>4.3.1 Mixed-effects regression model</td>
</tr>
<tr>
<td>4.3.2</td>
<td>4.3.2 Diffusion Model &amp; Data Simulations</td>
</tr>
<tr>
<td>4.4</td>
<td>4.4 Discussion</td>
</tr>
<tr>
<td>5</td>
<td>5 Practice-induced changes in bi-directional interference between visual short-term memory and externally-driven visual perception: asymmetric learning and transfer reveals multiple mechanisms</td>
</tr>
<tr>
<td>5.1</td>
<td>5.1 Introduction</td>
</tr>
<tr>
<td>5.2</td>
<td>5.2 Experiment</td>
</tr>
<tr>
<td>5.2.1</td>
<td>5.2.1 Method</td>
</tr>
<tr>
<td>5.3</td>
<td>5.3 Results</td>
</tr>
<tr>
<td>5.3.1</td>
<td>5.3.1 Discrimination Threshold Results</td>
</tr>
<tr>
<td>5.3.2</td>
<td>5.3.2 Orientation Recall Results</td>
</tr>
<tr>
<td>5.4</td>
<td>5.4 Discussion and model</td>
</tr>
<tr>
<td>6</td>
<td>6 General discussion</td>
</tr>
<tr>
<td>6.1</td>
<td>6.1 Future Directions</td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
<tr>
<td>Appendix A: Supplementary information for chapter 3</td>
<td>148</td>
</tr>
<tr>
<td>A.1</td>
<td>A.1 Additional details on the model fits</td>
</tr>
<tr>
<td>A.2</td>
<td>A.2 Motion aftereffect</td>
</tr>
<tr>
<td>A.2.1</td>
<td>A.2.1 MAE Procedure</td>
</tr>
<tr>
<td>A.2.2</td>
<td>A.2.2 Discussion</td>
</tr>
<tr>
<td>Appendix B: Supplementary information for chapter 4</td>
<td>153</td>
</tr>
<tr>
<td>B.1</td>
<td>B.1 Evaluation of alternative mixed-effects regression models</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Top section: Motion-aftereffect (MAE) duration patterns predicted by the selective reweighting (SRW) and representation modification (RM) hypotheses of perceptual learning. ( M ) denotes the MAE duration at pretest and ( c ) is an effect-size parameter. Middle and bottom sections: Coefficients of the planned linear contrasts for the two types of analysis of variance (ANOVA) discussed in Section 2.2.1. *Note: The predictions for the main effect of Session assume (incorrectly) that the response criterion does not change from pretest to posttest.</td>
</tr>
<tr>
<td>3.1</td>
<td>Descriptive statistics of the discriminability ( d' ) for easy and difficult stimulus pairs and for all diffusion-model parameters. The learning and specificity indices were calculated from group-averaged data (±80% bootstrap confidence interval; see note 2 for details). The trend analyses were performed on the individual data using orthogonal contrasts (linear and quadratic). The error term had ( df = 182 ) in all ANOVAs. ( \eta_p^2 ) = partial effect size = variance of the contrast relative to itself plus the error. * ( p &lt; .100 ). ** ( p &lt; .010 ). *** ( p &lt; .001 ).</td>
</tr>
<tr>
<td>4.1</td>
<td>Summary of coefficient estimates from equation (4.3.1) and their conceptual descriptions for the best fitting model. The parameters of primary interest are ( \beta_1 ), which captures the influence of the tilt after-effect (TAE) on response and ( \beta_3 ), which reflects the interaction between VSTM and the TAE. Inherent bias ( \beta_0 ) and VSTM ( \beta_2 ) were both expected to be non-significant because of a pre-experiment calibration routine and theoretical reasoning, respectively.</td>
</tr>
<tr>
<td>4.2</td>
<td>Mean, standard deviation, and quantiles for posterior distributions of diffusion model (DM) parameter ( z ) and DM regression coefficients from equations (4.3.3), (4.3.4), and (4.3.5).</td>
</tr>
</tbody>
</table>
5.1 Estimated parameters from Equation (5.3.1) fit to the orientation discrimination thresholds depicted in Figure 5.2. $\Delta_{\text{start}}$ reflects discrimination thresholds at the beginning of the training phase, $\Delta_{\text{end}}$ is an estimate of peak performance at the end of training, and $\tau$ represents the time course of learning. ........................................ 100

5.2 Visual short-term memory recall precision $p$ as estimated by Equation (5.3.1). $p_{\text{start}}$ represents initial recall precision at the beginning of the training phase, and $p_{\text{end}}$ is an estimate of the final recall precision at the end of training. $\tau$ reflects the learning rate for each exponential curve. Despite improvements in precision for both congruency conditions, the difference between conditions remains despite practice. . . 104
## List of Figures

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The stimuli were filtered-noise textures moving behind a circular aperture. On discrimination trials, the texture moved in one of four possible directions (depicted by arrows) relative to an implicit reference direction (dotted line). The adapting stimuli on motion-aftereffect trials were the same, except that they moved for 10 s rather than 397 ms. [Note: The fixation dot at the center and the angles between the directions are exaggerated for visibility.]</td>
</tr>
<tr>
<td>2.2</td>
<td>Learning curves for the discrimination task. Group average of the 11 participants in Experiment 1. The error bars are 90% within-subject confidence intervals. The shaded areas mark the two motion aftereffect sessions.</td>
</tr>
<tr>
<td>2.3</td>
<td>Mean motion aftereffect (MAE) durations in Experiment 1. The large panel plots the group average. The error bars are 90% within-subject confidence intervals. The pre and post labels on the horizontal axis denote the pre- and posttest MAE sessions, respectively. The small panels plot the mean MAEs for each individual observer (using the same graphical layout). The error bars are 90% confidence intervals estimated from 21 replications per condition. The ( F ) statistic above each panel tests whether the trained direction (solid line) differs from the control direction (dashed line) at posttest. None of the tests is significant (critical ( F(1,160) = 3.90 )).</td>
</tr>
<tr>
<td>2.4</td>
<td>Learning curves for the discrimination task. Group average of the 16 participants in Experiment 2. The error bars are 90% within-subject confidence intervals. The shaded areas mark the two motion aftereffect sessions.</td>
</tr>
</tbody>
</table>
2.5 Mean motion aftereffect (MAE) durations in Experiment 2. The large panel plots the group average. The error bars are 90% within-subject confidence intervals. The pre and post labels on the horizontal axis denote the pre- and posttest MAE sessions, respectively. The small panels plot the mean MAEs for each individual observer (using the same graphical layout). The error bars are 90% confidence intervals estimated from 21 replications per condition. The $F$ statistic above each panel tests the Direction $\times$ Session interaction. Only 1 of the 16 tests is significant (critical $F(3, 160) = 2.66$).

3.1 Schematic illustration of the decision process in the diffusion model. Three stochastic paths result in a fast correct response, a slow correct response, and an error response. Because of the positive drift rate $v$, the correct (upper) boundary is reached more often than the incorrect (lower) boundary. RT, response time; $a$, boundary separation parameter; $z$, starting point parameter.

3.2 The stimuli were filtered-noise textures moving behind a circular aperture. The direction of motion could take four possible values relative to an implicit reference direction (dashed line). The angles are exaggerated for visibility.

3.3 Learning profiles for the group-averaged discriminability (a) and mean response times (b) in the raw data, and for various parameters of the diffusion model (cf.). The observers practiced motion-direction discrimination for eight blocks (black symbols) and then were tested on the same task with motion in the orthogonal direction (open symbols). The error bars are 90% within-subjects confidence intervals. Shaded areas mark two additional sessions of motion aftereffect measurements.

3.4 Quantile probability plots illustrating the wealth of data and the quality of the fit. Each panel has 22 empirical degrees of freedom: the proportions of errors and correct responses for the easy and difficult discriminations (plotted on the x-axis) and the .1, .3, .5, .7, and .9 quantiles of the corresponding response time distributions (stacked vertically on the y-axis). For example, the x-coordinate of the left-most, bottommost data point in the top panel indicates the initial error rate (.18) for easy stimuli. The y-coordinate indicates the leading edge (.1 quantile 530 ms) of the corresponding RT distribution. After 4 days of training (middle panel), the performance improves on both measures (.08 rate and 480 ms, respectively). This illustration is based on group-averaged data; the analyses in the text (and the predictions in Fig. 5) are based on model fits to individual data.
3.5 Scatterplots illustrating the quality of the fit to individual data. The diffusion model was fit separately in each block (297 fits = 27 observers 11 blocks). Each panel contains 594 points (= 2972 difficulty levels). RT, response time

4.1 Sequence of events for trials in our two conditions. In both conditions, observers were shown three sequentially presented Gabors—Sample, Target, and then Match). Observers completed a binary pseudo-discrimination task using the Target Gabor and recreated the orientation of the Sample using the Match Gabor either from memory (VSTM trials) or through direct comparison (no-load trials). An attentional task ensured observers actively viewed the Sample under both conditions (see method). A black circular mask hid the computer monitor’s edges from view to minimize orientation cues.

4.2 Simulated data (lines) and experimental data (points). The solid line represents simulated data under the no-load condition, and was informed by our experimental data. For the VSTM-load condition, data were generated under two different hypotheses. Under hypothesis 1, drift rate is allowed to vary under VSTM-load relative to no-load (dashed line), while the remaining diffusion parameters remain fixed. Under hypothesis 2 (dotted line), boundary separation is allowed to increase uniformly under VSTM-load relative to no-load, while the remaining diffusion parameters remain fixed. Data points represent response frequencies for binned orientations from our experimental data. Circles show no-load trials and Xs show VSTM-load trials. The inset highlights nonparametric kernel density estimates for the simulated response time distributions under no-load (single solid line) and VSTM-load (both broken lines). Although we hypothesize that drift rate changes as a function of memory load, it is clear that uniform changes in response criterion can produce equivalent changes in response probabilities and shifts in RT distributions. Error bars are 1 standard error of the mean.
4.3 In the top panel, responses in diffusion modeling are described as a result of stochastic evidence accumulation. Processes, of which several are depicted here, begin at the starting point \( z \). Perceptual evidence accumulates in a noisy fashion over time, eventually pushing the process to one of two decision boundaries. Response times are reflected in the time taken to reach a decision boundary. The overall rate of evidence accumulation, whether positive or negative, is captured in the model by the drift rate \( v \). Speed-accuracy trade-offs are addressed in the model by the boundary separation \( a \). A smaller \( a \) leads to quicker responses that are more error prone due to noise in the accumulation process. A larger \( a \) leads to slower, but more accurate responses. Bias is captured in the model by \( z \), wherein \( z = a/2 \) for an unbiased observer. The bottom panel depicts two competing hypotheses for explaining our data. Our hypothesis (H1) is that the contents of VSTM act as a concurrent stimulus, leading to an increase in the absolute value of the drift rate (dotted lines) relative to a no-load condition (solid lines). In an alternative hypothesis (H2), drift rate remains the same under both VSTM-load and no-load while the boundary separation increases under load.

4.4 Drift rate \( v \) regression error bands from equation (4.3.3). The steeper slope under VSTM-load (light-gray) relative to no-load (dark-gray) reflects an increased tilt-aftereffect under VSTM-load relative to no-load. In terms of the diffusion model, this effect is primarily due to changes in drift rate (hypothesis 1 from figure 4.3).

5.1 The trial sequence for both experimental conditions involved the sequential presentation of three Gabors—the “Sample,” “Target,” and “Match” Gabors—in left, center, and right positions of the display, respectively. Observers completed two concurrent tasks. In an orientation recall task, participants directly rotated the Match Gabor to the remembered orientation of the Sample Gabor previously displayed. During the memory retention interval, observers classified the displayed Target Gabor orientation as clockwise or counter-clockwise relative to an implicit oblique reference boundary. Depending on the experimental condition, the Sample and Target Gabors were presented along a similar (congruent) or orthogonal (incongruent) orientation. Visual and auditory feedback was provided for both tasks. A black circular mask covering the monitor minimized orientation cues from the experimental apparatus.
5.2 Discrimination thresholds across the across all test and training phases of the experiment. Easy and hard trials have been separated here for clarity, but were randomly intermixed in the actual study. Solid symbols represent pretest and posttest blocks and open symbols reflect training blocks. Blue indicates congruent trials in which the Sample and Target Gabors shared similar orientations, while red indicates incongruent trials in which the Sample and Target were orthogonal. The overlaid curves are exponential fits from Equation (5.3.1), as summarized in Table 5.1. Symbol shape reflects participant group membership. For example, the first filled red circle in each panel reflects a block of pretest incongruent trials. The following series of open blue circles reflects the same group of observers during the training phase in which all trials were of the congruent type. Error bars are 1 standard error of the mean.

5.3 Recall precision, defined as the reciprocal of the standard deviation of memory recall error, across all phases of the study. Symbol type and color follows the same conventions as Figure 5.2. Precision was found to improve for both congruency conditions as a result of practice. However, unlike observed changes in discrimination thresholds, the difference between congruency conditions remained fixed across training, strongly suggesting that the effects of memory masking are a persistent source of interference. Error bars are 1 standard error of the mean.

5.4 Schematic depicting the proposed selective reweighting memory (SR-Memory) model described in the text. The model’s task is to reconcile the asymmetric pattern of results observed in the discrimination and recall tasks. Despite the apparent difficulty of such an endeavor, we believe the model, which adds a bank of VSTM units and relatively assumption-free connectivity to the successful selective reweighting model of perceptual learning, can explain our results in a parsimonious manner.

A.1 Learning profiles for two diffusion-model parameters: across-trial variability in drift rate (A) and in starting point (B). The data format is as in Figure 3.3 in the main text. The observers practiced motion-direction discrimination for 8 blocks (black symbols) and then were tested on the same task at the orthogonal direction (open symbols). The error bars are 90% within-subject confidence intervals. Shaded areas mark two additional sessions of motion-aftereffect measurements.
CHAPTER 1

Introduction

Visual short-term memory (VSTM) is responsible for maintaining perceptual information in awareness for short periods of time after it is no longer physically driven by an external stimulus. This form of “working memory” has been found to be a fundamental player in learning and goal management, the integration of information across eye movements (Irwin, 1991), and deficits in VSTM have been implicated in visuospatial perceptual disorders (Logie & Della Sala, 2005). A commonly used test of VSTM known as “change detection” has proven to be an effective tool for estimating both the capacity of working memory, as well as the fidelity or resolution of memorandum contained in the transient store (see Vogel, Woodman, & Luck, 2001; Awh, Barton, & Vogel, 2007, for examples). For example, a typical change detection task consists of an array of visual items (often simple colored squares or circles) that is briefly presented. After a blank delay period, a second array of items is presented that is either identical to the first, or that differs in the properties of one or more of the array items. The observer is asked to press a button to indicate whether the second stimulus was identical to or different from the first. Capacity can easily be tested by varying the number of items in the array, with impoverished performance observed for greater set sizes (performance losses typically level out with set sizes above 3-5 items, well within the revised “magic number” capacity of working memory (Cowan, 2001; G. A. Miller, 1956)). Precision of working memory is typically quite high, and
can also be tested within the change detection paradigm by varying the amount of change in the second stimulus relative to the first. Despite its popularity and effectiveness as a methodological tool to investigate the nature of visual short-term memory, interpretation of observer responses is quite narrow. That is, behavioral errors in the change detection task are attributed solely to such factors as forgetting, guessing, failure to recall, and confusion due to inter- and intra-item similarity (see Zhang & Luck, 2008; Bays, Catalao, & Husain, 2009; Gorgoraptis, Catalao, Bays, & Husain, 2011, for examples of a popular modeling approach that reduces errors into these constituent components). In other words, mistakes are assumed to arise from limitations in the processes associated with working memory—encoding, storage, and retrieval.

Another fundamental action is at work in the change detection task, however, that is ignored in VSTM research—perceptual discrimination. To determine whether a stimulus is different from a previously presented stimulus, a comparison must be made. To simplify discussions of change detection data, this comparative apparatus is assumed to be perfect, so that any errors in behavior are said to arise from the merits or deficiencies of memory (or its abject failure, forgetting). To be fair, simplifications such as lumping discrimination under the vague umbrella terminology of “recognition” are often necessary when studying such complex systems.

Yet we know from research in other domains that discrimination is an imperfect process. For example, it is well known that repeated exposure to discrimination tasks leads to improvements in observers’ ability to distinguish differences between two stimuli such as contrast, spatial frequency, orientation, and direction of motion (see Fahle & Poggio, 2002, for a review). This amelioration from exposure is known as perceptual learning, and unlike other forms of declarative learning, leads to improvements that often cannot be verbalized. Additionally, in the domain of perceptual
learning, discrimination data are often interpreted under the framework of signal detection theory (Macmillan & Creelman, 2005), wherein observer bias and criterion shifts lead to predictable changes in response behaviors despite the apparent simplicity of most discrimination tasks. It is a reasonable concern to ask if, and in what ways the intricacies of discrimination leads to response artifacts in VSTM change detection tasks, particularly under conditions of subtlety wherein the two stimuli differ only slightly along some dimension. Ignoring the contributions of the discrimination apparatus can lead to artifactual error in the interpretation of experimental results at the behavioral level.

Curiously, the “same-different” task, commonly used in the perceptual learning domain, bears a striking resemblance to the change detection task used in studies of working memory. In a typical same-different trial, observers are shown two stimuli separated temporally and must decide whether they were the same or different from one another (e.g., Luck & Vogel, 1997; Pashler, 1988). Broadly speaking, this is identical the change detection task from VSTM studies. Same-different tasks have been used to investigate perceptual learning of a variety of visual properties. Ball and Sekuler (1982, 1987) used the same-different task to measure performance on motion discrimination of a field of spatially random dots that moved uniformly along both cardinal and oblique orientations. Similar investigations have also been applied to the discrimination of contrast (Nachmias & Sansbury, 1974), luminance (Lasley & Cohn, 1981), and spatial frequency (Fiorentini & Berardi, 1981, 1980). Unlike the approach of VSTM research, the focus of analysis in studies of perceptual learning is not one of capacity, but one of discriminability. Differences between stimuli are commonly quite small; in this sense they are more similar to studies of VSTM precision, or fidelity of visual memoranda.

With the same-different task we find the compliment to problems identified in the
change detection task above. In studies of perceptual learning, discrimination rather than memory is the theoretical focus. Because of this, behavioral response characteristics are interpreted in terms of discriminability. This often means that task successes and failures are attributed solely to some mixture of discriminability (as estimated by $d'$), response bias, and criterion settings (speed-accuracy trade-offs (Pachella, 1974) are well described by adjustments in observer response criteria, which control whether accuracy or speed is favored). Implicit in this approach is the assumption that visual short-term memory, which is clearly required for the same-different task, is perfect—a property of VSTM known to be untrue. Other related perceptual discrimination tasks also rely on the tacit help of working memory. For example, Matthews, Liu, Geesaman, and Qian (1999) used a protocol in which two single line stimuli were presented sequentially with a temporal gap between. Participants were instructed to indicate whether the orientation of the second line was clockwise or counterclockwise relative to the first. Proper encoding, storage, and retrieval from VSTM are prerequisite to success in this task, yet these concepts are rarely discussed. In this particular example, there is no discussion of memory whatsoever in the article.

Some care has been taken with respect to forgetting, albeit indirectly so. A common analytic approach in perceptual learning studies using both the method of constant stimuli and adaptive methods of stimulus control is to present the observer with stimuli that vary along some dimension of interest (e.g., contrast), measuring their psychometric response function. For a stimulus quality such as contrast, we would expect performance to follow a sigmoid function as stimuli progress from low to high contrast. It is known that stimulus-independent errors lead to biases in the estimation of psychometric functions (Wichmann & Hill, 2001), and it is advised to include a lapse rate parameter, which reduces the peak asymptote of the fit to account for these errors. Memory errors—encoding and retrieval errors, as well as forgetting—are
likely well-accounted for by addressing lapse rate in this way. However, recent find-
ings in the mechanisms of VSTM have begun to suggest the potential for interactions
between the contents of VSTM and the perception of externally driven visual input.
These interactions, if present, would strongly indicate that perceptual discrimination
is subject to VSTM’s influence in ways that extend beyond mere “lapses” in memory
and attention. In fact, the current work will demonstrate that it is the precise re-
membered properties in working memory that can systematically alter the perception
of sensory input. Worse, these interactions are most likely to affect response behavior
for subtle discriminations. Given that investigations of visual phenomena are often
conducted at or near threshold levels by design, it is critical that we understand the
conditions under which VSTM contamination violates our assumptions of “perfect”
memory during the comparison process of perceptual discrimination. The primary
goal of this collection of studies is to begin to identify when and how these interactions
occur, including the mechanisms responsible.

To understand why this phenomenon is relatively new and under-appreciated
it is important to consider the historical landscape of working memory. Broadly
speaking, working memory has been conceived of as a temporary buffer into which
information is selectively transferred for safe-keeping. The most prominent influence
on this “standard model” of working memory is the multiple component model put
forth by A. D. Baddeley and Hitch (1974) which posits that working memory consists
of an auditory and visuospatial buffer—the phonological loop and the visuospatial
sketchpad, respectively. Time has seen the creation of additional subdivisions of these
buffers, such as the “visual cache” and “inner scribe” components of the visuospatial
sketchpad (A. D. Baddeley & Logie, 1999), but the basic structural assumptions have
remained largely unchanged. An added assumption, whether unspoken or not, is that
the buffer protects its contents from being overwritten by subsequent visual input.
This is in stark contrast to sensory buffers whose contents are immediately replaced by any new externally-driven input. Support for the standard model of working memory is extensive with tacit or explicit support of some form in many modern accounts of visual short-term memory (see Vogel et al., 2001; Woodman & Vogel, 2005, for recent examples).

A central tenant to the standard model is the role of the prefrontal cortex (PFC) in the control of working memory. Early evidence for the role of the PFC came from studies of monkey physiology in which sustained activity was found in individual neurons of the PFC during the memory retention interval (Fuster, 1973). Because the PFC has reentrant connections with sensory systems, the motor cortex, and other cortical and subcortical areas, Goldman-Rakic (1987, 1990) was the first to suggest that the temporary workspace buffers of Baddeley and Hitch were the emergent result of the PFC's ability to (1) cross-coordinate different brain areas and (2) maintain information across temporal gaps through sustained activity. A great deal of research on working memory has been inspired by the standard model of working memory (see Courtney, 2004, for a comprehensive review). The result of these efforts suggests that the PFC's role in working memory is functionally organized according to the type of information represented in different cortical areas, with a modular architecture of increasing hierarchical complexity toward more anterior regions of the PFC.

Despite the intuitive appeal of the standard model of working memory, recent discoveries have begun to challenge this conventional approach. It is known, for example, that the perception of external visual input during the memory retention interval can systematically degrade or bias the contents of visual short-term memory (Magnussen & Greenlee, 1992). This phenomenon of “memory masking” is disruptive to recall fidelity proportionate to the similarity between the mask and the memorandum (Magnussen & Greenlee, 1999). For example, in discrimination of spatial frequency,
a Gabor with a spatial frequency difference of 1–1.5 octaves compared to a memo-
risized Gabor will be more disruptive to recall than a Gabor with an identical spatial
frequency (Magnussen, Greenlee, Asplund, & Dyrnes, 1991). In addition, Magnussen
et al. (1991) found that varying the difference in orientation between the memory
sample and masker had no interfering effect when spatial frequency was the dimen-
sion of interest. That is, memory masking only occurs along task-relevant stimulus
channels. Such dependencies provide compelling evidence that memory masking is
the result of interactions between the neural substrates of VSTM and sensory rep-
resentations. These results in and of themselves are not particularly surprising, as
they support the sensible intuition that externally-driven sensations should dominate
our perceptual experience, lest VSTM create hallucinations. Still, it is clear that the
idea of an isolated VSTM buffer sequestered from the contamination of visual input
is rendered untenable.

Further work in VSTM has cast additional, and far more incriminating doubt on
the standard model of working memory. The most telling indication is the prolifera-
tion of dissociable neural subsystems responsible for the maintenance of informa-
tion in working memory. The work by Magnussen and Greenlee demonstrated that
even low-level physical properties such as orientation, spatial frequency, as well as
motion direction and velocity operate in parallel along separate relatively isolated
memory channels. Subsequent work has found similar dissociations between manip-
ulable versus nonmanipulable objects (Mecklinger, Gruenewald, Besson, Magnié, &
Von Cramon, 2002) and faces versus houses (Ranganath, Cohen, Dam, & D’Esposito,
2004). Similar results have been obtained in working memory of tactile, verbal, and
seems reasonable to assume that if the brain can represent it, the brain can also
demonstrate working memory for it” (p.29). This raises a serious concern for the
standard model, wherein memory buffers are said to incorporate individual modules in the PFC. Certainly, such a proliferation of subsystems of memory violates the parsimonious value of the standard model, as the current view seems to suggest that the PFC must redundantly encode increasingly more types of information from all over the brain. Even worse, recall fidelity is known to be extraordinarily precise under many conditions (Regan, 1985; Magnussen & Greenlee, 1992; Vogels & Orban, 1986; Lee & Harris, 1996; Burnham & Clark, 1955). It is unclear how frontal networks, which are involved in the coordination of many simultaneous cognitive processes including all other sensory modalities, might alone encode such detailed information without the aid of specialized sensory areas of the brain. Furthermore, the inefficiency of redundantly encoding sensory representations in a central memory location in the PFC would seem to violate the distributedness that makes the neural connectivity of the brain so efficient in many other aspects.

A competing theoretical account of VSTM known as the “sensory recruitment” hypothesis (also known as the “emergent-property” view of working memory) is gaining support (Awh & Jonides, 2001; D’Esposito, 2007; Postle, 2006). According to this hypothesis, the neural mechanisms of working memory are at least partially subserved by the same sensory areas involved in the encoding of sensory input. This hypothesis does not attempt to dismiss existing evidence (Curtis & D’Esposito, 2003; Pasternak & Greenlee, 2005) implicating the involvement of fronto-parietal networks in visual short-term memory tasks. Loosely speaking, the sensory recruitment hypothesis posits that the role of frontal areas is restricted to controlling the deployment and persistence of attention according to task needs. Working memory then, is the emergent result of attention-driven sustained activity across the entire brain. In this way, working memory is coordinated by the PFC, but represented by whichever specialized brain areas are already involved in encoding the information. This model of working
memory reconciles the aforementioned theoretical challenges of the standard model in part by removing the need for a centralized collection of modules responsible for representing an ever-growing list of memory buffers. It also can account for a growing body of modality-specific working memory signatures found in sensory areas that the standard model cannot. For example, recording single-cell activity in monkeys during working memory tasks has revealed sustained activity in brain areas responsible for the encoding of sensory input (Pasternak & Greenlee, 2005; E. K. Miller, Li, & Desimone, 1993; Chelazzi, Miller, Duncan, & Desimone, 1993). That is, the same cells responsible for encoding a “live” stimulus remain active during the retention interval, and the contents of VSTM can be reliably predicted by leveraging machine learning techniques against the observed activity levels. Recent studies using functional magnetic resonance imaging (fMRI) have found similarly predictive sustained activity in population-level measures of visual areas in human observers, as early as visual areas V1–V3 (Ester, Anderson, Serences, & Awh, 2013; Konstantinou, Bahrami, Rees, & Lavie, 2012; Ester, Serences, & Awh, 2009; Serences, Ester, Vogel, & Awh, 2009; Sligte, Scholte, & Lamme, 2009; Harrison & Tong, 2009; Baumann, Endestad, Magnussen, & W., 2008). Furthermore, work in transcranial magnetic stimulation (TMS) has provided converging evidence. For example, applying magnetic pulses during the memory retention interval to areas V1–V2 (Cattaneo, Bona, & Silvanto, 2012), extrastriate visual cortex (Saad & Silvanto, 2013b), as well as the lateral occipital complex and the posterior parietal cortex (van de Ven & Sack, 2013; van de Ven, Jacobs, & Sack, 2012; Soto, Llewelyn, & Silvanto, 2012) can impair working memory in observers. Further, phosphenes, which reflect the functional properties of stimulated areas, can be induced with TMS, leading to the illusory perception of features from memoranda in VSTM when visual areas are targeted (Silvanto & Cattaneo, 2010).

The sensory recruitment hypothesis of working memory provides a reasonable
and intuitive explanation for the effects of memory masking—if memory and perception share neural resources, we might expect cross-talk under circumstances in which similar stimuli are being represented on shared hardware. A major theoretical component of the current investigation involves answering the reciprocal question to memory masking. Specifically, if external visual input can disrupt VSTM, can the contents of VSTM in turn influence the perception of visual stimuli during the retention interval? Phenomenologically, this effect, if present, must be small, otherwise we would experience perceptual aberrations in every day life stemming from the contents of our memory. However, inside the experimental laboratory highly contrived psychophysical tasks are routinely designed to measure observers’ abilities at near-threshold levels. It is precisely in these environments that VSTM’s influence on perception might be disruptive in a measurable and non-negligible way. Because many common tasks involve the implicit use of VSTM under perceptual uncertainty, and because theoretical explanations are frequently based on the interpretation of responses across fragile distinctions in perception without considering VSTM, a more careful investigation of VSTM-perception interactions is warranted.

The present work will demonstrate that VSTM can, in fact, interfere with the perception of externally-driven stimuli, and will make an attempt at understanding how and when these interactions occur. Taken together, the current studies make a compelling case for including working memory as a necessary and explicit component of theories and models of perceptual learning. The culmination of this work is a novel specification of a model that can explain both forms of VSTM-perception interactions, including how they differentially change under conditions of extended practice.
CHAPTER 2

Motion aftereffect duration is not changed by perceptual learning: Evidence against the representation modification hypothesis

2.1 Introduction

Visual perceptual learning is defined as practice-induced improvement in visual tasks (see Fahle & Poggio, 2002; Fine & Jacobs, 2002, for reviews). It has been documented in orientation discrimination (e.g., Dosher & Lu, 1998), Vernier acuity (e.g., Fahle & Edelman, 1993), visual search (e.g., Ahissar & Hochstein, 1997), texture discrimination (e.g., Karni & Sagi, 1991), face identification (e.g., J. M. Gold, Bennett, & Sekuler, 1999), and motion detection and discrimination (e.g., Ball & Sekuler, 1987; X. Huang, Lu, Tjan, Zhou, & Liu, 2008; Law & Gold, 2008; Z. Liu, 1999; A. A. Petrov & Hayes, 2010; Watanabe, Náñez, & Sasaki, 2001). The learning effects are typically long-lasting and (partially) specific to the particular stimuli used in training (e.g., Ahissar & Hochstein, 1996, 1997; Ball & Sekuler, 1987; R. E. Crist, Kapadia, Westheimer, & Gilbert, 1997; Fahle & Edelman, 1993; Z. Liu, 1999).

The mechanisms of perceptual learning are still poorly understood and are a topic of active research (e.g., Gilbert, Sigman, & Crist, 2001; Lu, Yu, Watanabe, Sagi, &
Levi, 2009). Two prominent hypotheses in the field are *representation modification (RM)* and *selective reweighting (SRW)*. The representation modification hypothesis attributes the behavioral improvement to changes in the early visual representations (e.g., Gilbert et al., 2001; Karni & Sagi, 1991; Schoups, Vogels, Qian, & Orban, 2001). It is advanced on the basis of the stimulus specificity of the learning effect, which is consistent with the stimulus-specific tuning of neurons in the early sensory areas. The RM hypothesis has been implemented in two related models (Schwabe & Obermayer, 2005; Teich & Qian, 2003) that have accounted for some neurophysiological correlates of perceptual learning (Schoups et al., 2001; Yang & Maunsell, 2004). In other sensory modalities, abundant evidence for training-dependent cortical map plasticity has been found in the primary somatosensory and auditory cortices (see, e.g., Buonomano & Merzenich, 1998; Das, 1997, for reviews).

However, an alternative explanation is equally consistent with the stimulus specificity of learning (A. A. Petrov, Dosher, & Lu, 2005). It is possible that the system learns which features of the redundant, multifaceted early representations are most diagnostic for the task at hand and strengthens the read-out connections from the units encoding these features (Dosher & Lu, 1998; Mollon & Danilova, 1996). This selective reweighting hypothesis has been implemented in numerous models that account for a range of behavioral (e.g., Dosher & Lu, 1999; Lu, Liu, & Dosher, 2010; A. A. Petrov et al., 2005; A. A. Petrov, Dosher, & Lu, 2006; Seung & Sompolinsky, 1993; Sotiropoulos, Seitz, & Seriès, 2011; Vaina, Sundareswaran, & Harris, 1995) and neurophysiological (Law & Gold, 2008, 2009) data. The learning effect is stimulus-specific because each read-out connection is anchored in a stimulus-specific unit (A. A. Petrov et al., 2005).

Because each read-out connection is anchored in a sensory unit, the RM and SRW hypotheses cannot be distinguished on the basis of stimulus specificity. This distinction must be made on the basis of task specificity. There is abundant evidence that
perceptual learning is task-specific (e.g., Ahissar & Hochstein, 1993; Ball & Sekuler, 1987; X. Huang et al., 2008; Meinhardt, 2002; Shiu & Pashler, 1992). However, most of these studies involve pairs of tasks that depend on completely unrelated stimulus dimensions. For example, (Shiu & Pashler, 1992) used brightness discrimination and orientation discrimination with line stimuli that varied in both brightness and orientation. Such experimental designs demonstrate the role of attention in learning (Ahissar & Hochstein, 2002), but provide little information about the locus of plasticity. This is because the two tasks depend on stimulus dimensions that are encoded by non-overlapping populations of neurons. The use of the same stimulus set for both tasks does not guarantee that the same sensory representations are engaged in both (A. A. Petrov et al., 2005). When the representations overlap little, both RM and SRW hypotheses predict little transfer across tasks.

Our goal in this article is to test the RM hypothesis. Two conditions are required for a stringent test (A. A. Petrov et al., 2005): First, we need tasks that engage the same (or at least strongly overlapping) sensory representations. Second, the tasks must depend on different (or at most weakly overlapping) read-out connections. These two requirements are hard to satisfy simultaneously. We are aware of only three studies that satisfy them to some degree (R. E. Crist et al., 1997; Fahle, 1997; Saffell & Matthews, 2003). Two experiments demonstrated lack of transfer of learning across orientation discrimination and Vernier discrimination (R. E. Crist et al., 1997; Fahle, 1997). While a good case can be made that Vernier discrimination depends on the orientation-selective neurons in V1 (e.g., Weiss, Edelman, & Fahle, 1993; Wilson, 1986), this remains an assumption and Vernier hyperacuity can also be modeled on different principles (e.g., Cao & Grossberg, 2005). The third experiment (Saffell & Matthews, 2003) demonstrated lack of transfer of learning across
motion direction discrimination and speed discrimination. Convergent evidence indicates that most neurons sensitive to visual motion are tuned conjunctively for both direction and speed (see, e.g., Britten, 2004, for review). Thus, both tasks engage the same population of motion-sensitive neurons and the lack of transfer poses a problem for the RM hypothesis. However, these data cannot rule out one prominent form of representation modification—selective increase of the slope of the tuning curves (Schoups et al., 2001). It is possible that direction-discrimination training sharpens the direction tuning of certain critical neurons without affecting their speed tuning, whereas speed-discrimination training sharpens the latter but not the former. More generally, selective sharpening seems compatible with task-specific learning of any two stimulus dimensions. A more stringent test of the RM hypothesis requires a pair of dissociable tasks based on a single dimension.

The main idea of this article is to use adaptation to probe the early representations of visual motion direction. Adaptation is a valuable tool for studying the visual system (see, e.g., Clifford, 2002; Clifford et al., 2007, for reviews). Here, we test whether practicing a direction-discrimination task affects the strength of the motion aftereffect in the trained direction relative to a control direction.

The motion aftereffect (MAE) is a well-known visual illusion (see Mather, Verstraten, & Anstis, 1998; Mather, Pavan, Campana, & Casco, 2008, for reviews). Prolonged exposure to motion in a given direction causes a subsequent illusory percept of motion in the opposite direction. The aftereffect is stimulus-specific (see Thompson, 1998, for review). For example, it shows both spatial- and temporal-frequency tuning (e.g., Bex, Verstraten, & Mareschal, 1996; Cameron, Baker, & Boulton, 1992; Schofield, Ledgeway, & Hutchinson, 2007). Converging evidence from psychophysics (e.g., Nishida & Sato, 1995; Verstraten, van der Smagt, Frederiksen, & van de Grind,
single-unit recording (e.g., Kohn & Movshon, 2004), brain imaging (e.g., Taylor et al., 2000), transcranial magnetic stimulation (e.g., Théoret, Kobayashi, Ganis, Di Capua, & Pascual-Leone, 2002), and other methods (see Mather et al., 2008, for review) indicates that MAE is not a monolithic phenomenon but “an amalgam of neural adaptation at several visual cortical sites” (Mather et al., 2008, p. 481). Our present experiments use two types of aftereffects: static (sMAE) and dynamic (dMAE). The adapting stimulus—moving filtered-noise texture—is the same for both types. They differ in the test stimulus during the post-adaptation phase: a static texture frame for sMAE versus dynamic visual noise for dMAE. These two types of aftereffects have different properties and seem to arise at different levels of the motion-processing pathway (e.g., Nishida & Ashida, 2000; Nishida & Sato, 1995; Verstraten et al., 1999; see Culham et al., 1998; Mather et al., 2008, for reviews). In particular, areas V1, V2, and V3 are implicated in static MAE (e.g., Maruya, Watanabe, & Watanabe, 2008; Taylor et al., 2000) and MT is implicated in both static and dynamic MAE (e.g., Kohn & Movshon, 2004; Théoret et al., 2002; Tootell et al., 1995). We tested both types to probe for representation modification across these areas.

Various theoretical explanations of the MAE have been proposed (see Mather & Harris, 1998; Mather et al., 2008; Vidnyánszky, Blaser, & Paphthomas, 2002, for reviews). More generally, there is an extensive literature on the mechanisms of motion adaptation and related phenomena such as the direction aftereffect (see, e.g., Clifford, 2002; Kohn, 2007, for reviews). Some of these topics are discussed briefly in Section 2.5 below. While the details differ, all models agree that motion adaptation impacts the neuronal populations involved in processing and representing visual motion. In the closely related domain of orientation processing, the influential model of Teich and Qian Teich and Qian (2003) proposes a common mechanism for both adaptation and perceptual learning—sharpening of the orientation tuning curves.
by means of changing the lateral connections in a recurrent network. According to this proposal, “adaptation in behaving subjects may be viewed as a short-term form of learning” (Abstract Teich & Qian, 2003).

In sum, motion direction discrimination and MAE seem to rely on strongly overlapping sensory representations and are based on a single stimulus dimension. The representation modification hypothesis predicts that extensive discrimination practice near a particular direction will change the neuronal representations for this direction. We further hypothesize that this representational change will cause a detectable change in the strength of the MAE in the trained direction relative to an untrained control.

Turning to the selective reweighting hypothesis, it is important to choose the two experimental tasks that rely on different read-out connections. Only then do the SRW and RM predictions diverge. The MAE task is very similar to detection.\(^1\) We use reported MAE duration as an index of strength (Pantle, 1998). The observers pressed a key when the illusory motion was no longer detectable. There is mounting psychophysical (Hol & Treue, 2001; Jazayeri & Movshon, 2007; Phinney, Bowd, & Patterson, 1997; Regan & Beverley, 1985), neurophysiological (e.g., Purushothaman & Bradley, 2005; Raiguel, Vogels, Mysore, & Orban, 2006), and computational (e.g., Jazayeri & Movshon, 2006; A. A. Petrov et al., 2005; Seung & Sompolinsky, 1993) evidence that detection and fine\(^2\) discrimination rely on different read-outs. Thus, we use fine discrimination training in our experiments. The observers practiced to discriminate small differences in motion direction from trial to trial. We hypothesize

---

\(^1\)There are some subtle but important differences between detecting real and illusory motion. They are discussed in Section 2.5.

\(^2\)As opposed to coarse discrimination of motion in opposite directions, which is similar to detection (A. A. Petrov & Hayes, 2010).
that the fine-discrimination read-out differs from the MAE read-out. Therefore, if perceptual learning occurs via selective reweighting of the read-out connections from unchanging representations, fine discrimination practice should have no effect on the MAE.

The RM and SRW hypotheses thus make opposite predictions about the interaction of fine-discrimination practice and MAE. These predictions were tested in two experiments.

### 2.2 Experiment 1

The first experiment uses static and dynamic MAE to probe for practice-induced changes at different levels of the visual hierarchy. We use a moving filtered-noise texture for the adapting stimulus in both cases. The test stimulus is a static texture frame on sMAE trials and multiple frames of dynamic visual noise on dMAE trials. The experimental schedule begins with a MAE pretest session, followed by 4 discrimination sessions practicing one particular direction, followed by a MAE posttest session. A final discrimination session with the orthogonal direction verifies the specificity of the learning effect on the fine discrimination task. This experimental design produces the following three binary factors: Session (pretest vs. posttest) × Direction (trained vs. control) × Type (static vs. dynamic). All three factors are crossed within each participant. The Direction factor is counterbalanced between participants.

The question of main interest is whether there is a statistically significant interaction between the Session and Direction factors. The RM hypothesis predicts such interaction whereas the SRW hypothesis predicts no interaction. We use analysis of variance (ANOVA) to test for statistical significance. A technical difficulty arises at this point because the SRW prediction amounts to asserting the null hypothesis in
the ANOVA (Cohen, 1992; Keppel & Wickens, 2004). To make such null result meaningful, the test must have sufficient statistical power to detect the effect if, in fact, the RM hypothesis is correct and practice does induce direction-specific changes in the MAE. This concern is addressed in four ways: First, the adapting stimuli on the MAE sessions are as similar as possible to the training stimuli on the discrimination sessions. Second (and most important), the statistical test is performed individually for each participant. This leverages the power across multiple participants. The probability to fail to detect a true effect (“Type II error”) decreases exponentially with the number of independent tests. Thus, even though the individual tests may have modest power, the power of the combined test increases dramatically. Third, we use a priori planned contrasts to maximize the power of each individual test to detect the specific kind of interaction predicted by the RM hypothesis—see Section 2.2.1 for details. Fourth, we calculate explicitly the smallest effect size that our leveraged test is expected to detect at the conventional significance level ($p < .05$).

2.2.1 Method

Observers

The participants were 11 students at the Ohio State University with normal or corrected-to-normal vision. They were naïve to the purposes of the experiment and were paid $6 per hour plus a bonus contingent on their accuracy.

Stimuli and Apparatus

The stimuli were filtered-noise textures moving coherently at a constant speed of 12 deg/s behind a circular aperture (Figure 2.1). The filter had a Gaussian cross-section along the frequency axis in the Fourier domain and was radially symmetric at all directions. The spectral power peaked at 3 cpd and the bandwidth was 4
Figure 2.1: The stimuli were filtered-noise textures moving behind a circular aperture. On discrimination trials, the texture moved in one of four possible directions (depicted by arrows) relative to an implicit reference direction (dotted line). The adapting stimuli on motion-aftereffect trials were the same, except that they moved for 10 s rather than 397 ms. [Note: The fixation dot at the center and the angles between the directions are exaggerated for visibility.]
octaves (full width at half height). A new texture was generated on every trial by applying the same filter to a fresh sample of independent, identically distributed Gaussian noise. Consecutive, overlapping frames were cut out from a larger texture patch and presented at 96 Hz on a 21” NEC AccuSync 120 color CRT using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) in Matlab (The MathWorks, 2004). The direction of motion was manipulated by rotating the individual frames via OpenGL calls to the ATI Radeon HD2600 Pro graphics card. The aperture was implemented by a circular mask that was fully transparent at radius $R = 9.5$ cpd, fully opaque at $R = 10.5$ cpd, and ramped down linearly in between. A small red fixation dot (14.2 min in diameter) marked the stimulus center at all times. The monitor gamma function was estimated via a psychophysical matching procedure (cf., Colombo & Derrington, 2001) and was verified with a Minolta 1° luminance meter. A software lookup table defined 255 evenly spaced luminance levels between $L_{\text{min}} = 1.5 \text{ cd/m}^2$ and $L_{\text{max}} = 32.5 \text{ cd/m}^2$. The display was the only light source in the room and was viewed binocularly with the natural pupil from a chin rest located $\approx 93$ cm away. At that distance, 1 degree of visual angle subtended $\approx 42$ pixels ($1024 \times 768$ resolution).

**Discrimination Task and Procedure**

The fine direction discrimination task was defined with respect to a *reference direction* $\theta$ that was set implicitly for each block. The actual motion direction took four possible values: $(\theta - 3.5)$, $(\theta - 2)$, $(\theta + 2)$, and $(\theta + 3.5)$ degrees from vertical. Each block presented 120 stimuli of each kind. The instructions designated the first two directions as “counterclockwise” and the other two as “clockwise” and the observers made a binary discrimination choice by pressing one of two keys on the computer keyboard.
Each trial began with a brief beep. The texture appeared 500 ms later, moved for 397 ms, and disappeared. The beep onset always preceded the texture onset by exactly 500 ms, and thus could serve as a reliable attentional cue. A bonus system helped to motivate the observers and provided feedback. The reward for each correct response was a bonus point. The penalty for each error was the loss of a bonus point, an unpleasant beep, and the addition of 250 ms to the 800-ms intertrial interval. The cumulative bonus was displayed prominently at all times and was converted to dollars and cents after the last session.

**MAE Task and Procedure**

Each MAE trial consisted of three phases: adaptation, MAE test, and reset. The adaptation duration was 10 s, the MAE duration was the dependent variable, and the reset duration was adjusted to make the trial total exactly 30 s. The adapting stimuli were the same as in the discrimination task, except that their duration was \( \approx 25 \) times longer. The adaptation duration (10 s) was determined in a pilot experiment and was designed to avoid ceiling and floor effects. The observers were instructed to fixate the stationary red dot in the middle of the display throughout the adaptation period. As an incentive for doing so, a simple attentional task provided an opportunity to score bonus points. The motion direction alternated between \((\theta - 3)\) and \((\theta + 3)\) at random intervals and the participants pressed a key whenever they detected a change. A variable number of such changes occurred on a trial and the bonus points scored for detecting them were visible throughout the adaptation period.

The test phase presented stimuli of two kinds: static and dynamic. On static (sMAE) trials, the test stimulus was a stationary frame of the same texture (Figure 2.1). It appeared to move in direction \(-\theta\) due to the aftereffect. On dynamic (dMAE) trials, the test stimulus was dynamic visual noise. That is, each frame was a
mosaic of small square tiles of size $4 \times 4$ pixels. (Individual-pixel noise tended to blend into near-uniform grey.) The intensity of each tile was drawn independently from a Gaussian distribution and resampled 96 times per second. The semitransparent circular mask was applied throughout. This stimulus contained no globally coherent motion energy and yet it appeared to move due to the aftereffect. The observers were asked to press a key when the apparent motion stopped. The re-appearance of the bonus (which was not displayed during the test phase) signaled that the response had been registered.

The third and final phase was identical for both sMAE and dMAE trials. Dynamic noise was presented for the reminder of the duration in an effort to reset the motion processing system and minimize adaptation carry-overs across trials (von Grunau, 2002). It also removed the incentive to report short MAE durations because all trials lasted for 30 s regardless of the participant’s response. A brief alert beep signaled the beginning of the next trial.

**Experimental Design and Presentation Schedule**

The experiment involved two reference directions: $\theta = -50$ and $+40$ degrees from vertical. We avoided the diagonal directions to discourage verbalizable decision strategies. For 6 participants, the trained direction was $-50$ and the control direction was $+40$; these values were reversed for the other 5 participants. Each participant completed 2 MAE sessions (on days 1 and 6) and 5 discrimination sessions (on days 2–5 and 7). The MAE sessions consisted of 84 trials, whereas the discrimination sessions consisted of 960 trials. Each MAE session was divided into 7 blocks and the observers were encouraged to rest between the blocks. The MAE duration was

---

3 On a few (<2%) of the trials, no response was made before the deadline. These cases were attributed to lapses of attention and omitted from the analysis. Many of these lapses came from one particular observer.
measured repeatedly for each combination of reference direction (trained vs. control) and type (static vs. dynamic). Each block presented 3 replications of each combination, in random order. The discrimination sessions were divided into 2 blocks, which were further subdivided into “miniblocks.” The reference direction \( \theta \) was the same throughout the 4 practice sessions—this is what made it the trained direction for this individual. Each miniblock presented a counterbalanced, randomized sequence of trials with small clockwise and counterclockwise deviations around \( \theta \) (see Section 2.2.1).

After the MAE posttest on day 6, there was one final discrimination session. The first miniblock (120 trials) of this session revisited the trained reference direction and transitioned back to the discrimination task. Then \( \theta \) was switched to the orthogonal direction to assess the specificity of the discrimination learning effect.

**Power Analysis**

The question of main interest is whether the MAE duration differs significantly between the trained and control directions at posttest. The conventional way to answer this question is to test the Session \( \times \) Direction interaction in a within-subject ANOVA. However, the statistical power of this test would be relatively low. We needed to maximize the power to give the representation modification hypothesis a fair hearing. We conducted a detailed power analysis to make a null effect interpretable. Power calculations required an explicitly specified alternative hypothesis (Cohen, 1992; Keppel & Wickens, 2004). This gave a distinctly Bayesian flavor to our frequentist analysis. We examined a family of such hypotheses defined by an effect-size parameter \( c \).

The crossing of the Direction and Session factors produce 4 conditions: “trained at pretest,” “control at pretest,” “trained at posttest,” and “control at posttest.” The first two are not expected to differ because the reference directions are (essentially)
Table 2.1: Top section: Motion-aftereffect (MAE) duration patterns predicted by the selective reweighting (SRW) and representation modification (RM) hypotheses of perceptual learning. \( M \) denotes the MAE duration at pretest and \( c \) is an effect-size parameter. Middle and bottom sections: Coefficients of the planned linear contrasts for the two types of analysis of variance (ANOVA) discussed in Section 2.2.1.

*Note: The predictions for the main effect of Session assume (incorrectly) that the response criterion does not change from pretest to posttest.

symmetrical and counterbalanced between participants. Let \( M \) denotes the MAE duration for both pretest conditions as shown in the top section of Table 2.1. Let us temporarily\(^4\) assume that the MAE in the control direction does not change between pretest and posttest. Thus, the control MAE duration at posttest is also \( M \).

The trained direction at posttest is the critical condition. The RM hypothesis predicts a change there, whereas SRW predicts the same MAE across the board (Table 2.1). It is convenient to express the change in relative units by introducing the multiplicative parameter \( c \). It can be positive or negative. For example, \( c = -0.10 \) represents the hypothesis that perceptual learning causes a 10% drop in MAE duration.

\(^4\)The final analysis does not depend on this assumption; it is introduced here for expository convenience only.
A conventional ANOVA would partition the variance in our data into two main effects and one interaction effect. Each of them has 1 degree of freedom in our design and thus can be expressed as a planned contrast (Keppel & Wickens, 2004). The corresponding coefficients are listed in Table 2.1. The rightmost column shows the predictions of the RM hypothesis. The predicted MAE change is dispersed across all three tests, which dilutes the statistical power of any one of them. To focus the power, we replaced tests 2 and 3 by the simple-effect tests labeled 4 and 5 in Table 2.1. This concentrated the predicted change into a single test—the one that compares the MAE durations for the trained and control directions at posttest. The coefficients for this critical test are listed in the last row in the table. Note that contrasts 1, 4, and 5 form a mutually orthogonal set. Thus, we can use the same error term and the same significance level as the conventional partitioning (Keppel & Wickens, 2004). The Type factor (static vs. dynamic) is orthogonal to all these contrasts. Although not shown in Table 2.1, it was included in the analysis and the associated variance was subtracted from the error term. We adopted significance level $\alpha = .05$ for all power calculations.

To increase the power still further, a separate ANOVA was performed for each individual participant. These analyses capitalized on the fact that each cell of our experimental design contained 21 replications per observer. A group-level ANOVA would use the cell means only and thus miss the valuable information contained in these multiple replications. The outcomes of the individual tests were combined using the Bernoulli formula for repeated independent trials (Feller, 1957).
2.2.2 Results and Discussion

Discrimination Data

Two discriminability values (d’, Macmillan & Creelman, 2005) were calculated for each observer in each block: for the easy (θ ± 3.5) and difficult (θ ± 2) pair of stimuli. Figure 2.2 plots the group-averaged learning curves. The classic pattern was clearly replicated—the performance improved with practice but the improvement was partially specific to the trained direction, hence the drop after the switch to the orthogonal (control) direction on the last session.

The response times (RTs) also improved with practice. The RT distributions were fitted by a model describing the process of accumulating sensory evidence over time and making a binary decision. We used the diffusion model (Ratcliff, 1978; Ratcliff...
This analysis was based entirely on the discrimination data and is published separately (A. A. Petrov, Van Horn, & Ratcliff, 2011, summarized in Section 2.4 below). The decision-making aspects of the discrimination task are not relevant for the interpretation of the MAE data. Figure 2.2 contains the information needed for our present purposes. The training manipulation was successful—by the time of the MAE posttest (depicted by the shaded area), the $d'$ for the trained direction seemed to increase by approximately 50% relative to its initial level.

**MAE Data**

Figure 2.3 plots the mean MAE durations for the group average (large panel) and for the 11 individual observers (small panels). As expected, there were no significant differences between the two reference directions at pretest. The simple effect of Direction at pretest (contrast 4 in Table 2.1) was not significant in the group-level data ($F(1,10) < 1$, n.s.). Nor was it significant in 10 of the 11 individual ANOVAs. It did reach significance ($F(1,160) = 4.21, p = .042$) for 1 observer, but this appears to be a Type I error. Recall that the two directions ($\theta = -40$ and $+50$) were symmetrical with respect to the vertical and there is no reason to expect MAE differences prior to the discrimination training.

The main question is whether this symmetry was broken after training with one direction but not the other. The data in Figure 2.3 show no evidence for any significant asymmetry at posttest. The two directions (plotted with o’s and ×’s, respectively) continued to elicit very similar MAEs. The simple effect of Direction at posttest (contrast 5 in Table 2.1) was not significant in the group-level analysis ($F(1,10) < 1$, n.s.). More importantly, none of the 11 individual tests revealed any significant differences either. The individual $F$ values for this critical test are printed above their corresponding panels in Figure 2.3.
The average statistical power of the individual tests was estimated as follows. The mean square error (MSE) of each individual ANOVA was divided by the (squared) grand mean MAE for the corresponding observer. The median of these normalized error terms was .10 in our sample (mean = .13, SD = .10). The effect-size parameter $c$ in Table 2.1 is expressed in the same normalized units (Section 2.2.1). We calculated the power for a range of effect sizes on the basis of the median normalized MSE using standard formulas (Keppel & Wickens, 2004). For the critical test (contrast 5 in Table 2.1), we obtained power estimates $P = \{.050, .109, .304, .586, .827\}$ for effect sizes $c = \{0, \pm .05, \pm .10, \pm .15, \pm .20\}$, respectively.

This is the main result of the present article, so let us examine it more closely. Consider the hypothesis that the MAE did, in fact, change by 10% ($c = .10$). According to the above estimate, there is approximately 30% chance to detect a change of this size in the data from one (median) observer. In other words, the majority of the tests (70%) will fail to detect it (Type II error). Clearly, a negative outcome on one individual test does not warrant any strong conclusions. (The null result of the group-level analysis is similarly inconclusive.) However, the probability of observing 11 such failures on 11 independent tests is only $p = .0185$. Thus it is very unlikely that the collective pattern in Figure 2.3 could occur if the hypothesis $c = .10$ were true. The leveraged test has enough statistical power to reject this hypothesis at the 5% significance level. Bigger changes ($c \geq .15$) can be rejected strongly ($p < .0001$). Very small MAE changes ($c = .05$) are compatible with these data ($p = .28$).

Consider now the selective-reweighting prediction that MAE did not change at all ($c = 0$). The probability to incorrectly reject a true null hypothesis (Type I error) is $\alpha = .05$ for each individual test. The probability of 0 rejections on 11 tests is $(1 - \alpha)^{11} = .569$. Thus, our data pattern is 30 times more likely to occur under
Figure 2.3: Mean motion aftereffect (MAE) durations in Experiment 1. The large panel plots the group average. The error bars are 90% within-subject confidence intervals. The pre and post labels on the horizontal axis denote the pre- and posttest MAE sessions, respectively. The small panels plot the mean MAEs for each individual observer (using the same graphical layout). The error bars are 90% confidence intervals estimated from 21 replications per condition. The $F$ statistic above each panel tests whether the trained direction (solid line) differs from the control direction (dashed line) at posttest. None of the tests is significant (critical $F(1, 160) = 3.90$).

selective reweighting than under representation modification involving 10% MAE change.

The main effect of Session (contrast 1 in Table 2.1) was significant for 8 individuals. The MAE duration at posttest compared to pretest increased for 4 of them and decreased for the other 4. Because of the cancelation of these inconsistent shifts, the effect did not reach significance in the group-level ANOVA ($F(1, 10) = 0.33$). We attribute these individual differences to the subjective choice of the response criterion. Recall that the task was to press a key when the apparent motion stopped. It seems
that the observers had adopted different criteria and changed them across sessions in idiosyncratic ways. The lack of significant Session × Direction interaction (as tested by contrasts 4 and 5 above) indicates that these criteria were applied consistently for both directions.

The Type factor had a significant main effect for 9 individuals. Moreover, it was the one effect that reached significance at the group level \( F(1, 10) = 18.2, p < .002 \). The MAE lasted significantly longer on dynamic trials than on static ones (Figure 2.3). We attribute this to the random fluctuations in the dynamic-noise stimuli. The Session × Type interaction was significant for 6 observers but not for the group \( F(1, 10) = 2.01, \text{n.s.} \). The general pattern was that static and dynamic MAE differed more at pretest than at posttest. One possible interpretation of this interaction is that the response criteria became better calibrated as observers gained experience with the MAE task. The Direction × Type interaction was not significant for any individual participant, again suggesting consistent response criteria for both directions. The 3-way interaction (S × D × T) was significant for only 1 of the 11 observers \( F(1, 160) = 4.21, p = .042 \), probably a Type I error.

In conclusion, the data from Experiment 1 suggest that the MAE duration did not change (or changed very little) with discrimination practice. This challenges the representation modification hypothesis. There is, however, one important theoretical possibility that was not addressed in this experiment. The test stimulus in MAE appears to move in the direction that is opposite to the direction of the adapting stimulus. Given that the trained direction was used only for the adapting stimulus in Experiment 1, the illusory motion was never aligned with the trained direction. It was opposite to it. Previous research (e.g., Ball & Sekuler, 1987) indicates that the improvement in motion direction discrimination does not transfer to the opposite direction. This opens the possibility that representation modification did occur but
was not detected because Experiment 1 did not test the MAE in the relevant direction. We conducted a second experiment to evaluate this possibility.

2.3 Experiment 2

This experiment was identical to Experiment 1 except that the MAE duration was measured in 4 directions: $\theta = -50, +40, +130,$ and $-140$ relative to vertical. Also, all dynamic tests were dropped to make room for the static tests in the new directions.

2.3.1 Method

Observers

Sixteen new participants were recruited from the same population and were paid the same hourly rate and bonus as in Experiment 1.

Stimuli, Task, and Procedure

The stimuli and apparatus were the same as in Experiment 1. The fine discrimination task and procedure were also the same. The MAE sessions contained static trials only—21 replications in each of the 4 directions. There were no dynamic trials and thus the MAE session length (and hence statistical power) was the same as in Experiment 1.

Experimental Design and Presentation Schedule

Half of the participants trained with $\theta = -50$ and the other half with $+40$. There was a switch to the orthogonal upward direction on the last session, as in Experiment 1. The two downward directions (+130 and $-140$) were never used for discrimination. They were only used for the adapting stimuli on MAE trials.
Power Analysis

Because the Direction factor has 4 levels (and hence 3 degrees of freedom), it cannot be formulated as a single linear contrast. A set of 3 such contrasts would be necessary. This negates the advantages of the alternative analysis that was used in Experiment 1. Here we used the conventional partitioning of the variance into main effects and interactions. An ANOVA was performed for each individual, the power was calculated for a range of effect sizes using standard formulas (Keppel & Wickens, 2004), and the outcomes of the individual tests were combined as in Experiment 1.

2.3.2 Results and Discussion

Discrimination Data

The $d'$ learning curves for the discrimination task are plotted in Figure 2.4. It closely replicates the results of Experiment 1 (Figure 2.2). The $d'$ for the trained direction (solid symbols) improved with practice but the improvement transferred little to the control direction (open symbols).

MAE Data

Figure 2.5 shows the mean aftereffect durations for the group average (large panel) and for the 16 individual observers (small panels). The results of Experiment 1 were replicated. Most importantly, the Direction $\times$ Session interaction was not statistically significant for 15 of the 16 observers. The individual $F$ values for this critical test are printed above the corresponding panels in Figure 2.5. One observer did show a significant interaction ($F(3, 45) = 3.22, p < .025$) but their data (lower right-hand corner in Fig. 2.5) suggest that it was due to an anomalously low MAE for the control direction at pretest—an apparent Type I error. The group-level test did not reach significance ($F(3, 45) = 0.84$).
Figure 2.4: Learning curves for the discrimination task. Group average of the 16 participants in Experiment 2. The error bars are 90% within-subject confidence intervals. The shaded areas mark the two motion aftereffect sessions.

The normalized mean square error of each individual ANOVA was calculated as in Experiment 1. The median of the new sample was MSE = .11 (mean = .12, SD = .06), replicating our earlier results. On the basis of this error term, we calculated the power to detect the Direction × Session interaction in the data from one (median) observer. The power estimates were $P = \{.050, .083, .202, .387, .605\}$ for effect sizes $c = \{0, \pm .05, \pm .10, \pm .15, \pm .20\}$, respectively. The probability of obtaining exactly 1 significant outcome on $n$ independent tests was calculated by the formula $p = nP(1-P)^{n-1}$ (Feller, 1957). The results were $p \leq \{.37, .36, .11, .004, .00001\}$ for $n = 16$ and the effect sizes listed above.

Experiment 2 thus had slightly less statistical power than Experiment 1. The hypothesis that MAE changes by 15% or more can be rejected strongly ($p < .004$) but the hypothesis $c = .10$ cannot ($p = .11$). Still, the data pattern is 3.4 times more
likely to occur under the null hypothesis \((c = 0)\) than under a 10% change. This reinforces the main conclusion of Experiment 1: The MAE duration did not change (or changed very little) with discrimination practice. The new result is that this was demonstrated for adapting stimuli moving not only in the trained direction but in the opposite direction as well.

The lack of main effect of Session at the group level was replicated too \((F(1, 15) = 1.55, p = .23)\). Averaged across all 4 directions, the MAE duration at posttest compared to pretest increased significantly for 2 individuals and decreased significantly for 9 others. This replicated the finding that different observers adopt different response criteria and change them in idiosyncratic ways. Finally, the Direction factor did not have a significant main effect for the group \((F(3, 45) = 0.16)\) but did for 8 individuals, mostly due to differences between the upward and downward directions.

### 2.4 Decision-making Aspects

The present article focused on the MAE, whereas a companion article (A. A. Petrov et al., 2011) focused on the discrimination sessions. Because of their homogeneity, the data from Experiments 1 and 2 were analyzed together. A *learning index* (Fine & Jacobs, 2002) was calculated to quantify the increase of \(d'\) relative to its initial level: \(LI = (d'_{8} - d'_{1})/d'_{1}\), where the subscripts denote block numbers (cf. Fig. 2.2 and 2.4). It was\(^5\) \(LI = .55 \pm .08\) for the average learning curve in the combined sample. A *specificity index* (Ahissar & Hochstein, 1997) quantified the disruption caused by the switch to the control direction: \(SI = (d'_{8} - d'_{9})/(d'_{8} - d'_{1})\). It was \(SI = .60 \pm .10\) (A. A. Petrov et al., 2011). In words, the average \(d'\) improved by approximately 55%

\(^5\)±80% bootstrap confidence interval, see A. A. Petrov et al. (2011) for details.
after 4 days of discrimination practice but 60% of this improvement were specific to the trained direction.

As the response times (RTs) also improved with practice, the $d'$ indices underestimate the true learning effect. The diffusion model (DM) is a valuable tool for analyzing joint accuracy and RT data ((Ratcliff, 1978); see (Ratcliff & Mckoon, 2008), for review). Just as signal detection theory (Macmillan & Creelman, 2005) converts hits and false alarms into theoretically motivated estimates of discriminability and bias, the DM converts hits, false alarms, and RT distribution statistics into estimated parameters of various processing components. The DM accounted very well for the detailed discrimination data and re-expressed the regularities in these data in terms of 7 parameters per observer per block (A. A. Petrov et al., 2011). The drift rate parameter is especially important for our present purposes because it quantifies the sensory evidence feeding into the decision process. By the end of training, the average drift rates were two times greater than their initial level ($LI = .99 \pm .23$). This accounted for both the observed $d'$ increase (Fig. 2.2 and 2.4) and part of the RT decrease. The learning effect was partially specific to the trained direction ($SI = .68 \pm .09$).

### 2.5 General Discussion

The present study tested a prediction of the representation modification hypothesis of perceptual learning. According to this hypothesis, practicing a direction-discrimination task induces stimulus-specific changes in the cortical representation of visual motion. In turn, these changes predict stimulus-specific changes in the motion

---

6The DM analysis suggested a second learning mechanism that improved the timing of the decision-process onset relative to the stimulus onset. It is not discussed here because it does not affect the interpretation of the MAE data.
aftereffect. In two experiments, fine-discrimination practice caused large stimulus-specific improvements in $d'$ but no significant stimulus-specific changes in MAE duration. Power analysis indicated that the data were approximately 100 times\textsuperscript{7} more likely given the hypothesis of no MAE change than the hypothesis of a 10% relative change.

Section 2.1 cited convergent evidence that the MAE arises in the cortical areas that process and represent visual motion. Recall that areas V1, V2, and V3 are implicated in static MAE (e.g., Maruya et al., 2008; Taylor et al., 2000) and MT is implicated in both static and dynamic MAE (e.g., Kohn & Movshon, 2004; Théoret et al., 2002; Tootell et al., 1995). The absence of significant stimulus-specific changes in either static or dynamic MAE suggests that little or no change occurred in the cortical representations of visual motion up to and including area MT. This challenges the RM hypothesis.

This inference depends on two premises: 
1) that our tasks engaged the same population of motion-sensitive neurons and 
2) that modifying these internal representations would cause detectable changes in the MAE. We consider (and reject) two alternative interpretations of our results that negate these premises in turn.

The first alternative interpretation points out, correctly, that overlap at the level of cortical areas may not necessarily imply overlap in the specific neurons that determine the behavioral outcome in our two experimental tasks. There is mounting evidence (e.g., Hol & Treue, 2001; Jazayeri & Movshon, 2007; Purushothaman & Bradley, 2005) that the neurons that are most diagnostic for fine discrimination are those whose direction preferences are shifted away from the discrimination boundary. The tuning curve of these \textit{flanking neurons} has the steepest slope at the boundary.

\textsuperscript{7}These odds (or \textit{Bayes factors}, Kass & Raftery, 1995) combine multiplicatively across data sets. They were 30 and 3.4 in Experiments 1 and 2, respectively.
Figure 2.5: Mean motion aftereffect (MAE) durations in Experiment 2. The large panel plots the group average. The error bars are 90% within-subject confidence intervals. The pre and post labels on the horizontal axis denote the pre- and posttest MAE sessions, respectively. The small panels plot the mean MAEs for each individual observer (using the same graphical layout). The error bars are 90% confidence intervals estimated from 21 replications per condition. The $F$ statistic above each panel tests the Direction $\times$ Session interaction. Only 1 of the 16 tests is significant (critical $F(3,160) = 2.66$).
The psychophysical demonstrations of this (Clifford, Wyatt, Arnold, Smith, & Wenderoth, 2001; Hol & Treue, 2001; Phinney et al., 1997; Regan & Beverley, 1985) are particularly relevant here because they involve adaptation (see Clifford, 2002, for review). Fine-discrimination thresholds increase by up to 60% for motion directions 20–30 deg on either side of the adaptor (Hol & Treue, 2001; Phinney et al., 1997) and for orientation angles 10–15 deg on either side (Clifford et al., 2001; Regan & Beverley, 1985). The discrimination threshold at the adapted direction or orientation decreases by up to 20% (Clifford et al., 2001; Phinney et al., 1997; Regan & Beverley, 1985). The latter improvement has been attributed to disinhibition of the flanker neurons when the neurons tuned for the boundary are suppressed by adaptation (Phinney et al., 1997). Other explanations are also possible (e.g., Clifford et al., 2001; Seriès, Stocker, & Simoncelli, 2009). In one way or another, the effect seems to arise from the lateral interactions among the tuned mechanisms.

The detection thresholds, on the other hand, are most elevated at the adapted direction (e.g., Hol & Treue, 2001) or orientation (e.g., Regan & Beverley, 1985). This unimodal pattern is qualitatively different from the Mexican-hat pattern of change in the fine-discrimination thresholds. This is why Section 2.1 cited these studies in support of the claim that the two tasks in our experiments rely on different read-out connections.

In light of this evidence, why should we expect an interaction between fine-discrimination training and the MAE under the representation modification hypothesis? For concreteness, consider the preeminent RM proposal: Perceptual learning sharpens the tuning curves of the flanker neurons on either side of the trained reference direction (Schwabe & Obermayer, 2005; Schoups et al., 2001; Teich & Qian, 2003; Yang & Maunsell, 2004). The neurons tuned for the reference direction itself
need not be modified. Given that the adapting stimuli on our MAE trials were very close to the reference direction, why should we expect a change in MAE triggered by modifications in the flanker neurons 20-30 deg away?

The answer is threefold: First, there are well documented lateral interactions among the neurons in the early sensory areas (e.g., Gilbert, 1992). The horizontal collaterals in V1, for example, contribute to orientation sensitivity (e.g., Nelson, Toth, Sheth, & Sur, 1994; Somers, Nelson, & Sur, 1995), contrast gain control (e.g., Heeger, 1992), and context integration (e.g., Gilbert, Ito, Kapadia, & Westheimer, 2000). As discussed above, release from lateral inhibition is one likely mechanism of the improvement of the discrimination threshold at the adapted direction (Phinney et al., 1997). The RM models (Schwabe & Obermayer, 2005; Teich & Qian, 2003) explicitly attribute perceptual learning to plasticity in the recurrent lateral connections. Second, adaptation affects the flanker neurons too—a phenomenon known as flank adaptation (e.g., Kohn & Movshon, 2004); see (Kohn, 2007), for review). The effects of adaptation are not confined to the neurons that respond most strongly to the adaptor. Thus, if practice had modified significantly the response properties of the flanker neurons on both sides of the reference direction in our experiments, it is not unreasonable to expect a significant modification in the strength and/or time course of adaptation along this direction as well.

Finally, the detection of illusory motion is characterized by additional smoothing and integration compared to real motion. There is strong evidence that “the perceived global motion direction during motion aftereffects results from local vector averaging

---

8Indeed, no such modifications have been reported in single-cell recordings to date, except that (Ghose, Yang, & Maunsell, 2002) found slightly fewer V1 neurons whose optimal orientation was near the trained orientation. This resulted in a small but significant decrease in the V1 population response to the trained orientation at the trained location compared to a location in the opposite, untrained hemifield.
of the co-localized motion-direction signals induced by adaptation” (Abstract, emphasis added Vidnyánszky et al., 2002). A straightforward demonstration of this is that adaptation to bivectorial transparent motion normally leads to univectorial non-transparent MAE in the direction opposite to the vector sum of the adapting directions (e.g., Verstraten, Fredericksen, & van de Grind, 1994). This phenomenon reveals the inadequacy of the classic opponent-process (or ratio) model (Sutherland, 1961), which attributes the MAE to an adaptation-induced imbalance between the responses of the units tuned to the adapted direction and to its opposite. This problem motivated the development of the distribution shift model (Mather, 1980) and its successor, the automatic gain-control model (Grunewald, 1996; Grunewald & Lankheet, 1996; van de Grind, Lankheet, & Tao, 2003; van de Grind, van der Smagt, & Verstraten, 2004). The details of these models are beyond the scope of this article (see Mather & Harris, 1998; van de Grind et al., 2003, for reviews). It suffices for our purposes that they all agree that the MAE involves multiple channels and that the mutual (dis)inhibition among them plays a major role. In the case of bivectorial adaptation, the inhibitory interactions fuse the aftereffect into one single direction. Now, the RM hypothesis postulates changes in the channels flanking the discrimination boundary. Given the stimulus-specificity of perceptual learning, the response properties of these flanker channels should be different from those of the surrounding channels. This two-prong modification pattern induced by training is qualitatively similar to the two-prong suppression pattern induced by bivectorial adaptation. This gives reason to expect that the strength of the (univectorial) MAE in the direction opposite to the trained boundary would differ from the untrained control. This also suggests an interesting follow-up experiment that uses bivectorial transparent motion (e.g., $\theta \pm 25$ deg) for the adapting stimulus on MAE trials.
In summary, the first alternative interpretation of our results does not seem convincing upon close examination. It appears that our two experimental tasks did engage highly overlapping populations of neurons, including the lateral connections between them. It must be acknowledged that the main conclusions of this article depend on this conjecture and that further research is needed to clarify the role of the lateral connectivity. Note also that the same connectivity is exactly the plasticity site proposed by the RM hypothesis.

The other alternative interpretation disputes the second premise of our reasoning. Why should we expect that modifying these representations would cause detectable MAE changes? Suppose, for the sake of argument, that there really are common plasticity mechanisms for adaptation and learning (Teich & Qian, 2003). Then the MAE post-test in the trained direction reflects the combined effect of learning plus adaptation, whereas that in the control direction reflects the effect of adaptation alone. If the adaptation effect were much stronger than the learning effect, adaptation plus learning would not differ significantly from adaptation alone. This would explain the observed lack of significant differences in the MAE data.

To answer this question we need to compare the effects of adaptation and learning. We use the change in discrimination threshold as common currency to estimate the approximate relative strengths of these two factors. Recall that the learning index for the average $d'$ data is $\approx 55\%$. To a first approximation (cf. Fine & Jacobs, 2002), this corresponds to a commensurate decrease in discrimination threshold. On the other hand, the two studies that measured the effect of adaptation on motion discrimination (Hol & Treue, 2001; Phinney et al., 1997) report $\approx 60\%$ increases in the fine discrimination thresholds for the flanker directions (and up to 20% decrease for the adapted direction). Note that the adaptation protocols in these studies were
much stronger than ours. Phinney et al. (1997), for example, used an initial adaptation period of 5 min followed by top-up periods of 10 sec interspersed between the discrimination trials. In contrast, our adaptation protocol was calibrated to avoid ceiling and floor effects, and included reset periods at the end of each trial. Our adaptation effects are thus probably smaller than 60%. Still, a conservative estimate is that the behavioral effects of adaptation and learning have approximately equal strength. This casts serious doubt on the second alternative interpretation of our results.

In conclusion, while we cannot rule out the possibility that a small amount of representation modification did occur but went undetected, our experimental technique seems powerful enough to rule out modifications that could account for a non-trivial portion of the large behavioral improvement. Most of the increase in $d'$ (and diffusion drift rate) apparently stems from some other plasticity mechanism(s).

This conclusion agrees with the results of a recent study (Law & Gold, 2008) that recorded from approximately 250 individual MT neurons from two monkeys before and during extensive training on a motion direction-discrimination task. Despite dramatic improvements in the behavioral thresholds, no significant changes in firing rates or tuning properties were observed in the MT sample, contrary to the prediction of the RM hypothesis. To our knowledge, Law and Gold (2008) is the only single-cell recording study of motion perceptual learning, but there are five similar studies of orientation-discrimination learning (R. B. Crist, Li, & Gilbert, 2001; Ghose et al., 2002; Raiguel et al., 2006; Schoups et al., 2001; Yang & Maunsell, 2004). One of them (Schoups et al., 2001) found modest changes in V1 but this result did not replicate in two subsequent attempts (Ghose et al., 2002; Yang & Maunsell, 2004). Two studies (Raiguel et al., 2006; Yang & Maunsell, 2004) found changes in area V4. While stronger than those in V1, they were still insufficient to account for the
massive behavioral improvements (see A. A. Petrov et al., 2005; Raiguel et al., 2006, for reviews). Again, these data seem inconsistent with the RM hypothesis.

The selective reweighting (SRW) hypothesis, on the other hand, offers a straightforward and natural account. It attributes the behavioral improvement to synaptic plasticity in the read-out connections to higher areas. The pattern of activation (or the tuning properties) in early representational areas such as V1, V2, or MT is not expected to change. Rather, such changes are predicted for higher areas, particularly those involved in determining the response. This is precisely the pattern observed by Law and Gold (2008)—no significant changes in MT coupled with pronounced changes in LIP, a sensory-motor area. Moreover, the changes in LIP accumulated as learning progressed and their time course correlated with the time course of the behavioral improvement (Law & Gold, 2008). A neural network model accounted for the detailed neurophysiological recordings in terms of selective reweighting of the connections between MT and LIP (Law & Gold, 2009). A closely related model had been developed on the basis of human behavioral data (Dosher & Lu, 1998, 1999; Lu et al., 2010; A. A. Petrov et al., 2005, 2006; Sotiropolous et al., 2011; Vaina et al., 1995).

Our present data lends further support to the SRW hypothesis. On this account, the $d'$ learning curves in Figures 2.2 and 2.4 stem from selective reweighting of the read-out connections. Areas V1, V2, and MT do not change and neither does the motion aftereffect. The signal-to-noise ratio of the sensory input to the decision-making areas does change, however, because task-relevant inputs gain importance, whereas task-irrelevant inputs lose it (A. A. Petrov et al., 2005). This accounts for the twofold increase in the drift rate parameter of the diffusion-model fit to the RT data (A. A. Petrov et al., 2011). The MAE duration is not affected by the discrimination
practice because it relies on a different set of read-out connections, as discussed above. These non-overlapping read-outs explain the task specificity of the learning effect.

To our knowledge, the present result is the first demonstration that MAE changes little (or not at all) after practicing a direction-discrimination task. There is one prior study (Vidnyánszky & Sohn, 2005) that did find significant MAE changes after practicing an unrelated task. Vidnyánszky and Sohn (2005) aimed to investigate the effect of practice on the efficiency of attentional selection. To that end, they constructed bivectorial transparent motion displays in which red dots moved horizontally and green dots moved vertically (or vice versa). The observers were instructed to attend one of the motion planes and ignore the other. The task was to detect brief increases of the attended dot population’s luminance. The MAE duration was measured before and after practice on the luminance task. Four of the 6 participants showed significant decreases in the MAE evoked by the motion signal which was neglected during the seven practice sessions. Abstract Vidnyánszky and Sohn (2005) concluded that “attentional suppression of task-irrelevant stimuli becomes more efficient with practice.” The effect size varied between 20% and 70% across the 4 observers. This indicates that MAE duration is a sensitive measure that can be affected by practice manipulations. Our experiment had more than enough statistical power to detect a MAE change of this magnitude.

In conclusion, the present study establishes a new empirical constraint on theories of perceptual learning. Any model of motion direction-discrimination learning in particular must be able to produce 55% $d'$ improvement (and 99% drift-rate improvement, (A. A. Petrov et al., 2011)) with less than 10% change in either static or dynamic MAE. The selective reweighting hypothesis predicts this pattern, whereas it poses a great challenge to the representation modification hypothesis.
CHAPTER 3

Dissociable perceptual-learning mechanisms revealed by diffusion-model analysis

3.1 Introduction

Performance on perceptual tasks improves with practice. This perceptual learning occurs in all sensory modalities (Fahle & Poggio, 2002) and is of great theoretical interest (Lu et al., 2009) and practical importance (Polat, 2009). However, its mechanisms are still poorly understood. The signature property of perceptual learning is its stimulus specificity: The improvement is (partially) restricted to stimuli similar to those used in training (see, e.g., Ahissar & Hochstein, 1997). This indicates that (part of) the neural substrate of the learning effect resides in the early stages of the sensory processing pathway (Karni & Sagi, 1991), which may involve changes in the early sensory areas (see Gilbert et al., 2001, for a review) and/or the read-out connections to decision areas (Dosher & Lu, 1998; Law & Gold, 2008; A. A. Petrov et al., 2005).

Most perceptual learning studies use accuracy (or, conversely, sensitivity) as their dependent variable (Fahle & Poggio, 2002). Response times (RTs) are typically ignored, or sometimes the mean RTs are analyzed but errors are ignored (e.g., Ding,}

The text of this chapter is taken from A. A. Petrov et al., 2011
Song, Fan, Qu, & Chen, 2003). Either approach neglects the well-known speed-accuracy trade-off (Pachella, 1974). Also, such analyses use only one data point per block per stimulus type. This restricted empirical base tends to give rise to restrictive theoretical accounts that attribute all learning to one cortical site (e.g., Karni & Sagi, 1991) or to one learning rule operating at task-dependent levels of the processing hierarchy (Ahissar & Hochstein, 2004).

It seems highly unlikely, however, that perceptual learning is a monolithic phenomenon. Rather, even the simplest task engages multiple brain systems, and the overall behavioral improvement arises from multiple contributions. Univariate data tend to obscure this heterogeneity, whereas richer data sets reveal it. For example, dissociable learning mechanisms have been identified using event-related potentials (Ding et al., 2003), fMRI (Li, Mayhew, & Kourtzi, 2009; Vaina, Belliveau, des Roziers, & Zeffiro, 1998), and external-noise manipulations (Dosher & Lu, 1998).

Figure 3.1: Schematic illustration of the decision process in the diffusion model. Three stochastic paths result in a fast correct response, a slow correct response, and an error response. Because of the positive drift rate $v$, the correct (upper) boundary is reached more often than the incorrect (lower) boundary. RT, response time; $a$, boundary separation parameter; $z$, starting point parameter.
The present study pioneers the use of RT distributions for studying perceptual learning. A typical RT distribution can be described approximately by five quantiles (Ratcliff, 1979) that divide the probability mass into six bins. This makes 12 bins in total, for correct and error responses. Since the total mass is fixed, there are 11 degrees of freedom per block. Obviously, this carries much more information than does accuracy alone, but it presents an analytic challenge. We use the diffusion model (DM; Ratcliff, 1978) to analyze such data. This is analogous to the use of signal detection theory (Macmillan & Creelman, 2005) to convert hits and false alarms into theoretically motivated estimates of discriminability and bias. Analogously, DM converts hits, false alarms, and RT distribution statistics into estimated parameters of various processing components.

DM characterizes the process of making simple two-choice decisions (see Ratcliff & Mckoon, 2008, for a review). The core of the model is a diffusion process that describes the stochastic accumulation of evidence for two competing responses (Fig. 3.1). The process terminates when the accumulated evidence reaches one of two decision boundaries (or criteria). The better the information about the stimulus, the larger the drift rate $v$ in the correct direction. Due to within-trial variability in evidence accumulation, processes with the same mean drift rate terminate at different times (producing RT distributions) and sometimes at the wrong boundary (producing errors). The model has seven free parameters: mean drift rate $v$, across-trial variability $\eta$ in drift rate, boundary separation $a$, mean starting point $z$ between the boundaries ($0 < z < a$), across-trial range $s_z$ in starting point, mean nondecision time $T_{er}$, and across-trial range $s_t$ in nondecision time. The first five parameters affect both accuracy and speed in the model. For example, increased drift rate produces higher accuracy and faster RTs, whereas increased boundary separation produces higher accuracy but slower RTs, all else being equal. The two nondecision parameters ($T_{er}$, $T_{er}$...
and \( s_i \) affect the RTs only. They describe the combined duration of processes such as stimulus encoding, memory access, and response execution. All parameters are estimated simultaneously by fitting the model to behavioral data. DM is tightly constrained, particularly in experimental designs involving stimuli at multiple difficulty levels. The proportions correct and the shapes of all RT distributions across all difficulty levels must be accounted for with a fixed parameter set within a block. Only the drift rate \( v \) is allowed to vary as a function of difficulty. DM has been tested and validated extensively. For instance, speed-related instructions affect the boundary separation parameter, whereas stimulus manipulations affect the drift rate (see Ratcliff & McKoon, 2008, for a review). However, the effects of practice have been studied relatively little (Dutilh, Krypotos, & Wagenmakers, 2011; Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009; Ratcliff, Thapar, & McKoon, 2006), and the effects of perceptual learning are largely unknown.

Various perceptual learning mechanisms can be identified using the DM framework. First and foremost, the drift rates are expected to improve with practice (Dutilh et al., 2009; Ratcliff et al., 2006). The drift rates in perceptual tasks measure the quality of the sensory input to the decision process. Most theories of perceptual learning (e.g., Ahissar & Hochstein, 2004; Karni & Sagi, 1991; A. A. Petrov et al., 2005) predict a drift-rate increase. This increase affects both accuracy and RT. The diffusion analysis captures both effects in a single measure, and thus can detect weaker learning effects. It can also estimate the variability attributable to speed-accuracy trade-offs across observers and across sessions. Removing this variability from the error term improves the power of the analysis still further (C. Liu & Watanabe, 2010).

The DM framework also allows us to study the temporal aspects of perceptual learning. Our hypothesis is that observers may learn to deploy attention during the
critical period of the trial sequence. Attention plays an important role in perceptual learning (Ahissar & Hochstein, 2002) and in learning more generally (Kruschke, 2003). Though relatively neglected in the literature, the temporal aspects of attention are as important as its spatial aspects (Large & Jones, 1999). In particular, the decision mechanism must be timed relative to the stimulus onset (Purcell et al., 2010; Ratcliff & Smith, 2010). If triggered before sensory evidence becomes available, the diffusion process merely accumulates noise. If triggered too late, valuable evidence can be lost. Therefore, one way to improve performance is to calibrate the onset of the decision process. To test this synchronization hypothesis, we conducted an experiment in which a beep reliably preceded the stimulus onset. The critical prediction was that the variability of nondecision times would be high at first and decrease significantly with practice. Moreover, this decrease should transfer across stimuli as long as the temporal structure remains the same on each trial. In contrast, the improvement in
drift rates was expected to be partially stimulus specific. The experiment included a transfer test to assess the specificity of learning. The task was a visual motion-direction discrimination.

3.2 Method

3.2.1 Participants

A total of 27 university students with normal or corrected-to-normal vision were paid $6/h plus a bonus contingent on their accuracy.

3.2.2 Stimuli and apparatus

The stimuli were filtered-noise textures moving behind a circular aperture (see Fig. 3.2; diameter 10, speed 12/s). The filter had a Gaussian cross-section along the frequency axis in the Fourier domain (peak frequency 3 cycles/deg at all orientations, full width at half height 4 octaves). On each trial, the filter was applied to a fresh sample of independent, identically distributed Gaussian noise. All stimuli were generated in MATLAB and presented on a 21-in. NEC Accusync 120 color CRT (96 frames/s, mean luminance 16.6 cd/m², chinrest at 93 cm, free viewing in a darkened room). Each trial began with a brief beep. The texture appeared 500 ms later, moved for 397 ms, and then disappeared. The beep onset always preceded the texture onset by exactly 500 ms, and thus could serve as a reliable attentional cue.

3.2.3 Task and procedure

The direction discrimination task was defined with respect to a reference direction $\theta$. The actual motion direction took four possible values: $(\theta - 3.5)$, $(\theta - 2)$, $(\theta + 2)$, and $(\theta + 3.5)$ degrees from vertical. Each block randomly presented 120 stimuli of
Figure 3.3: Learning profiles for the group-averaged discriminability (a) and mean response times (b) in the raw data, and for various parameters of the diffusion model (cf). The observers practiced motion-direction discrimination for eight blocks (black symbols) and then were tested on the same task with motion in the orthogonal direction (open symbols). The error bars are 90% within-subjects confidence intervals. Shaded areas mark two additional sessions of motion aftereffect measurements.
Table 3.1: Descriptive statistics of the discriminability $d'$ for easy and difficult stimulus pairs and for all diffusion-model parameters. The learning and specificity indices were calculated from group-averaged data (±80% bootstrap confidence interval; see note 2 for details). The trend analyses were performed on the individual data using orthogonal contrasts (linear and quadratic). The error term had $df = 182$ in all ANOVAs. $\eta^2_p = \text{partial effect size} = \text{variance of the contrast relative to itself plus the error.}$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Learning</th>
<th>Specificity</th>
<th>Linear Trend</th>
<th>Quadratic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Index LI</td>
<td>Index SI</td>
<td>$F$</td>
<td>$\eta^2_p$</td>
</tr>
<tr>
<td>Easy discriminability $d'_1$</td>
<td>.55 ± .08</td>
<td>.55 ± .09</td>
<td>152***</td>
<td>.74</td>
</tr>
<tr>
<td>Difficult discriminability $d'_2$</td>
<td>.65 ± .10</td>
<td>.66 ± .12</td>
<td>165***</td>
<td>.75</td>
</tr>
<tr>
<td>Mean discriminability $d'$</td>
<td>.55 ± .08</td>
<td>.60 ± .10</td>
<td>202***</td>
<td>.79</td>
</tr>
<tr>
<td>Mean reaction time RT</td>
<td>−.27 ± .03</td>
<td>.37 ± .08</td>
<td>251***</td>
<td>.82</td>
</tr>
<tr>
<td>Easy drift rate $v_1$</td>
<td>0.96 ± 0.23</td>
<td>.66 ± .10</td>
<td>91***</td>
<td>.62</td>
</tr>
<tr>
<td>Difficult drift rate $v_2$</td>
<td>1.05 ± 0.25</td>
<td>.72 ± .09</td>
<td>95***</td>
<td>.64</td>
</tr>
<tr>
<td>Mean drift rate $v$</td>
<td>0.99 ± 0.23</td>
<td>.68 ± .09</td>
<td>98***</td>
<td>.64</td>
</tr>
<tr>
<td>Mean nondecision time $T_{er}$</td>
<td>−.20 ± .04</td>
<td>.22 ± .10</td>
<td>62***</td>
<td>.53</td>
</tr>
<tr>
<td>Nondecision time range $s_t$</td>
<td>−.56 ± .06</td>
<td>.00 ± .08</td>
<td>161***</td>
<td>.75</td>
</tr>
<tr>
<td>Boundary separation $a$</td>
<td>−.09 ± .07</td>
<td>n/a</td>
<td>7.32**</td>
<td>.10</td>
</tr>
<tr>
<td>Drift variability $\eta$</td>
<td>.37 ± .21</td>
<td>n/a</td>
<td>6.00*</td>
<td>.08</td>
</tr>
<tr>
<td>Starting point range $s_z$</td>
<td>−.03 ± .25</td>
<td>n/a</td>
<td>0.37</td>
<td>.00</td>
</tr>
</tbody>
</table>

The instructions designated the first two as “counterclockwise” and the other two as “clockwise.” The observers pressed a key with their left hand to respond “counterclockwise,” and another key with their right hand to respond “clockwise.” We used four stimuli in a binary task in order to prevent same–different comparisons with the previous trial and to constrain the DM.

Since the task was quite monotonous, many students from our participant pool
tended to sacrifice accuracy for speed. To prevent guessing and keep the observers engaged, the procedure rewarded accuracy and penalized excessively fast RTs. The reward for each correct response was a bonus point. The penalty for each error was the loss of a bonus point, an unpleasant beep, and the addition of 250 ms to the 800-ms intertrial interval. The cumulative bonus was displayed prominently at all times. The penalty for excessively fast (< 250 ms) RTs was a “slow down” message that forced the participant to wait for 1,500 ms. RTs between 250 and 500 ms incurred a silent penalty of 2*(500−RT) ms. Thus, the fastest way to complete a trial was to produce a correct response in exactly 500 ms.

Each participant completed a total of 4,800 discrimination trials in 10 blocks across five sessions. Two additional sessions—before Block 1 and after Block 8—measured the motion aftereffect (MAE; Mather et al., 1998) in the trained, test, and two control directions\(^1\). Fourteen participants trained with \(\theta = -50\) on Blocks 1–8 and on a “mini-block” consisting of the first 120 discrimination trials on the last session. These participants then tested with \(\theta = +40\) on Blocks 9 and 10. The other 13 participants followed the same schedule but trained on \(\theta = +40\) and tested on \(\theta = -50\).

### 3.2.4 Data analysis

The data for clockwise and counterclockwise stimuli were pooled because this distinction had no statistically significant effects. The discriminability (\(d'\)) was calculated for the easy (\(\Delta = 7\)) and difficult (\(\Delta = 4\)) pairs in each block. Seven DM parameters (easy \(v\), difficult \(v\), \(a\), \(T_{er}\), \(s_t\), \(\eta\), and \(s_z\); \(z = a/2\)) were estimated for each observer in each block. An iterative algorithm minimized the \(\chi^2\) discrepancy between the predicted and observed quantile RTs (Ratcliff & Tuerlinckx, 2002).

\(^{1}\)The MAE sessions do not affect the interpretation of the present results.
We used trend analysis to test the statistical significance of the learning effects during Blocks 1–8. We also calculated two quantitative indices: The *learning index* LI = \( \frac{X_8 - X_1}{X_1} \) for a variable \( X \) quantified the improvement by the end of training relative to the initial performance (Fine & Jacobs, 2002). The *specificity index* SI = \( \frac{X_8 - X_9}{X_8 - X_1} \) quantified the disruption caused by the switch to the orthogonal direction in Test Block 9 (Ahissar & Hochstein, 1997).

### 3.3 Results

Both discriminability and mean RT improved with practice (Fig. 3.3) and showed highly significant linear and quadratic trends (Table 3.1). The \( d' \) profiles for easy and difficult discriminations were approximately proportional to each other \( d'_{ez} \approx k_d d'_{diff} \) with \( k = 1.65 \pm 0.13 \), in agreement with published data (A. A. Petrov et al., 2005). The learning effects were partially specific to the trained reference direction, although the degree of specificity differed significantly for the two dependent measures. The specificity index was \( SI = .60 \pm .10 \) for \( d' \) and \( .37 \pm .08 \) for the mean RT.

The DM achieved good fits, evident in the quantile probability plots in Fig. 3.4 and the scatterplots in Fig. 3.5. The former show the proportions of correct and error responses (on the x-axis) and the corresponding RT distributions (summarized by the .1, .3, .5, .7, and .9 quantiles on the y-axis). The model (circles) tracks the data (x’s) well\(^3\). The scatterplots show that the model can reconstruct the data for each individual on each block to a good approximation. The quality of the fit, coupled

\(^2\)All indices throughout the text and in Table 3.1 are reported as \( I \pm CI \), where \( I \) is the index calculated from the group-averaged data in Figure 3.3, and \( \pm CI \) is the 80% bootstrap confidence interval. We estimated the variance of the group-level indices by resampling the participants with replacement into 1,000 “groups” and repeating the calculation for each group.

\(^3\)The quantitative measure \( (\chi^2) \) of goodness of fit confirms this. See the online supplement for details.
with past research (Ratcliff & Mckoon, 2008) validating the DM in conditions similar to ours, suggests that the DM parameters offer a concise characterization of the underlying cognitive processes. There were statistically significant learning effects for all DM parameters except the starting point variability $s_z$ (Table 3.1). The twofold improvement in drift rate (Fig. 3.3c) indicates that the quality of the sensory input to the decision process increases with practice. The learning index for the drift rate $v$ (LI$= 0.99 \pm 0.23$) was significantly⁴ higher than the $d'$ learning index ($0.55 \pm 0.08$). This is because $v$ reflects learning in both accuracy and speed. The improvement was largely (but not entirely) specific to the trained reference direction (SI$= .68 \pm 0.09$).

The parameters describing the distribution of nondecision times across trials also improved significantly. The mean nondecision time $T_{er}$ decreased by 20% on average (Fig. 3.3d). The specificity index for $T_{er}$ ($0.22 \pm 0.10$) was significantly⁵ lower than that for the mean overall RT ($0.37 \pm 0.08$). This is because the improvement in overall RT stems in part from the stimulus-specific increase in drift rate. The nondecision variability $s_{t}$ is of particular interest. As predicted by the synchronization hypothesis, it was high at first (283 ms during Block 1, Fig. 3.3f) and decreased steeply to 120 ms by the end of training. Moreover, the improvement transferred fully to the orthogonal direction of motion (SI$= .00 \pm 0.08$).

There was a small but statistically significant decreasing linear trend in the boundary separation parameter $a$ (Table 3.1). This suggests a slight adjustment in the speed-accuracy trade-off. The drift-rate increase apparently offset this adjustment.

⁴Paired-samples bootstrap $z = 2.8, p < .01$.  
⁵Paired-samples bootstrap $z = 2.5, p < .013$
Figure 3.4: Quantile probability plots illustrating the wealth of data and the quality of the fit. Each panel has 22 empirical degrees of freedom: the proportions of errors and correct responses for the easy and difficult discriminations (plotted on the x-axis) and the .1, .3, .5, .7, and .9 quantiles of the corresponding response time distributions (stacked vertically on the y-axis). For example, the x-coordinate of the leftmost, bottommost data point in the top panel indicates the initial error rate (.18) for easy stimuli. The y-coordinate indicates the leading edge (.1 quantile 530 ms) of the corresponding RT distribution. After 4 days of training (middle panel), the performance improves on both measures (.08 rate and 480 ms, respectively). This illustration is based on group-averaged data; the analyses in the text (and the predictions in Fig. 5) are based on model fits to individual data.
and prevented a drop in accuracy. Finally, there was a marginally significant decrease in the across-trial variability in drift rate (η) but no significant changes in the variability in starting point (sz). See Appendix A for details.

3.4 Discussion

This article makes two contributions: methodological and substantive. The methodological one is to demonstrate the applicability of the diffusion model to perceptual learning research. This research typically involves simple two-choice tasks, RTs faster than 1 s, and thousands of trials—precisely the conditions that DM is best suited for. The model accounted for all behavioral data—22 measurements per block in our experiment—with seven parameters. Figure 3.5 demonstrates that this reduction occurs with little loss of information. Moreover, the DM parameters have theoretically motivated and empirically validated interpretations in terms of component processes (see Ratcliff & McKoon, 2008, for a review). Neurophysiological correlates of several such processes have been found, narrowing the gap between brain and behavior (see J. I. Gold & Shadlen, 2007, for a review).

The DM analysis confers several advantages, all of which stem from its access to more detailed data. First, the drift-rate parameter v is sensitive to improvements in both accuracy and speed. This produces stronger learning effects, manifested here in the high learning index for v. Second, DM accounts for speed-accuracy trade-offs (Dutilh et al., 2009; Ratcliff et al., 2006). The associated variability can be partialed out of the error term (C. Liu & Watanabe, 2010). The third and most important advantage is that DM reveals phenomena that cannot be reached by traditional methods. This leads to the substantive contribution of this article.

We identified two distinct learning mechanisms with markedly different specificities. The first mechanism improves the quality of the sensory input to the decision
Figure 3.5: Scatterplots illustrating the quality of the fit to individual data. The diffusion model was fit separately in each block (297 fits = 27 observers 11 blocks). Each panel contains 594 points (= 2972 difficulty levels). RT, response time
process, and manifests itself in increased drift rates. This improvement is partially stimulus specific and is compatible with most theories of perceptual learning, including representation modification (Gilbert et al., 2001; Karni & Sagi, 1991) and selective reweighting (Dosher & Lu, 1998; Law & Gold, 2008; A. A. Petrov et al., 2005). The second mechanism manifests itself in decreased and less variable nondecision times.

As discussed in the introduction, triggering the decision process too early or too late impairs performance. Recent theoretical work has suggested that the diffusion process can be gated (Purcell et al., 2010) or disinhibited (Ratcliff & Smith, 2010) at the time in which usable sensory evidence becomes available at the decision-making areas. One intriguing interpretation of our data is that the observers improved the timing of this internal gating operation. We speculate that during the first session, this timing was good on some trials but too slow on others. This inflated the nondecision time variability and degraded the mean drift rates. Apparently, the slow nondecision times were eliminated with practice. Under this synchronization hypothesis, the nondecision time variability $s_t$ decreased as the observers learned the temporal relationship between the beep and the stimulus onset. Because this relationship was independent of motion direction, the $s_t$ decrease transferred fully to new stimuli.

The nondecision time in the DM covers the combined duration of stimulus encoding and response execution. Probably both improved with practice. Although the present data cannot differentiate their relative contributions, it seems unlikely that the speed-up can be attributed entirely to motor factors. The drop in $T_{cr}$ (Fig. 3.3d) was partially stimulus specific, and the 160-ms drop in $s_t$ (Fig. 3.3f) seems too large relative to the duration of simple RTs (Luce, 1986).

We used a combination of bonuses and “slow down” messages to prevent fast guessing. Pilot data indicated that without such incentives, some students from
our participant pool tended to respond so quickly that their accuracy was barely above chance. Not surprisingly, there was little improvement with practice. Our procedure minimized this behavior and produced robust learning effects. Still, it must be acknowledged that our results may not generalize well to less motivated observers.

In conclusion, perceptual learning is not a monolithic phenomenon. Two learning mechanisms with different properties seem to be at work in our study and can be dissociated with the aid of the diffusion model.
CHAPTER 4

The contents of visual working memory systematically bias perception

4.1 Introduction

Visual short term memory is the mechanism by which task-specific perceptual information is selectively chosen and retained across relatively short time periods. Despite the ubiquitous role of VSTM in everyday life (e.g., by aiding the process of smoothing otherwise erratic visual input between saccadic eye movements) and in many common psychophysical tasks, much uncertainty remains regarding the mechanisms responsible for these transient memory representations. Certainly the goal-oriented nature of VSTM reflects an executive prefrontal cortical component, as commonly found in studies of working memory (Curtis & D’Esposito, 2003; E. K. Miller et al., 1993). It is unclear, however, how higher-level executive brain regions involved in multi-modality coordination can alone account for the fidelity often exhibited by VSTM (Magnussen & Greenlee, 1992).

A growing body of research suggests that VSTM is subserved by interactions between large cross-cortical networks, at least some of which partially overlap with sensory areas involved in the original encoding of visual input. The “sensory-recruitment” model of VSTM (Awh & Jonides, 2001; D’Esposito, 2007; Postle, 2006), which posits the role of early visual areas in the maintenance of VSTM information, has found
increasing support in several forms, including sustained neural activity in nonhuman primate cells selective for stimulus information during the memory retention interval (Pasternak & Greenlee, 2005; E. K. Miller et al., 1993; Chelazzi et al., 1993), and stimulus-specific sustained activity in human observers in areas as early as V1-V4 as evidenced by functional MRI (Ester et al., 2013; Konstantinou et al., 2012; Ester et al., 2009; Serences et al., 2009; Sligte et al., 2009; Harrison & Tong, 2009). Further electrophysiological studies in humans have found consistent ERP markers during VSTM maintenance known as “contralateral delay activity” (CDA) (Anderson, Vogel, & Awh, 2011; Ikkai, McCollough, & Vogel, 2010; McCollough, Machizawa, & Vogel, 2007; Vogel & Machizawa, 2004), and evidence from transcranial magnetic stimulation (TMS) studies support claims for an interaction between VSTM and early visual cortex (van de Ven & Sack, 2013; Saad & Silvanto, 2013b; van de Ven et al., 2012; Soto et al., 2012). Research on visual imagery is converging on similar results. For example, both observer behavior (Farah, 1989; Ishai & Sagi, 1997) and activity in visual cortex has been found to be similar for viewing an oriented line, holding that stimulus in VSTM, and even simply visualizing the line (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Cattaneo et al., 2012; Cattaneo, Vecchi, Pascual-Leone, & Silvanto, 2009). Further, Silvanto and Cattaneo (2010) demonstrated that stimulus features retained in VSTM can appear in TMS-induced phosphenes when stimulation occurs over cortical regions involved in representing the original stimuli. Finally, research on visual field maps suggests that VSTM may have retinotopic representations in early visual areas (see Barton & Brewer, 2013, for a review).

Visual short term memory has a highly limited capacity (e.g., Barton, Ester, & Awh, 2009; Bays & Husain, 2008; Zhang & Luck, 2008; Awh et al., 2007; Alvarez & Cavanagh, 2004; Vogel et al., 2001; Luck & Vogel, 1997), and items in VSTM interfere with one another as a function of their similarity. That is, observers are able to
more accurately remember, for example, the orientation and spatial frequency of two stimuli, compared to storing two different orientations simultaneously (Magnussen & Greenlee, 1999). Interestingly, the fact that VSTM suffers from these deficiencies is consistent with the idea that the mechanisms of visual working memory share overlapping resources with sensory input, as current visual input should get the lion’s share of resources, and memory should not be so phenomenologically strong so as to impinge on perception in a hallucinatory manner. Evidence from “memory masking” studies has previously demonstrated the dominance of perception over VSTM, as well as the interaction of the two systems. For example, Magnussen & Greenlee, 1992 found that the velocity of a stimulus presented during the memory retention interval could interfere with observers’ ability to make judgments about velocity information held in VSTM.

Research on memory masking has established that the neural economics of sharing resources has some limiting consequences for VSTM. A related question is what, if any, effect the contents of visual working memory have on external visual input gathered during the retention interval. The answer to this question is not clear, as even neural overlap does not guarantee that weak persistent activation in VSTM can muster enough influence to overcome the much stronger effect of external input. Several recent studies have begun to address this question with results suggesting that the contents of VSTM can systematically modulate visual input. For example, VSTM appears to both create and affect existing tilt aftereffects (Saad & Silvanto, 2013b, 2013a; Scocchia, Cicchini, & Triesch, 2013), create the motion repulsion illusion (Kang, Hong, Blake, & Woodman, 2011), bias the perception of visual motion direction (Mendoza, Schneiderman, Kaul, & Martínez-Trujillo, 2011), and reduce detection sensitivity (Konstantinou et al., 2012). However, there is disagreement over whether VSTM facilitates or inhibits the perception of stimuli.
Additionally, the aforementioned studies used a common experimental methodology in which behavior under working memory load is compared to a no-load condition. A typical analysis for this approach is to compare differences in response accuracy or thresholds without reporting or analyzing response times (or, at best, conducting an analysis of accuracy and RT independently). However, it is known that VSTM load increases response times (RT) on concurrent tasks (Woodman, Vogel, & Luck, 2001). Such changes in response behavior typify the speed-accuracy trade-off (SAT), and have a long history outside of vision science (Pachella, 1974). The SAT describes observers’ ability to either favor accuracy by slowing down, or speed by allowing for proportionally more mistakes.

To properly model the SAT, it is necessary to jointly analyze response times and accuracy, as neither measure is fully diagnostic in isolation. This presents a problem for previous studies, as it is unclear whether reported changes in responses were due to a perceptual effect or due to known changes in response times between load conditions. In the following study, we address these problems by using the Ratcliff Drift Diffusion Model (DDM) (Ratcliff, 1978). The DDM is a well established model of decision processes that distills response time distributions and accuracy into a collection of theoretically well-justified parameters corresponding to known psychological properties (for a review, see Ratcliff & McKoon, 2008). Similar DDMs have been leveraged in innumerable studies for the last 3 decades to model the SAT, and have recently been employed to investigate the mechanisms behind the learning of simple visual tasks (A. A. Petrov et al., 2011).

In the current study we test the hypothesis that the contents of VSTM can systematically interfere with the perception of subsequent visual input on a concurrent task while also addressing an alternative explanation not previously discounted by earlier studies. Specifically, we directly measure changes in response bias due to the
modulatory effects of VSTM on the tilt aftereffect (TAE). Consistent with previous studies, we found that VSTM can influence the encoding of external visual input. However, similar to Woodman et al. (2001), we found a uniform increase in response times for observers under VSTM load relative to a no-load condition. Importantly, we address this previously neglected confound by jointly analyzing response bias and response time distributions using the Drift Diffusion Model (DDM). Using the DDM, we describe both Monte Carlo simulations and theoretical arguments for how changes in RT alone can account for the apparent effects of VSTM in previous studies. Critically, we found that VSTM-induced changes in perceptual information led to changes in behavior above and beyond those due to decision-related changes in response behaviors. Further, our results strongly suggest a facilitatory effect of VSTM on the tilt aftereffect.

4.2 Experiment

4.2.1 Method

Participants

27 naive university students participated in this study. All subjects had normal or corrected to normal vision. 11 observers participated for credit in an introductory psychology course. The remaining 16 participants were given an hourly monetary rate plus an additional bonus contingent on their accuracy.

Stimuli and Apparatus

Stimuli consisted of Gabors that subtended 6.1 degrees of visual angle and had a spatial frequency of 2.3 cycles per degree. A linearly-ramped circular mask, fully transparent at radius $R = 5.1^\circ$ and fully opaque at $R = 6.1^\circ$, imposed an aperture over each stimulus such that each Gabor’s outer edge blended smoothly into the...
uniform gray background color of the monitor. Each trial consisted of three sequential presentations of Gabor patterns that differed only in their orientation and location on screen, which we have labeled in order of presentation as “Sample”, “Target”, and “Match.”

True, gravitational vertical was a relevant orientation during each trial, and several steps were taken to ensure comparability across subjects. First, a black circular mask was placed over the computer monitor with an inner diameter of 28.5 cm and an outer diameter of 68.5 cm. The purpose of this mask was to minimize participants’ ability to use nearby orientation cues (e.g., the edges of the square monitor, company logos printed on the experimental equipment, etc.) for completing the task. Second, at the beginning of each session, observers completed 6 calibration trials in which a custom rotary input knob was used to orient an on-screen Gabor to vertical. Each subject’s calibration average was used as “true” vertical for the remainder of the session. This subjective vertical value $\delta^s$ was added to all stimulus orientations, shifting the orientation toward each subjects’ subjective sense of gravitational vertical.

Observers used all three stimuli to complete two concurrent tasks. Sample Gabors served as both a memory cue for a VSTM task and as adapters similar to tilt after-effect studies. Sample orientations were normally distributed ($\sigma = 2^\circ$), with per-trial orientations determined by $\theta + \delta^s$, where $\theta$ was centered on one of three values, $-15^\circ, 0^\circ$, or $+15^\circ$. Target Gabors were always presented at vertical ($0^\circ + \delta^s$).

Images were created on an Apple iMac 2.66Hz Intel Core 2 Duo with an ATI Radeon HD2600 Pro graphics card using Matlab (The MathWorks, 2004) and the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 20” NEC AccuSync 120-BK CRT Monitor with a resolution of 1024x768, a refresh rate of 96 Hz, and a mean luminance of 16.6 cd/m$^2$. The monitor was calibrated before initial use using a Minolta 1° luminance meter, and the look-up table
Figure 4.1: Sequence of events for trials in our two conditions. In both conditions, observers were shown three sequentially presented Gabors—Sample, Target, and then Match). Observers completed a binary pseudo-discrimination task using the Target Gabor and recreated the orientation of the Sample using the Match Gabor either from memory (VSTM trials) or through direct comparison (no-load trials). An attentional task ensured observers actively viewed the Sample under both conditions (see method). A black circular mask hid the computer monitor’s edges from view to minimize orientation cues.

was adjusted to correct for the gamma function of the CRT. There were no other light sources in the room, and observers’ heads were held fixed by a chin rest at a distance of 927 mm from the screen, causing one degree of visual angle to span \( \approx 43 \) pixels.

**Procedure**

Participants initiated the trial sequence with a key press. Three stimuli, the Sample, Target, and Match Gabors, were displayed one after another in that order at three
different screen positions, left (L), center (C), and right (R), respectively. Each stimulus was preceded with a 500 msec fixation dot to facilitate controlled eye movements. Observers completed two concurrent tasks during each trial. The task of primary interest was a simple counter-clockwise/clockwise (CCW/CW) judgment made based upon the apparent orientation of the Target Gabor. Subjects were instructed to indicate by button press whether the Target Gabor was presented CCW or CW of “true” vertical. Importantly, even though the instructions claimed that the Target would be rotated away from vertical in either direction by some small amount, in actuality the Target Gabor was always presented at $\delta^\circ_s$ (i.e., gravitational vertical, as defined by each subjects’ calibration). As such, the task measured response bias directly rather than discriminability.

This pseudo-discrimination task was flanked by an orientation match task that differed according to two experimental conditions: a VSTM condition and a no-load condition in which memorization was not required to complete the task. Each subject completed conditions on separate sessions, with condition order counterbalanced across subjects. In a trial, the Sample was presented at location L for 1500 msec near one of three orientations (see Stimuli and Apparatus). For the VSTM condition, participants were asked to hold the orientation of the Sample in visual short-term memory. Observers were encouraged to forget the Sample stimulus in the no-load condition. Next, the Target, oriented at $\delta^\circ_s$, was displayed for 250 msec at location C. Subjects indicated a CCW or CW response by button press. Finally, the Match was presented at screen position R at a starting orientation determined by adding a uniformly chosen value of $\pm[30,60]^\circ$ to the previous orientation of the Sample. Observers were instructed to use a custom-built rotary input knob to align the Match to the orientation of Sample Gabor held in VSTM. For the no-load condition, the original Sample reappeared onscreen at position L simultaneously with the Match.
Gabor, enabling participants to complete a simple perceptual match. To ensure that subjects were viewing the Sample equally across the VSTM and no-load conditions, a simple letter detection task was embedded within the Sample’s presentation. A faint letter ”E” or ”F” appeared at the center of the Sample Gabor. Subjects pressed the corresponding letter on the keyboard to demonstrate that they had seen the stimulus.

Feedback consisted of a mixture of auditory tones, visual indicators, and accuracy-based bonus points. An audible tone was played for incorrect responses on the letter detection task. Visual feedback (750 msec), dependent on the veridical orientation of the Sample $M_v$ and the observer’s response $M_s$, was provided on the orientation match task for both conditions. Immediately after response collection, a black dashed line appeared over the Match indicating the veridical orientation ($M_v$) of the original Sample. Simultaneously, a solid colored line indicating the observer’s response ($M_s$) appeared in color, allowing for quick assessment of error. The color of the line was determined by error magnitude $e = |M_v - M_s|$, with green feedback for $e <= 5^\circ$, yellow for $5^\circ < e <= 10^\circ$, and red for $e > 10^\circ$. Participants earned bonus points for each color at a rate of 5, 2, and 0 respectively.

Because “correct” and “incorrect” were undefined for the CCW/CW task, genuine feedback was impossible. Instead, a quasi-probabilistic feedback method was utilized, the behavior of which varied dependent on the orientation of the Sample. Feedback on this condition was theoretically motivated. For trials where the Sample’s orientation $\theta = 0^\circ$, we would expect no tilt aftereffect. For this condition, an unbiased observer would have a CW response rate of 50%. It is possible, however, that some subjects would exhibit a response bias under this condition. To discourage this behavior, we implemented a sliding window that tracked the participant’s response (CCW or CW) and whether or not they had received feedback for the previous 11 trials (an odd number to prevent ties). If the response for the current trial matched the most
frequent response from the sliding window, they received feedback (this feedback was throttled at 80% even under biased behavior to prevent discouragement). By contrast, for trials in which the Sample orientation $\theta = -15^\circ, +15^\circ$, we expected systematic response bias as a function of Sample orientation. To avoid influencing these response characteristics, infrequent (20%) feedback was applied in a purely probabilistic manner. At the end of the study, bonus points were converted to an additional cash bonus to encourage motivation.

4.3 Results

Observers were able to successfully maintain and recall Gabor orientations on the VSTM task with high levels of fidelity. For analysis purposes, memory recall data were categorically binned according to whether the Sample stimulus’ orientation was centered on $\theta = -15^\circ$, $0^\circ$, or $+15^\circ$. Median recall error, defined as VSTM response - Sample orientation, was $5.54^\circ$, $3.46^\circ$, $5.80^\circ$ for Sample orientations centered on $-15^\circ$, $0^\circ$, and $+15^\circ$, respectively. Performance for orientations near vertical was anticipated to be better due to the well-known oblique effect (Campbell, Kulikowski, & Levinson, 1966; Orban, Vandenbussche, & Vogels, 1984; Furmanski & Engel, 2000). In no-load condition trials, subjects performed a simple perceptual orientation match which provided a measurement of response sensitivity on our apparatus. No-load response errors were $2.11^\circ$, $1.12^\circ$, and $1.8^\circ$. Assuming identical response sensitivity across conditions, VSTM orientation recall errors were approximately $3.43^\circ$, $2.34^\circ$, and $4.00^\circ$. 
Figure 4.2: Simulated data (lines) and experimental data (points). The solid line represents simulated data under the no-load condition and was informed by our experimental data. For the VSTM-load condition, data were generated under two different hypotheses. Under hypothesis 1, drift rate is allowed to vary under VSTM-load relative to no-load (dashed line), while the remaining diffusion parameters remain fixed. Under hypothesis 2 (dotted line), boundary separation is allowed to increase uniformly under VSTM-load relative to no-load, while the remaining diffusion parameters remain fixed. Data points represent response frequencies for binned orientations from our experimental data. Circles show no-load trials and Xs show VSTM-load trials. The inset highlights nonparametric kernel density estimates for the simulated response time distributions under no-load (single solid line) and VSTM-load (both broken lines). Although we hypothesize that drift rate changes as a function of memory load, it is clear that uniform changes in response criterion can produce equivalent changes in response probabilities and shifts in RT distributions. Error bars are 1 standard error of the mean.

Although all stimuli on the CCW/CW task were displayed at vertical, observers were unaware of the ruse and responded as if the Target Gabor was tilted. This setup provided a measure of response bias as a function of Sample Gabor orientation. Figure 4.2 shows the probability of a clockwise (CW) response as a function of Sample orientation, binned by 15° intervals, at a group level analysis. At a first pass, two
trends seem evident. First, as the orientation of the Sample increased, CW responses decreased. This expected result is indicative of a repulsive tilt aftereffect. That is, the orientation of the Sample served as an adapting stimulus, causing observers to perceive the vertically-oriented Target as tilted in the opposing direction. Second, this repulsive effect appears to be exaggerated for the VSTM condition relative to the no-load condition. The observed exaggeration of the TAE under VSTM-load is the result of primary interest, and is discussed at length below. An analysis of the letter discrimination task confirmed that subjects were viewing the Sample Gabor equally under both conditions, ruling out the possibility that differences in response bias across conditions were due to nonequivalent adaptation. Accuracy on the letter discrimination task was 84.5% and 85.1% for the VSTM and no-load conditions, respectively.

4.3.1 Mixed-effects regression model

To assess statistical significance of these observed response patterns, we fit a family of logistic mixed-effects models to the binary response data. Linear mixed-effect models provide robust, unbiased fits that allow for simultaneous estimation of fixed (group) and random (individual subject) effects. Such analyses provide significant improvements over traditional linear regression methods that require either separate fits for each subject, or require averaging over individual differences, treating the data as a single, homogeneous group. We used the following best-fitting general linear mixed-effects logistic regression to model responses on the CCW/CW task. Note that several competing model specifications were tested and compared to our winning model, justifying its selection and the significance of the parameters (see Appendix B).
\[ \text{logit}(P(CW)) = \beta_0 + b_{i0} + (\beta_1 + b_{i1}) \cdot \theta + (\beta_2 + b_{i2}) \cdot c + \beta_3 \cdot \theta \cdot c + \epsilon_{ij} \quad (4.3.1) \]

\[ b_{i0} \sim \mathcal{N}(0, \sigma_0^2), b_{i1} \sim \mathcal{N}(0, \sigma_1^2), b_{i2} \sim \mathcal{N}(0, \sigma_2^2), \]

\[ b_{i3} \sim \mathcal{N}(0, \sigma_3^2), \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \quad (4.3.2) \]

The fixed (group averaged) effects include overall response bias \( \beta_0 \), change in response bias \( \beta_1 \) due to Sample orientation \( \theta \), change in response bias \( \beta_2 \) due to load condition \( c \), and change in response bias \( \beta_3 \) due to an interaction between Sample orientation \( \theta \) and load condition \( c \). Random effects include relative changes in bias for subject \( i \) due to inherent bias \( b_{i0} \), Sample orientation \( b_{i1} \), and experimental condition \( b_{i2} \). A random error component \( \epsilon_{ij} \) for subject \( i \) on trial \( j \) is assumed to be normally distributed and uncorrelated with the random effects. The value of \( \theta \) is continuous, ranging from approximately \(-15^\circ\) to \(+15^\circ\), and represents the possible orientations that the Sample Gabor can assume. The value of \( c \) is coded as either \( +0.273 \) for the VSTM condition or \( -0.273 \) for the no-load condition.

In the model each coefficient represents changes in the probability to respond clockwise due to inherent bias or one of several experimental effects: tilt aftereffect adaptation, VSTM interference, or an interaction between VSTM and adaptation (see Table 4.1). Individual subject random effects captured stimulus non-specific baseline variations in overall response bias. Such differences are expected, and were not significant at the group level \( (\beta_0 = 0.067, p = .70) \), suggesting that our calibration routine for defining “gravitational vertical” for each observer was successful. As our memory sample also acted as an adapting stimulus, we expected tilt aftereffects with and without working memory load. The significance of model coefficient \( \beta_1 = \)
<table>
<thead>
<tr>
<th>Response Change Due To</th>
<th>Coefficient</th>
<th>Estimate</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inherent bias</td>
<td>$\beta_0$</td>
<td>0.067</td>
<td>.70</td>
</tr>
<tr>
<td>TAE</td>
<td>$\beta_1$</td>
<td>-0.041</td>
<td>$3.1 \times 10^{-7}$</td>
</tr>
<tr>
<td>VSTM</td>
<td>$\beta_2$</td>
<td>-0.354</td>
<td>.13</td>
</tr>
<tr>
<td>VSTM/TAE interaction</td>
<td>$\beta_3$</td>
<td>-0.017</td>
<td>.002</td>
</tr>
</tbody>
</table>

Table 4.1: Summary of coefficient estimates from equation (4.3.1) and their conceptual descriptions for the best fitting model. The parameters of primary interest are $\beta_1$, which captures the influence of the tilt aftereffect (TAE) on response and $\beta_3$, which reflects the interaction between VSTM and the TAE. Inherent bias $\beta_0$ and VSTM $\beta_2$ were both expected to be non-significant because of a pre-experiment calibration routine and theoretical reasoning, respectively.

$-0.041 \ (p < 3.1 \times 10^{-7})$ reflects these adaptation-dependent changes in response frequency. Under our coding, the negative sign of the coefficient indicates a repulsive tilt aftereffect. That is, the vertically-aligned Target Gabor is perceived to be tilted away from the orientation of the Sample Gabor. The value of $\beta_1$ suggests that for each positive increment of 1° added to the adapting stimulus, the probability of responding “clockwise” is reduced by $\approx 0.009\%$. The effect of condition (VSTM versus no-load, independent of Sample orientation) was found to be not significant ($\beta_2 = -0.354, \ p = .13$). This result was anticipated for theoretical reasons, as VSTM load agnostic of memory contents is an ill-specified concept unlikely to manifest as meaningful variations in the data. Note that these results so far are not novel, but are a methodological requirement for the interpretation of the remaining analysis.

The parameter of theoretical interest in our model is the interaction term between VSTM and Sample orientation ($\beta_3 = -0.017, \ p = 0.002$). The significance of this value suggests that working memory load had a modulating effect on the perception of the Target stimulus, and the sign indicates that this effect of VSTM led to an apparent
strengthening or facilitation of the TAE. Specifically, concurrent TAE adaptation and VSTM load reduced the probability of responding “clockwise” by ≈ 0.011% for every degree of Sample orientation, an approximate 22% increase in response bias over and above the TAE. To verify the significance of this effect, we tested our result against a null model which lacked an interaction term, but was otherwise identical to our best-fitting model. A likelihood ratio comparison confirmed that the addition of the interaction term was indeed an improvement over the null model ($\chi^2(1) = 9.17, p = .002$).

At a first look these results confirm our hypothesis that the contents of visual short-term memory can influence external visual input during the retention interval. However, as discussed above, it is common for response time distributions on concurrent tasks to vary relative to no-load conditions. This often manifests as slower RTs under VSTM load. Such changes in response time can be symptomatic of underlying perceptual and/or decision related changes in information processing. Next, we establish why a joint analysis of bias/accuracy and response time distributions is necessary and confirm what previous studies of VSTM interference have not—that our effect is perceptual and not an artifact of criterion adjustment.

### 4.3.2 Diffusion Model & Data Simulations

To understand how our results (and related results from other VSTM-interference studies) are confounded with observer response characteristics at the previous level of analysis, it is useful to consider the pattern of results from within the theoretical framework of diffusion modeling. The Ratcliff diffusion model (DM; Ratcliff, 1978) is a stochastic, sequential-sampling model for explaining binary decisions (see Figure 4.3). The DM is characterized by seven parameters (see Ratcliff & Mckoon, 2008, for a review). In the model, evidence from perceptual input accumulates until the
one of two decision boundaries (or criteria) is reached. Response time distributions are generated from this noisy process and are affected by several parameters in the model. Rate of evidence accumulation is determined by the quality or strength of inputs, and is captured in the model by the drift rate $v$ (with across-trial variability $\eta$), and can be interpreted as an indication of evidence for a particular response. Speed-accuracy trade-offs are explained in the model by the boundary separation (or threshold) parameter $a$. A smaller value of $a$ leads to quicker responses that are more susceptible to the noise in the accumulation process. Higher thresholds lead to more accurate responses with slower response time distributions. Mean bias across trials is captured by the starting point of evidence accumulation $z$ with across-trial variation $s_z$. For an unbiased observer, $z = a/2$. Non-decision related components (e.g., response movement coordination and action, perception, etc.) are captured by a single parameter $T_{er}$ (with across-trial range $s_{t}$). All seven parameters are estimated simultaneously, and model convergence is jointly determined by accuracy and the full response time distribution (as opposed to a simple summary statistic such as median response time).

Consider now the DM parameters in the context of our experimental protocol. A Target Gabor presented at “gravitational vertical” $\delta^G_S$ (see method), preceded by a neutral Sample Gabor also presented at $\delta^G_S$ should appear vertically oriented to an observer (i.e., there should be no TAE). Under this presentation sequence, an unbiased observer completing our pseudo-discrimination task should have a drift rate $v \approx 0$. However, Sample Gabors in this study were presented at oblique orientations. As such, a CCW Sample orientation will cause the $\delta^G_S$ oriented Target Gabor to appear tilted CW (due to a TAE), leading to a positive drift rate in the diffusion model. Inversely, a CW Sample orientation will lead to a negative drift rate due to a CCW TAE.
Figure 4.3: In the top panel, responses in diffusion modeling are described as a result of stochastic evidence accumulation. Processes, of which several are depicted here, begin at the starting point $z$. Perceptual evidence accumulates in a noisy fashion over time, eventually pushing the process to one of two decision boundaries. Response times are reflected in the time taken to reach a decision boundary. The overall rate of evidence accumulation, whether positive or negative, is captured in the model by the drift rate $v$. Speed-accuracy trade-offs are addressed in the model by the boundary separation $a$. A smaller $a$ leads to quicker responses that are more error prone due to noise in the accumulation process. A larger $a$ leads to slower, but more accurate responses. Bias is captured in the model by $z$, wherein $z = a/2$ for an unbiased observer. The bottom panel depicts two competing hypotheses for explaining our data. Our hypothesis (H1) is that the contents of VSTM act as a concurrent stimulus, leading to an increase in the absolute value of the drift rate (dotted lines) relative to a no-load condition (solid lines). In an alternative hypothesis (H2), drift rate remains the same under both VSTM-load and no-load while the boundary separation increases under load.
Our motivating hypothesis for this study is that the contents of VSTM will have a modulatory effect on the perception of the Target Gabor, similar to the effects of adaption. If correct, this means that VSTM-manipulation will produce changes in the quality of perceptual input, captured by drift rate \( v \). However, drift rate is not the only parameter in the DM that can affect response frequencies. In fact, an alternative explanation is that drift rate remains constant between VSTM and no-load conditions while boundary separation \( a \) changes as a function of load. In the rare case that studies report response times, slower RTs on the order of several hundred milliseconds are typically recorded for VSTM versus no-load conditions. Such slowdowns can be represented in the DM by a widening of \( a \), leading to slower and more accurate responses. Assuming a constant TAE magnitude (and equivalently, a fixed \( v \)) across load-condition in our study, a larger \( a \) in the VSTM condition would lead to less noise-induced collisions with decision boundaries. This would manifest itself as more frequent CW responses for positive drift rates and more frequent CCW responses for negative drift rates even if \( v \) remained constant across load-condition. Behaviorally, the appearance would be that the TAE was strengthened when it had not, in fact, changed at all.

To demonstrate this possibility, we simulated two data sets, manipulating \( v \) in one case and \( a \) in the other. For both data sets, the remaining DM parameters were identical across all conditions \((T_{er} = 0.3, z = a/2, s_v = 0, s_z = 0, s_t = 0)\). In one simulation, only \( v \) was allowed to vary while \( a \) was fixed across all conditions. For the second simulation, \( a \) varied as a function of VSTM load \((a = 1.33\) for no-load and \(a = 2\) for VSTM-load, identical for all values of Sample orientation). As Figure 4.2 makes clear, both DM parameterizations are able to account for the behavioral data found in our study. It is clear that an analysis of response bias alone cannot tease apart the two competing explanations for our data, nor can an analysis of response
time alone. To determine the source of the observed increase in TAE magnitude, both response bias and the entire response time distribution must be jointly considered.

![Drift rate regression error bands](image)

Figure 4.4: Drift rate $v$ regression error bands from equation (4.3.3). The steeper slope under VSTM-load (light-gray) relative to no-load (dark-gray) reflects an increased tilt-aftereffect under VSTM-load relative to no-load. In terms of the diffusion model, this effect is primarily due to changes in drift rate (hypothesis 1 from figure 4.3).

For the purpose of addressing the confounding explanation from above, we estimated DM parameters using a hierarchical Bayesian diffusion model (Wiecki, Sofer, & Frank, 2013). Similar to our logistic regression analysis, this specification of the DM allowed for simultaneous estimation of group (of primary interest here) and individual subject parameters. Additionally, we were able to estimate the DM parameters as a function of a regression model similar to our logistic regression analysis. This method allowed us to measure the trial-by-trial effect of Sample orientation on the parameters of the model. Parameters of primary interest were $v$, $a$, $T_{er}$, and $z$. We
specified several linear models in combination for $v$, $a$, and $T_{er}$. For our full model for drift rate we used a coding scheme similar to our logistic regression model:

$$v = \beta_{v0} + \beta_{v1} \cdot \theta + \beta_{v2} \cdot c + \beta_{v3} \cdot \theta \cdot c + \epsilon_{vj},$$

$$j = 1, \ldots, n, \epsilon_{vj} \sim \mathcal{N}(0, \sigma^2), \quad (4.3.3)$$

where drift rate for the current trial $j$ is determined by inherent bias $\beta_{v0}$, Sample orientation $\beta_{v1}$, load-condition $\beta_{v2}$, an interaction between load-condition and Sample orientation $\beta_{v3}$, and noise in the accumulation process $\epsilon_{vj}$.

As explained above, it is likely that subjects adjust their boundary separation $a$ between load-conditions. It is unlikely, but also possible that our effect could be driven in part by VSTM-load dependent changes in non-decision related components $T_{er}$ (Rinkenauer, Osman, Ulrich, Müller-Gethmann, & Mattes, 2004), so we specify an additional regression for this possibility. However, there is little theoretical reason to expect that observers are capable of adjusting $a$ or $T_{er}$ as a function of Sample orientation. Taken together, we anticipate the following linear models for these parameters:

$$a = \beta_{a0} + \beta_{a2} \cdot c + \epsilon_{aj}, j = 1, \ldots, n, \epsilon_{aj} \sim \mathcal{N}(0, \sigma^2), \quad (4.3.4)$$

$$T_{er} = \beta_{t0} + \beta_{t2} \cdot c + \epsilon_{tj}, j = 1, \ldots, n, \epsilon_{tj} \sim \mathcal{N}(0, \sigma^2), \quad (4.3.5)$$

where the meaning and subscript numbering for the beta coefficients has been borrowed from equations (4.3.1) & (4.3.3) to facilitate comparison between models. A single posterior distribution for the bias parameter was estimated, as $z$ was assumed to be constant across conditions.

Table 4.2 highlights summary statistics for samples taken from the posteriors of $z$ and the coefficients from equations (4.3.3), (4.3.4), and (4.3.5). Several key results
Table 4.2: Mean, standard deviation, and quantiles for posterior distributions of diffusion model (DM) parameter $z$ and DM regression coefficients from equations (4.3.3), (4.3.4), and (4.3.5).

<table>
<thead>
<tr>
<th>Coefficient/Parameter</th>
<th>mean</th>
<th>std</th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{v_0}$</td>
<td>0.0877</td>
<td>0.1523</td>
<td>-0.2040</td>
<td>-0.0084</td>
<td>0.0805</td>
<td>0.1830</td>
<td>0.3919</td>
</tr>
<tr>
<td>$\beta_{v_1}$</td>
<td>-0.0259</td>
<td>0.0011</td>
<td>-0.0280</td>
<td>-0.0266</td>
<td>-0.0259</td>
<td>-0.0251</td>
<td>-0.0236</td>
</tr>
<tr>
<td>$\beta_{v_2}$</td>
<td>-0.2149</td>
<td>0.0538</td>
<td>-0.3193</td>
<td>-0.2477</td>
<td>-0.2149</td>
<td>-0.1788</td>
<td>-0.1121</td>
</tr>
<tr>
<td>$\beta_{v_3}$</td>
<td>-0.0152</td>
<td>0.0042</td>
<td>-0.0237</td>
<td>-0.0180</td>
<td>-0.0152</td>
<td>-0.0123</td>
<td>-0.0070</td>
</tr>
<tr>
<td>$\beta_{a_0}$</td>
<td>1.3250</td>
<td>0.0688</td>
<td>1.1927</td>
<td>1.2796</td>
<td>1.3235</td>
<td>1.3695</td>
<td>1.4711</td>
</tr>
<tr>
<td>$\beta_{a_2}$</td>
<td>0.0507</td>
<td>0.0200</td>
<td>0.0137</td>
<td>0.0369</td>
<td>0.0507</td>
<td>0.0634</td>
<td>0.0928</td>
</tr>
<tr>
<td>$\beta_{a_3}$</td>
<td>0.2529</td>
<td>0.0143</td>
<td>0.2245</td>
<td>0.2439</td>
<td>0.2526</td>
<td>0.2615</td>
<td>0.2837</td>
</tr>
<tr>
<td>$\beta_{t_0}$</td>
<td>0.0039</td>
<td>0.0026</td>
<td>-0.0010</td>
<td>0.0021</td>
<td>0.0039</td>
<td>0.0057</td>
<td>0.0093</td>
</tr>
<tr>
<td>$\beta_{t_3}$</td>
<td>0.4843</td>
<td>0.0068</td>
<td>0.4702</td>
<td>0.4800</td>
<td>0.4843</td>
<td>0.4891</td>
<td>0.4978</td>
</tr>
</tbody>
</table>

address our investigation of the alternative hypothesis that shifts in criterion across load-condition drive the behavioral results of our and previous studies. First, drift rate coefficients mirror our logistic regression results. Namely, $\beta_{v_1} = -0.0259$, 95% credible interval (CI) [-0.0280, -0.0251], indicates that we have a repulsive TAE. In DM terms, $v$ decreases as the Sample Gabor orientation becomes more clockwise, leading to less collisions with the “CW” response boundary. The interaction term $\beta_{v_3} = -0.0152$, 95% CI [-0.0237, -0.0123] once again strongly suggests that VSTM had a modulatory effect on the perception of the Target Gabor, varying systematically with the Sample Gabor orientation. Note that 0 lies outside of the 95% CI, further reinforcing the finding. Importantly, although similar in spirit to our logistic regression interaction analysis, this result is a marked improvement as the $\beta_{v_3}$ term reflects changes in the quality of perceptual evidence entering the decision process, as captured by the drift rate. By comparison, the previous logistic regression result was at the behavioral level, reflecting only a change in response frequency. The cause of the reported change was unclear prior to our DM analysis.

A second pattern of results further strengthens this conclusion. The coefficient
$\beta_{a0} = 1.3250$, 95% CI [1.1927, 1.3695], provides an estimate of the boundary separation $a$ for our no-load condition. Critically, $\beta_{a2} = 0.0507$, 95% CI [0.0137, 0.0634] suggests a small, but measurable increase in the boundary separation for the VSTM condition relative to the no-load condition, confirming that our alternative hypothesis is a potential confound for studies of this type. In our case, our experimental design minimized differences in RT across load-conditions, and the combined analysis confirms that the drift rate was affected by the contents of VSTM over and above the effects due to changes in boundary separation. Due to significantly larger differences in RT between load-condition in previous studies (or worse, unreported RT information), it is not clear whether previous effects were due to changes in response thresholds, changes in perceptual input to decision areas, or a combination of both. The analysis reported here provides a powerful method for teasing apart these confounding explanations for several reasons. First, the DDM provides more statistical power by measuring the full effect more directly, whereas regression-based techniques only gets at the lower bound of the effect captured by accuracy. In our case, the DDM extracted substantially more (39% versus the 22% estimated by logistic regression). Second, the DDM addresses the speed-accuracy trade-off, which is a particularly problematic confound in this area of research.

Finally, to further ensure the significance the interaction term $\beta_{v3}$, we estimated a second “null” model for comparison with our winning model. The null model had no interaction term, but otherwise shared the same coefficients and specification. Model comparison revealed our full model to be preferred over the null model (DIC values for the full and null models were 11922.09 and 11932.86, respectively). Such a comparison confirms that while boundary separation $a$ can account for some small amount of the effect, changes in drift rate due to VSTM-perception interactions primarily drive our results.
4.4 Discussion

The results of this study lend substantive support to the growing consensus that early sensory areas are involved in VSTM maintenance of previously encoded visual input. Observers completed a concurrent pseudo-discrimination task that directly measured response bias while either under VSTM load or no-load. We found this protocol to be very effective at eliciting several important changes to response characteristics. First, the orientation of the memory sample caused a reliable and consistent TAE in the opposing direction. This result is not novel, but was intentional in our design for the purpose of establishing baseline adaptation effects for our design. Second, VSTM load further affected response bias in a direction identical to TAE adaptation (a facilitatory interaction), leading to an approximately 22% increase in the strength of the TAE. Importantly, this effect was found by comparison to a no-load condition in which observers engaged in an otherwise identical pseudo-discrimination task. An attentional task ensured that observers visually engaged all stimuli regardless of experimental condition, and this control helped to rule out attention as the mechanism driving our results. Furthermore, work by Mendoza et al., 2011 strongly suggests that VSTM and attention separately influence perception, and in an additive manner.

This result is consistent with several recent studies that find interactions between VSTM content and external visual input (Saad & Silvanto, 2013a, 2013b; Scocchia et al., 2013; Kang et al., 2011; Mendoza et al., 2011), but extends previous research further by addressing a previous neglected confound in the interpretation of similar results. Specifically, we employed the use of hierarchical Bayesian drift diffusion modeling to demonstrate that (1) criterion shifts can create the illusion of systematic VSTM-perception interaction at the behavioral level found in previous studies, and (2) the results of our study were due to VSTM-induced perceptual changes over and above a small effect due to criterion adjustment.
Our findings are complimentary to several recent publications that ask similar questions. Of particular note is work by Scocchia et al., 2013 which found results similar to our own. In addition to finding a qualitatively similar pattern of results at the behavioral level, the authors provided an analysis of response time and accuracy data. Our work here extends this work in several critical ways. First, rather than measure the TAE indirectly through accuracy, our pseudo-discrimination task provided a direct measure of response bias as a function of VSTM load. Second, and more importantly, our joint analysis of bias and response time distributions via diffusion modeling addresses an alternative criterion-adjustment account capable of explaining their results. In fact, they report slower discrimination RTs under VSTM load relative to no-load ($\approx 753$ msec vs. $\approx 640$ msec across both experiments, respectively). In light of our Monte Carlo simulations above, it is clear that such RT differences can fully explain their accuracy effect without appealing to VSTM interference of a perceptual nature. It is also worth mentioning that despite their reported speedup on the second of two discriminations under VSTM load, both discriminations under VSTM load were clearly slower than their no-load counterparts, a pattern of results consistent with our alternative hypothesis. By contrast, our design resulted in small differences in RT between VSTM load and no-load trials. Further, our DDM analysis systematically reduced RT distributions to their constituent parts, allowing us to simultaneously account for the criterion-adjustment hypothesis while confirming that VSTM did, in fact, change the perception of observed stimuli.

The work described here bears on two other recent studies that asked a different, but related question. Both studies investigated the effect of VSTM on the TAE. The first (Saad & Silvanto, 2013a) utilized a sequential design wherein a standard adaptation-then-test TAE protocol directly followed a VSTM task. They found that VSTM strengthened the TAE when the memory cue and adapter had congruent
orientations, but reduced the TAE when they were incongruent. Critically, the two
tasks in their study did not overlap. Thus, although complementary to our findings,
their results are not directly comparable to ours. In a second study by the same
authors (Saad & Silvanto, 2013b), a concurrent design found that VSTM reduced
the TAE even when the memory sample and adapter had identical orientations (see
however, experiment 1c, which featured a protocol similar to our VSTM-condition
yielding TAE-inducing effects like our own). A critical difference between this design
and our own involved the adapting stimulus. In our case, the Sample Gabor served
dual roles as both memory cue and adapter, whereas adapter and cue were separate
stimuli in the aforementioned studies. Because the authors did not report response
times, it is difficult to reconcile these results with our current findings. What is clear
is that our results strongly suggest that the contents of VSTM act as a concurrent
stimulus, mildly affecting perception, but in a manner consistent with a physical
stimulus present in the visual field. Our pseudo-discrimination task occurred during
the VSTM retention interval, and so the effect of VSTM content was imposed on
the perception of the Target stimulus. By contrast, in the design of Saad & Silvanto,
2013b, the adapter was displayed during the retention interval, suggesting that VSTM
had a modulatory effect on the process of adaptation, which could account for the
reduction in TAE for VSTM-load versus no-load conditions.

In summary, these results are consistent with the sensory recruitment model of
VSTM. If stimulus information partially represented in perceptual areas of cortex
overlaps with the encoding of concurrent visual input, an interaction may alter the
contents of VSTM (i.e., memory masking), the perception of visual input, or a mixture
of both. For orientation information, this interaction can cause the contents of VSTM
to induce a TAE on subsequent stimuli. In this way, orientation information in VSTM
acts as a concurrent stimulus, even though it is no longer physically present.
CHAPTER 5

Practice-induced changes in bi-directional interference between visual short-term memory and externally-driven visual perception: asymmetric learning and transfer reveals multiple mechanisms

5.1 Introduction

To complete most tasks, including navigating the world around us in our daily lives, the visual system makes use of working memory that allows for the storage, retrieval, and manipulation of visual information (A. D. Baddeley & Hitch, 1974; Cowan, 1997; Logie, 2014; E. K. Miller, Erickson, & Desimone, 1996). Historically, the predominant view, known as the standard model of visual short-term memory (VSTM) (see Courtney, 2004, for a recent review) treated VSTM as a visuospatial sketchpad. This workspace, presumed to be mediated by areas such as the prefrontal and parietal cortex (Goldman-Rakic, 1987, 1990) outside of perceptual areas of the brain, has been predominantly treated as a protected buffer whose contents are only overwritten according to attention-driven task needs. This is in stark contrast to purely sensory representations that are immediately overwritten by any changes in visual input (Sperling, 1960; Phillips, 1974). The notion of a fully protected buffer was called into question with the discovery of so-called memory masking (see Magnussen & Greenlee, 1999; A. Baddeley, 2003, for reviews). Using inter-stimulus intervals
outside the range of sensory masking (Olzak & Thomas, 1986), Magnussen et al. (1991) discovered that stimuli presented during the memory retention interval could reduce subsequent judgments regarding the contents of VSTM. Additionally, memory masking was found to only be disruptive for relevant stimulus features. For example, if spatial frequency is the relevant dimension for discrimination judgments, a memory mask’s spatial frequency will interfere with a stimulus representation in VSTM, whereas the orientation of the masker will have no effect (Magnussen et al., 1991; Magnussen & Greenlee, 1992; Bennett & Cortese, 1996).

 Until recently, there was no reason to suspect an interaction between sensory representation and VSTM in the opposite direction to memory masking. Despite the critical role of VSTM in visuospatial tasks, it is reasonable to expect that transient memory stores should be subordinate to externally-driven sensory representations lest hallucination become a problem of everyday life. Evidence counter to this intuition has recently come to light. Early attempts at assessing VSTM’s effect on perception addressed the time course of processing visual items presented in the memory retention interval. Work in this area indeed found processing time to be affected by visual working memory (Downing, 2000; Moores & Maxwell, 2008; Pan & Soto, 2010; Soto, Heinke, Humphreys, & Blanco, 2005; Soto, Hodsoll, Rotshtein, & Humphreys, 2008; Turatto, Vescovi, & Valsecchi, 2008; Woodman & Luck, 2007). However, it was unclear whether observed processing delays were due to changes in the perception of visual input, or simply the result of cognitive load more broadly construed.

 More recent work has begun to identify changes in the perception of sensory information due to short-term memory representations. Kang et al. (2011) found that motion information in VSTM can produce a repulsive bias in the perception of visual motion in the same way that a physical stimulus can. Work by Mendoza et al. (2011) has ruled out the effect as purely attentional, finding that both task-driven
attention and working memory both contribute separately (and additively) to interference in the perception of visual motion. Similarly for motion’s static counterpart, orientation information in VSTM has been found to bias the perception of oriented gratings and generate tilt aftereffects (Saad & Silvanto, 2013b, 2013a; Scocchia et al., 2013). Our own work (Chapter 4) has confirmed that these orientation biases are due to perceptual effects, and are not the result of dual-task response time artifacts. Perceptual effects arising from the influence of VSTM have been identified in other visual domains as well. For example, Konstantinou et al. (2012) have found that visual detection sensitivity is reduced commensurate with a decrease in cortical response in early visual areas due to stimulus information held in short-term memory.

Complimenting these findings is a growing body of brain imaging and neurophysiological data supporting a close relationship between sensory processing and visual short-term memory. Converging work has found sustained neural activity in cells involved in the representation of sensory input during the memory retention interval in both monkeys (Chelazzi et al., 1993; E. K. Miller et al., 1993; Pasternak & Greenlee, 2005) and human observers (Ester et al., 2009; Harrison & Tong, 2009; Serences et al., 2009; Sligte et al., 2009; Konstantinou et al., 2012; Ester et al., 2013). Work in electrophysiological studies has identified event-related potential (ERP) markers during the maintenance of visual representations in VSTM known as “contralateral delay activity” (CDA). Further, support from transcranial magnetic stimulation research (Soto et al., 2012; van de Ven et al., 2012; Saad & Silvanto, 2013b; van de Ven & Sack, 2013) strongly implicates sensory areas of the brain in the short-term storage of visual information. While the presence of VSTM-related activity in sensory areas does not mean that there must be interactions between perception and memory in and of itself, it does complement existing studies finding such bi-directional interactions. Taken together, research on the neural locus of working memory and
the bidirectional interactions between sensory processing and memory has led to increased support for the “sensory recruitment” model of VSTM (Awh & Jonides, 2001; D’Esposito, 2007), which posits that working memory is an emergent property of the brain, arising from the collective action of goal-driven executive areas of the brain, and the sensory cortices responsible for encoding external input and their subsequent memory representations (see Postle, 2006, for an excellent review).

Although conceptually similar, it is unclear whether observed behavioral changes from memory masking, and the opposite effect—the influencing effects of VSTM on perception—arise from a common mechanism. To date, researchers have not asked this question, and existing behavioral data is incapable of providing an answer. Furthermore, although it is clear that mutual interference exists between the two systems, it is unknown whether the interactions operate as a fixed limit, or are attenuated by the effect of practice. The aim of the current study was to address these questions. To accomplish this, we implemented a novel experimental design to simultaneously test both forms of interference as a function of practice. Using a dual task design, we once again found strong evidence of bidirectional interference on a perceptual task (orientation discrimination) during the memory retention interval, as well as memory masking on a memory recall task. Training indeed reduced the effect of VSTM on perception. By contrast, however, practice had little to no effect on memory masking, revealing the dominance of visual input over transient memory stores. A novel model is proposed based on the framework of the selective reweighting model of A. A. Petrov et al. (2005) that explains our asymmetrical pattern of results.
5.2 Experiment

5.2.1 Method

Participants
20 naive university students participated in this study and were given an hourly monetary rate plus an additional bonus contingent on their accuracy to encourage motivated responses. All subjects had normal or corrected to normal vision.

Stimuli and Apparatus
The experimental display consisted of Gabors with a spatial frequency of 2.3 cycles per degree of visual angle presented against a uniform gray background. Gabor stimuli were blended smoothly into the background gray by means of a linearly-ramped circular aperture that transitioned from fully transparent to fully opaque from a radius range of $R = 5.1°–6.1°$. Three consecutive Gabor stimuli were presented on each trial left of center, center, and right of center, in that order. We label these stimuli as “Sample,” “Target,” and “Match,” respectively.

Stimulus orientation was the manipulation of interest for all Gabors. To minimize external orientation cues from the monitor and the experimental room, a black circular mask with an inner and outer diameter of 28.5 cm and 68.5 cm, respectively, hid all but the relevant area of the computer monitor at all times.

Images were created on an Apple iMac 2.66Hz Intel Core 2 Duo with an ATI Radeon HD2600 Pro graphics card using Matlab (The MathWorks, 2004) and the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 20” NEC AccuSync 120-BK CRT Monitor with a resolution of 1024x768, a refresh rate of 96 Hz, and a mean luminance of 16.6 cd/m$^2$. The monitor was calibrated before initial use using a Minolta 1° luminance meter, and the look-up table was adjusted to correct for the gamma function of the CRT. There were no other
Figure 5.1: The trial sequence for both experimental conditions involved the sequential presentation of three Gabors—the “Sample,” “Target,” and “Match” Gabors—in left, center, and right positions of the display, respectively. Observers completed two concurrent tasks. In an orientation recall task, participants directly rotated the Match Gabor to the remembered orientation of the Sample Gabor previously displayed. During the memory retention interval, observers classified the displayed Target Gabor orientation as clockwise or counter-clockwise relative to an implicit oblique reference boundary. Depending on the experimental condition, the Sample and Target Gabors were presented along a similar (congruent) or orthogonal (incongruent) orientation. Visual and auditory feedback was provided for both tasks. A black circular mask covering the monitor minimized orientation cues from the experimental apparatus.
light sources in the room, and observers’ heads were held fixed by a chin rest at a distance of 927 mm from the screen, causing one degree of visual angle to span $\approx 43$ pixels.

**Procedure**

On each trial, three Gabor stimuli were sequentially presented at screen positions left, center, and right. Observers were instructed to fixate each stimulus directly, and fixation dots preceded each stimulus location by 500 msec to facilitate controlled eye movements. Participants were instructed to complete two concurrent tasks, (1) orientation recall and (2) orientation discrimination. The orientation recall task used the first and third Gabors—the Sample and Match, respectively—and the discrimination task used only the Target Gabor, presented between the Sample and Match during the memory retention interval. Trial congruency was defined according to the similarity between the Sample and Target. For “congruent” trials, the average Sample orientation across trials was equal to the discrimination boundary of the Target Gabor. On “incongruent” trials, the mean Sample orientation was orthogonal to the Target discrimination boundary.

All participants completed a sequence of pretest, training, and then posttest, spread across 3 separate sessions. There were two experimental groups labeled as “train-congruent” and “train-incongruent” with 10 participants per group. For observers in the train-congruent group, Sample and Target orientations were congruent during the test phase and incongruent during pre- and posttest trials. Those in the train-incongruent group were administered the complimentary pattern of congruency. For both groups, congruency was switched by manipulating the orientation of the Sample Gabor only. That is, aside from small trial-to-trial deviations in the Target Gabor’s orientation around the discrimination boundary, the overall global
orientation of the Target never changed across the entire experiment. Trials were organized into blocks of 48 trials, with 3 blocks of pretest, 28 blocks of training, and 5 blocks of posttest, for a total of 1728 trials per participant.

The trial sequence was initiated by key press. Following initiation, the Sample was presented in a left screen position for 500 msec. Observers were instructed to memorize the orientation of the Gabor as precisely as possible. Sample orientations were drawn from one of two uniform distributions ranging from $-60^\circ$ to $-50^\circ$, or from $+30^\circ$ to $+40^\circ$ from vertical. Sample distributions were assigned according to an additional grouping to control for the effect of overall orientation (these conditions have no experimental value, but were included to counter-balance “left” versus “right” conditions). During pretest trials, observers received Sample orientations drawn from one of the two distributions (chosen according to the their group membership). During training sessions, Sample orientations were drawn from the other distribution. For posttest trials, observers once again received Sample orientations drawn from their pretest distributions. Stimulus offset was followed by a 2000 msec memory retention interval wherein the screen was a uniform gray color.

Next, the Target was presented at the center of the screen for 250 msec. Anytime after Target onset, observers had an unbounded response window in which to indicate by button press whether the orientation of the Target had been clockwise or counterclockwise of an implicit boundary orientation (the exact placement of which had to be learned with experience through feedback). The boundary orientation was fixed throughout the entire study on one of two values, $-55^\circ$ or $+35^\circ$ relative to vertical. Thus, for congruent trials, the reference boundary was equal to the center

---

1Both distributions were discrete uniform with a step size of $0.25^\circ$, the maximum resolution of our rotary input device.
of the uniform distribution from which the Sample orientation was drawn. For incongruent trials, the boundary was orthogonal with respect to the mean of the uniform distribution. Positive and negative feedback was provided visually—in the form of onscreen smiley and sad faces—and audibly as brief beeps for incorrect responses. Quick responses incurred a delay to ensure that the overall time between Target offset and Match onset was 2000 msec. Overly long responses could lead to > 2000 msec delays, which we wanted to avoid to ensure that the memory retention interval was identical across trials. To avoid such delays, participants were encouraged to respond as quickly as possible without sacrificing accuracy\(^2\).

Following response collection and any remaining delay on the discrimination task, the Match Gabor was presented in a right screen position (see Figure 5.1). The starting orientation of the Match was uniformly jittered (±15°) around either an orientation identical to (“congruent” trials) or around an orientation orthogonal to (“incongruent” trials) the boundary orientation of the Target discrimination task. The congruency of the Match followed the behavior of the Sample on each trial so that the Match’s starting position was in the same overall direction (clockwise versus counterclockwise, relative to vertical). Observers directly manipulated the orientation of the Match to the remembered orientation of the Sample. Continuous orientation responses were collected by using a custom made response knob that allowed for direct and intuitive control of the stimulus on screen. Response deadline on the recall task was unlimited. Visual feedback was provided contingent on response error (\(e\), the absolute difference between the response and the orientation of the Sample Gabor). Following response collection, a dashed black line was drawn directly on the Match showing the veridical orientation (i.e., the orientation of the Sample).

\(^2\)Our setup and instructions achieved this goal; mean and median response times across participants were 900 and 739 msec, respectively, well within our goal of < 2000 msec.
Superimposed on this was a colored line indicating the observer’s response orientation. This allowed for direct inspection of the recall error. Additionally, the color of the line was determined by the magnitude of the error. The color was green for \( e \leq 2^\circ \), yellow for \( 2^\circ < e \leq 4^\circ \), and red for \( e > 4^\circ \). Participants earned bonus points which converted to cash contingent upon this coloring scheme, with no points collected for “red” responses, an intermediate amount for “yellow” responses, and the most for “green” responses. This incentivized precision of recall responses.

**Adaptive Threshold Method**

For the Target Gabor, the difference in orientation between clockwise and counterclockwise stimuli \( \Delta \) was an independent variable controlled through adaptive thresholding. Across pretest, training, and posttest blocks thresholds were estimated for two difficulty conditions, “easy” and “hard,” which tracked target response accuracies of 84% and 69% correct, respectively. We used a single probit regression model (Equations (5.2.1) and (5.2.2)) to estimate orientation thresholds for both difficulty conditions.

\[
F(x|\alpha, \beta) = \int_{-\infty}^{\alpha+\beta x} \frac{1}{\sqrt{2\pi}} e^{-z^2} dz, \tag{5.2.1}
\]

which can be rewritten more simply as,

\[
Pr(correct) = \Phi(\alpha + \beta x), \tag{5.2.2}
\]

where \( \Phi \) is the normal cumulative distribution function, \( \alpha \) is the intercept, and \( \beta \) is the slope.

For each trial, the Target Gabor orientation was determined by first fitting Equation (5.2.2) to the response data from previous trials. The resulting psychometric function was then used to determine the corresponding orientation given the current trial’s difficulty level and target response (clockwise versus counterclockwise).
An even number of trials from each combination of difficulty and CW/CCW were randomly presented in each block.

The probit regression assumes that the fitted data come from a fixed source. That is, the estimation routine assumes it is tracking a relatively stationary response accuracy. However, learning was expected in our study, and thus the model was forced to track a moving target. To accommodate this additional complexity, we implemented a sliding window of past trials. For a given trial, the psychometric function for the current observer was estimated by using the previous 50 trials only. This ensured that the model was not burdened by (presumably poorer) performance from the distant past, allowing for threshold improvements as a function of practice. At the beginning of each experimental phase—pretest, training, and posttest—the sliding window was reset to remove trials from other conditions. To seed the model with trials for each sliding window, the beginning of each phase was populated with 1 block (48 trials) of a hardwired collection, in randomized order, of linearly-spaced orientations ranging from $\Delta = 0.5^\circ$ to $\Delta = 8^\circ$. From block 2 onward in each phase, orientations were fit adaptively as described above$^3$.

### 5.3 Results

The design of our study enabled the measurement of two dependent variables—orientation discrimination thresholds and orientation recall precision (VSTM fidelity)—on two separate tasks on each trial. We anticipated a primary effect of practice-induced learning on the discrimination thresholds; such improvements are the hallmark of perceptual learning studies (Fahle & Poggio, 2002). An additional result

---

$^3$On block 2 of each phase, the sliding window was populated with data from the 48 hardwired trials. Starting with block 2, the windows contents filled to 50 with the first two trials. Once full, the window maintained only the most recent 50 trials for the remainder of each phase.
of interest involved the effect of training on VSTM recall. It is less clear whether practice on orientation recall would lead to similar improvements in recall precision, as no clear consensus has been reached with respect to the effect of practice on visual short term memory (see introduction). Beyond the practice-related primary results, we were interested in any potential interactions between our two conditions. Our research (see Chapter 4), as well as previous studies (e.g., Saad & Silvanto, 2013b; Scocchia et al., 2013), suggests that the contents of VSTM representations can mutually interact with perceptual input. Because of this, we anticipated secondary effects due to interactivity between the two tasks. Sections 5.3.1 & 5.3.2 below address these interactions directly.

5.3.1 Discrimination Threshold Results

Figure 5.2 summarizes group-level discrimination thresholds across the entire study for all 20 participants. To generate each data point, psychometric functions were estimated for each observer on a block-by-block basis using the binary discrimination response data. Each data point represents the mean threshold value across all observers. Similar to the adaptive method used in the experimental setup, we used Equation (5.2.2) to estimate thresholds for “easy” and “hard” trials (84% and 69%, respectively). For analysis purposes, the size of blocks of experimental trials were chosen to satisfy two constraints. First, we wanted enough trials per block to provide reliable and valid estimates of the underlying psychometric function. Second, each block needed to be of sufficiently small size so as not to hide the effect of learning on discrimination thresholds. For training trials, our analysis block size was set to 96 trials, which provided stable estimates without reducing our resolution of the practice effect. For pretest and posttest trials, we were not interested in learning. Instead, the focus was on overall performance. Due to this, pre- and posttest trials were
Figure 5.2: Discrimination thresholds across all test and training phases of the experiment. Easy and hard trials have been separated here for clarity, but were randomly intermixed in the actual study. Solid symbols represent pretest and posttest blocks and open symbols reflect training blocks. Blue indicates congruent trials in which the Sample and Target Gabors shared similar orientations, while red indicates incongruent trials in which the Sample and Target were orthogonal. The overlaid curves are exponential fits from Equation (5.3.1), as summarized in Table 5.1. Symbol shape reflects participant group membership. For example, the first filled red circle in each panel reflects a block of pretest incongruent trials. The following series of open blue circles reflects the same group of observers during the training phase in which all trials were of the congruent type. Error bars are 1 standard error of the mean.
each collapsed into single blocks of 144 and 240 trials, respectively. The larger block sizes for pre- and posttest trials are justified as follows. Because we were interested in comparing *overall* performance on the posttest relative to pretest (learning was explicitly not a factor of interest during testing phases), it was more important to derive a highly reliable approximation of the overall thresholds for each phase, rather than any observed learning (we consider the pre- and posttest data more as point estimates of the underlying performance of each observer during these phases).

Learning was assumed to follow an exponential form according to existing published methods (Heathcote, Brown, & Mewhort, 2000). We fit the following modified exponential function, similar to the model used by A. A. Petrov and Hayes (2010), to the resulting group averaged threshold data for the training phase:

\[
\begin{align*}
    f_{m,n}(t; \Delta_{\text{start}}, \Delta_{\text{end}}, \tau) &= \Delta_{\text{start}} e^{-t/\tau} - e^{-n/\tau} + \Delta_{\text{end}} e^{-m/\tau} - e^{-n/\tau} \\
    &= \Delta_{\text{start}} e^{-t/\tau} - e^{-n/\tau} + \Delta_{\text{end}} e^{-m/\tau} - e^{-n/\tau}.
\end{align*}
\] (5.3.1)

The model estimates threshold orientation differences \(\Delta\) between clockwise and counterclockwise stimuli. The parameters \(\Delta_{\text{start}}\) and \(\Delta_{\text{end}}\) providing estimates of threshold discrimination levels for the reference periods *start* and *end*, which correspond to the beginning and end of the training phase. The parameter \(\tau\) controls the rate of the exponential law that interpolates the threshold \(f_{m,n}(t)\) at time \(t\) between the fixed start and end points of the function. This provides a reliable and theoretically justified method (Heathcote et al., 2000) for assessing the rate of learning, as well as performance levels at the beginning and end of training regimes. The full details and benefits of this modified exponential, which avoids extrapolations to infinity, can be found in A. A. Petrov and Hayes (2010).

Learning was observed in all conditions (see Figure 5.2). Observers’ discrimination thresholds improved for easy and difficult trials under both congruent and incongruent conditions. To facilitate comparison between performance on the different conditions depicted in Figure 5.2, we can use the model parameters from Equation
Table 5.1: Estimated parameters from Equation (5.3.1) fit to the orientation discrimination thresholds depicted in Figure 5.2. $\Delta_{\text{start}}$ reflects discrimination thresholds at the beginning of the training phase, $\Delta_{\text{end}}$ is an estimate of peak performance at the end of training, and $\tau$ represents the time course of learning.

<table>
<thead>
<tr>
<th>Condition</th>
<th>$\Delta_{\text{start}}$</th>
<th>$\Delta_{\text{end}}$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy congruent</td>
<td>14</td>
<td>7</td>
<td>0.6</td>
</tr>
<tr>
<td>Easy incongruent</td>
<td>18</td>
<td>7</td>
<td>1.6</td>
</tr>
<tr>
<td>Hard congruent</td>
<td>8</td>
<td>4</td>
<td>0.7</td>
</tr>
<tr>
<td>Hard incongruent</td>
<td>11</td>
<td>4</td>
<td>1.3</td>
</tr>
</tbody>
</table>

(5.3.1) as a common currency (see Table 5.1 for the full details). For easy trials, the average starting threshold was $\Delta_{\text{start}} \approx 16^\circ$. After practice, easy thresholds reached asymptotic levels around $\Delta_{\text{end}} \approx 7^\circ$. For hard trials, average initial thresholds were near $\Delta_{\text{start}} \approx 10^\circ$, with learned thresholds reaching asymptotic levels near $\Delta_{\text{end}} \approx 4^\circ$. Taken together, we observed an approximately 42% improvement in discrimination thresholds across all difficulty and congruency conditions.

Despite similarities in the overall learning profiles across conditions, we discovered several asymmetries when comparing across congruent and incongruent conditions. First, and most importantly, thresholds were worse (i.e., higher) under incongruent conditions relative to congruent trials. This was true for pretest-easy, where the mean threshold for the incongruent condition was $\Delta_{\text{pretest}} \approx 20^\circ$, compared to a congruent threshold of $\Delta_{\text{pretest}} \approx 15^\circ$. Performance on incongruent trials was also worse than congruent for pretest-hard ($\Delta_{\text{pretest}} \approx 10^\circ$ versus $\Delta_{\text{pretest}} \approx 8^\circ$, respectively).

Similar to pretest, we found thresholds were initially worse for incongruent-easy trials during the training phase, with $\Delta_{\text{start}} \approx 18^\circ$ for incongruent and $\Delta_{\text{start}} \approx 14^\circ$ for congruent-hard. The same pattern was found for hard trials, with $\Delta_{\text{start}} \approx 11^\circ$. 
for incongruent-hard and $\Delta_{\text{start}} \approx 8^\circ$ for congruent-hard. Despite initial differences in threshold across congruency conditions, practice on incongruent trials exhibited a faster time course of learning ($\tau \approx 1.6$ versus $\tau \approx 0.6$ for easy trials, and $\tau \approx 1.3$ versus $\tau \approx 0.7$ for hard trials). In fact, thresholds for incongruent stimuli reached equivalent asymptotic performance as congruent stimuli within $\approx 600$ trials ($600$ trials represents 45% of the entire training phase). Taken together, these results suggest that although incongruent stimuli are initially more difficult, the observed differences due to interference can be reduced entirely with practice. Furthermore, these asymmetries due to congruency effects support our hypothesis that the contents of VSTM can have systematic effects on the representation of visual input during the retention interval. It is important to note that even though incongruent stimuli lead to poorer discrimination performance, it is likely that congruent stimuli also interfere with visual input, relative to a baseline condition in which the VSTM system is not simultaneously taxed. For this reason it is not clear from our data whether practice can remove the interfering effect of VSTM on discrimination altogether. It is clear, however, that differences due to congruency are eliminated with practice.

To gain an understanding into the mechanisms driving practice-induced improvements in discrimination, it is common practice in the domain of perceptual learning to assess how well learned improvements transfer to related tasks. Typically, this is done through the same pretest, train, posttest regimen used in our study. Often, only a percentage of improvements will transfer, for example, to the same task carried out on stimuli displayed along an orientation orthogonal to the training direction. Such specificity of learning is often diagnostic to the underlying processes responsible (A. A. Petrov et al., 2011). We employed a similar tactic in this study. However, because the effect of congruency, and not the effect of perceptual learning per se, was the factor of interest, we manipulated the memory Sample’s orientation on pre-
Figure 5.3: Recall precision, defined as the reciprocal of the standard deviation of memory recall error, across all phases of the study. Symbol type and color follows the same conventions as Figure 5.2. Precision was found to improve for both congruency conditions as a result of practice. However, unlike observed changes in discrimination thresholds, the difference between congruency conditions remained fixed across training, strongly suggesting that the effects of memory masking are a persistent source of interference. Error bars are 1 standard error of the mean.

and posttest relative to training (leaving the discrimination Target Gabor reference boundary fixed across the entire study). This enabled us to restrict the interpretation of any specificity of learning to the effect of congruency, rather than to any effect due to the mechanisms of perceptual learning.

Despite reaching identical asymptotic performance during training, on posttest we observed incongruent thresholds much higher than those on congruent trials ($\Delta_{end} \approx 16$ for incongruent-easy versus $\Delta_{end} \approx 8$ for congruent-easy, and $\Delta_{end} \approx 9$ for incongruent-hard versus $\Delta_{end} \approx 4$ for congruent-hard). This pattern matched
that observed during pretest and at the beginning of the training phase. Importantly, transfer of learning from one congruency condition to another was contingent upon the conditions set during the training phase. That is, observers reached the same asymptotic performance during training regardless of congruency type, yet only those participants that trained on incongruent stimuli achieved similar performance on congruent stimuli at posttest. By contrast, despite reaching identical threshold performance during training, those observers trained on congruent stimuli exhibited a high degree of specificity of learning, as evidenced by their return to pretest performance levels after switching to incongruent stimuli on posttest. This transfer asymmetry suggests that the observed VSTM-perception interference patterns in this and previous studies is due in part to at least two different mechanisms. A theoretical explanation for this pattern of results is proposed later. First, however, it is important to ascertain whether the same asymmetries exist in our orientation recall results.

### 5.3.2 Orientation Recall Results

Figure 5.3 depicts the group-level orientation recall results for all 20 participants. The data shown illustrate orientation recall precision for pretest, training, and posttest across the entire study. We utilize the same blocking regime depicted in Figure 5.2 as described above. The dependent measure of interest here is recall precision. To calculate precision, we first determine the working memory error (i.e., the angular difference between the observer’s response and the orientation of the memory Sample Gabor) on each trial. With a single block of trials, precision is the reciprocal of the standard deviation of recall errors within the block:

\[ p = \frac{1}{\text{sd}(\text{error})}. \]  

(5.3.2)
Precision was calculated for each analysis block (1 pretest, 14 training blocks, and 1 posttest block) for each observer. The data points in Figure 5.3 represent the group-level averages of these points across all participants. As in Figure 5.2, we fit Equation (5.3.1) to the training data, assuming an exponential law of learning. Table 5.2 contains a numerical summary of the exponential fits. The data points in Figure 5.3 follow the same notation as in Figure 5.2. That is, circles represent observers from the train-congruent group and triangles represent observers from the train-incongruent group. Colors indicate congruency type, with blue indicating congruent conditions and red indicating incongruent trials.

<table>
<thead>
<tr>
<th>Condition</th>
<th>$p_{start}$</th>
<th>$p_{end}$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train congruent</td>
<td>4.4</td>
<td>5.6</td>
<td>8.3</td>
</tr>
<tr>
<td>Train incongruent</td>
<td>4.7</td>
<td>6.2</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Table 5.2: Visual short-term memory recall precision $p$ as estimated by Equation (5.3.1). $p_{start}$ represents initial recall precision at the beginning of the training phase, and $p_{end}$ is an estimate of the final recall precision at the end of training. $\tau$ reflects the learning rate for each exponential curve. Despite improvements in precision for both congruency conditions, the difference between conditions remains despite practice.

Three results are immediately clear in Figure 5.3. First, we see marked improvement in recall precision as a function of training. This result is true for both congruency conditions, and is novel in and of itself. The effect of practice on visual short-term memory has been unclear in previous research (Olson & Jiang, 2004; Chen, Eng, & Jiang, 2006; Sørensen & Kyllingsbæk, 2012), with few studies finding evidence of small improvements, and most finding no change at all. However, traditionally these studies were conducted using the change detection task. In such
studies, observers are typically shown an array of items, then, after a brief delay, are shown either the same array of items, or an array with one or more items changed along one or more dimensions of interest. For example, an observer could be briefly shown several colored patches. After a brief delay, an equal number of colored patches would be displayed again, any number of which could have changed color since the initial presentation. Generally, practice does not enable observers to achieve the same level of performance with larger array sizes. That is, practice on such tasks typically does not lead an observer to performance levels on 5 item arrays that they previously exhibited on 4 item arrays. This has led many researchers to conclude that VSTM cannot be improved with practice because the number of “items” in memory cannot be increased. However, our study utilized a continuous response collection regime, allowing for finer measurement of the underlying fidelity of memory stores. This allows for assessment of recall precision (see Equation (5.3.2)), a more sensitive measure. This added sensitivity revealed a strong learning component in our VSTM recall response data. Participants that trained on congruent stimuli had a change in memory precision of $p = 1.2$ units, and observers that trained on incongruent stimuli had precision improvements of $p = 1.5$ units, demonstrating similar practice-induced improvements in VSTM regardless of condition.

Second, despite recall improvements across all conditions, there is a persistent difference in recall precision between congruency conditions. At the beginning of the training phase, precision on train-congruent was $p = 4.4$, whereas precision on train-incongruent was $p = 4.7$. It is worth pointing out that this difference was expected due to known memory masking effects (Magnussen & Greenlee, 1992, 1999; Magnussen et al., 1991), in which stimulus properties (such as orientation, spatial frequency, etc.) similar to those of items held in memory interfere with an observer’s
ability to accurately recall information in VSTM. This initial difference helps to confirm that our experimental design generates our intended interference between VSTM and perceptual input. What was unknown, however, was what effect practice would have on the memory masking effect. In addition to greater precision, improvements in recall precision grew more quickly $\tau = 8.3$ for the Congruent condition versus $\tau = 1.9$ for the incongruent condition. This effect is likely due to the constraints imposed on recall due to memory masking. In the face of more interference in VSTM under congruent conditions due to memory masking effects, noise was increased in memory representations, slowing the rate of improvement. Further, at the end of the training phase precision on the congruent condition remained lower than the incongruent condition, with no evidence that the gap in memory fidelity had been reduced with practice ($p = 5.6$ and $p = 6.2$ for congruent and incongruent conditions, respectively). The observed fixed pattern of performance despite practice is in strong contrast to the observed improvements in discrimination thresholds, wherein differences due to congruency-based interference were eliminated with training. Again, we find another asymmetry indicating the presence of multiple mechanisms driving the complementary interference-patterns under investigation in this study.

Third, in addition to the striking differences in the learning profiles of recall precision compared to discrimination thresholds, we find another very important piece of evidence in our recall results. Across the entire training phase recall performance is worse for congruent stimuli. This is opposite to the results observed in our discrimination task, wherein performance was worse across all phases of the study for incongruent stimuli. It has been shown earlier in Chapter 4 that interference between the contents of VSTM and visual input is bidirectional. At that level of analysis, it was reasonable to conclude that observed interferences were driven by single underlying mechanism shared by both. However, in addition to the learning asymmetries
outlined above, the contrasting role of congruency in visual input-based tasks versus visual short-term memory tasks strongly indicates the influence of at least two mechanisms driving the apparently singular phenomenon of “interference.” Below we propose a theoretical explanation of how externally driven perceptual representations and internally managed memory representations might interact in ways that lead to two independent forms of interference.

5.4 Discussion and model

The results of this study provide novel insights into the nature of the underlying relationships between the representational systems of visual input as used in common perceptual discrimination tasks and those of visual short-term memory stores. Despite the interdependence of VSTM and many perceptual tasks (e.g., same-different, 2-alternative forced choice, and match to sample, which either implicitly or explicitly require short-term memory), qualitative and quantitative modeling accounts of perceptual processes (see e.g., Ahissar & Hochstein, 2004; Vaina et al., 1995; Teich & Qian, 2003; Law & Gold, 2009) and visual short-term memory processes (see e.g., Bowman & Wyble, 2007; Johnson, Spencer, & Schöner, 2009) are typically treated in complete isolation or, at best, by including a simplified caricature of the complementary system.

However, recent work regarding the nature of interactions between these two perceptual systems (see Chapter 4, as well as Saad & Silvanto, 2013b, 2013a; Scocchia et al., 2013; Kang et al., 2011; Mendoza et al., 2011; J. Huang & Sekuler, 2010) has illuminated the need for development of a more explicit model of the connectivity between externally driven visual representations and internally controlled visual memory stores. A successful candidate model would be constrained by converging behavioral patterns observed in perceptual discrimination and learning, as well as
data from visual short-term memory tasks. Ideally, incorporating an established model of perceptual processes would provide the added benefit of ensuring our data and theoretical position conform with known expectations of previous behavioral and modeling work. We believe the incremental, or selective reweighting model (SRM; A. A. Petrov et al., 2005) provides a suitable starting point for an explanation of the mechanisms driving previous findings of VSTM-perception interference, as well as the asymmetrical patterns of learning and interference observed in the current study.

The incremental reweighting model posits that improvements in perceptual discrimination occur due to small changes in the connections from early perceptual areas to decision-making areas of the brain. These changes to the connection weights are driven by Hebbian updating (Hebb, 1949) with and without external feedback. This is an important distinction from competing models of perceptual learning (i.e., the “representation modification” hypothesis family of models) in which the perceptual regions of the brain responsible for representing visual stimuli are assumed to change as a function of practice (see the reverse hierarchy theory of Ahissar & Hochstein, 1997, 2004, for a paradigmatic example). Under the representation modification hypothesis, learning leads to fundamental changes in the way task-relevant stimuli are encoded by the visual system. By contrast, the selective reweighting model assumes that perceptual representations are fixed. Changes (i.e., improvements typically) in response characteristics under the selective reweighting model arise from selective tuning of task relevant channels from perceptual areas to decision-making areas. Put simply, decision areas learn to strengthen outputs from perceptual channels that aid in task completion while simultaneously weakening perceptual outputs that do not contribute to successful response selection. Selective tuning of visual outputs is an important theoretical component of the selective reweighting model, and is important for our discussion below. Despite this, the mechanisms of the SRM are compatible
Figure 5.4: Schematic depicting the proposed selective reweighting memory (SR-Memory) model described in the text. The model’s task is to reconcile the asymmetric pattern of results observed in the discrimination and recall tasks. Despite the apparent difficulty of such an endeavor, we believe the model, which adds a bank of VSTM units and relatively assumption-free connectivity to the successful selective reweighting model of perceptual learning, can explain our results in a parsimonious manner.
with the representation modification hypothesis. It is conceivable for changes in stimulus encoding to occur as a result of practice, but the potential for this complimentary process does not detract from the theoretical arguments below.

Our results from the current study, as well as those described in Chapter 4, can be explained by combining the theoretical framework of the selective reweighting model with an additional memory mechanism. Figure 5.4 summarizes the overall architecture of the proposed model, referred to hereafter as the selective reweighting memory (SR-Memory) model. There are three major conceptual divisions in the model. First, we propose a perceptual bank of units (neurons) involved in the encoding of sensory input. That is, stimulus properties such as orientation are understood to be represented in a distributive fashion across a collection of orientation tuned units, each with a preferred orientation sensitivity.

Second, a decision area takes as input the activity levels of units in the perceptual bank. Here, for simplicity the decision area is reduced to a single decision unit that reduces the activity of the perceptual bank to a single binary (clockwise or counterclockwise) response. Modifiable connections enable full connectivity between each perceptual node and the decision unit. Weights control how influential each perceptual unit is in the decision process. Through discrimination practice, these weights are updated according to the Hebbian learning rule, leading to improvements in discriminability. The perceptual bank and the decision unit, along with the modifiable connection weights represent the existing selective reweighting model of A. A. Petrov et al., 2005.

Third, an additional bank of memory units serves the short-term memory requirements of the model. We assume that the perceptual and memory banks work in parallel via separate but interacting processes. We remain agnostic about the implementational details of the memory mechanism itself. Broadly speaking, there
have been two approaches to modeling visual working memory (see Postle, 2006, for a review)—the “standard model” (see Courtney, 2004, for a recent review) and the “sensory recruitment model” (e.g., Awh & Jonides, 2001; D’Esposito, 2007; Postle, 2006). Although our current results, as well as those discussed in Chapter 4, support the sensory recruitment model, it is important to note that our proposed mechanism works under either hypothesis. This flexibility in theoretical implementation is a strength of our proposed model, and allows us to move forward on discussions of the origins of interference between these two perceptual systems without having to settle this theoretical dispute. For our purposes, we assume that the memory bank contains a similar array of units tuned for the same stimulus properties as those involved in encoding visual input (in this case, orientation selectivity is the relevant dimension of interest). For ease of exposition, we assume that the memory bank units in Figure 5.4 are tuned to the same preferred orientations as their perceptual bank counterparts directly above in the figure. In this way, the far left unit in both banks has the same orientation tuning properties, and so on for each vertical pair of units. In reality, orientation tuning is likely far more distributed across units. Regardless of the physical layout of the connectivity, the logic still holds.

External visual input drives activity in the perceptual bank according to the orientation of the stimulus imposing itself on the model, leading to a particular pattern of activation for a given stimulus orientation. To store the representation in VSTM, attention opens a gating mechanism to allow the information to be “copied” into the memory bank. According to the standard model of VSTM, this involves storing the orientation information in the visuospatial sketchpad (presumably in a separate brain area). Under the sensory recruitment model of VSTM, we would assume that the memory bank involves similar/identical brain areas, with differences between the
memory and perceptual bank being purely functional (e.g., the memory and perceptual “banks” might be one and the same, with functional differences arising due to temporal asynchrony). Either way, the contents of perceptual bank selected for storage are assumed to be moved in parallel to the memory bank (see Shibuya & Bundesen, 1988). Once the perceptual information has been successfully “copied” to the memory bank, the attentional gate is closed, protecting the memory bank from further visual input during the memory retention interval.

In a perfect system, we might expect the information in the memory bank to be fully protected while the attentional gate is closed. However, we know from research on memory masking (Magnussen & Greenlee, 1992, 1999; Magnussen et al., 1991) that externally presented stimuli viewed during the memory retention interval degrades the fidelity of VSTM contents. Indeed, our own results (Figure 5.3) show strong evidence of such memory masking, demonstrating that the gating mechanism responsible for regulating what information makes it into VSTM is imperfect. We refer to this property as perceptual bleed. In short, perceptual bleed ensures that some small amount of information in the perceptual bank will leak into similarly-tuned units in the memory bank regardless of the observer’s goals.

In addition to both representational banks and the decision unit, the SR-Memory model has weighted connections from each unit in both banks to the decision unit. To accomplish our orientation discrimination task, we can imagine stimuli from the “clockwise” class as creating a pattern of activity on the perceptual bank, while “counterclockwise” stimuli produce a different distribution of activity (how different is determined by the mean difference in orientation between the two stimulus classes—Δ in our design). For a given stimulus, the job of the decision unit is to classify the pattern of activity in the perceptual bank as belonging to the clockwise or
counterclockwise distribution. Through training and iterative updating of the connection weights from each unit, the decision unit can improve its classification accuracy over time. Diagnostic units will have their weights strengthened, while ambiguous (or irrelevant) units will have their weights driven toward zero. This framework has proven to be a successful fit to perceptual discrimination data, and represents the core behavior of the selective reweighting model of perceptual learning (A. A. Petrov et al., 2005; Dosher & Lu, 1998, 1999; A. A. Petrov et al., 2006).

The SR-Memory model proposed here includes additional connectivity from each unit in the memory bank with the decision unit. Such connectivity would, for example, allow the decision unit to compare the orientation of an externally driven stimulus with a second stimulus stored in VSTM—a common exercise in psychophysics studies known as the same-different task. It is clear that the information in VSTM is phenomenologically weaker than externally driven visual input. Despite this, we contest that the nature of the connectivity of the memory bank to decision areas downstream remains functionally similar to the perceptual bank. This means that although the read-out connections to the decision unit from the memory bank are weaker, the decision unit treats the activity as if it has arisen from the perceptual bank. Thus, a poorly visible Gabor at a given orientation will drive activity in the decision unit in the same way that similarly oriented Gabor in VSTM will. Importantly, although the underlying architecture is assumed to be the same, training (as well as previous visual experience) changes the weights of connections, ensuring that memory and perceptual banks provide quite different inputs to the decision unit in practice.

To understand how the SR-Memory model predicts the results of this study, we can look first to the train-incongruent condition. At the beginning of the training phase, the Sample Gabor is viewed, stimulating activity in the perceptual bank according to the orientation of the Sample. This activity is immediately “copied” via
attentional gating into the memory bank for storage through the retention interval. During the retention interval, the observer is asked to view the Target Gabor and make a clockwise/counterclockwise discrimination judgment. The Target Gabor is presented at an orientation orthogonal to the Sample, leading to a pattern of activity in the perceptual bank represented by a collection of units largely unique from the units involved in the encoding of the memory Sample. The task of the decision unit is to classify the Target according to its orientation. However, activity from both the memory and perceptual banks feed into this decision process. Initially, the decision unit treats information from both banks equally. This, coupled with the large difference between activity levels leads to frequent response errors. With practice, the decision unit learns to identify units diagnostic to the discrimination task—in this case, specific units in the perceptual bank are helpful, whereas memory units encoding the Sample are not.

In addition to activity from the stored Gabor in VSTM, perceptual bleed from activity associated with the Target stimulus leaks into adjacent units in the memory bank (see Figure 5.4). Because perceptual bleed is uncontrolled, the leak is diffuse, leading to broad, non-specific activity in memory bank units tuned to similar orientations. Activity in the memory bank from this bleed also contributes to the decision process. Again, with practice the relatively unhelpful contribution from bleed-activated units becomes ignored by the decision unit.

For our train-congruent condition, the story is quite similar, but with an important difference. In this case, the memory Sample is on average rotated to the same orientation as the discrimination boundary of the Target Gabor (rather than orthogonal to it, as in the train-incongruent condition). Early in training, the decision unit again treats activity from both banks of units equally. In this case, however, the error caused by integrating over both banks is smaller because of greater agreement...
between banks. That is, although not perfectly correlated, the activity in the memory bank is closer to that of the perceptual bank on occasion. The finding that congruent conditions lead to better performance in all phases of our study is driven by this property of the model. Because the decision unit can quickly ignore unhelpful information from the memory bank, the performance differences due to congruency can be quickly eradicated, as seen by the quick and complete convergence of performance on our two conditions in the training phase.

At posttest we find that observers switching from train-congruent to test-incongruent perform much worse than those switching from train-incongruent to test-congruent. This is also explained quite easily by the model. Observers trained on incongruent stimuli have activity at multiple orientations along the memory bank. The orthogonally presented Sample and Target Gabors ensure activity on two largely non-overlapping populations of units. Additionally, perceptual bleed generates weak activity in the memory bank for units adjacent to the Target Gabor. Taken together, a large proportion of the memory bank units are active during the training phase, ensuring that the decision unit will drive their connection weights toward zero. By contrast, for train-congruent trials, the Sample Gabor activity falls in the same region in the memory bank as the activity generated by perceptual bleed. This activity is also ignored with practice. However, for train-congruent, units tuned to orthogonally-presented stimuli remain largely silent (i.e., inactive) throughout the training phase. This ensures that their connection weights remain relatively unchanged as a result of practice. During posttest when stimuli are now orthogonally presented, the decision unit is now faced with activity on memory units that were previously quiet. This leads to larger errors as the decision unit struggles to learn that this new source of information is non-diagnostic. Observers trained on incongruent stimuli do not have this problem. During posttest, stimuli are now congruent. This leads to activity in
the memory bank across units that have previously been activated due to perceptual bleed in the training phase. Because of this, the decision unit already “knows” to ignore activity from these units, leading to performance levels similar to the training phase.

Remarkably, despite its simplicity, our proposed model can also simultaneously account for the opposite pattern of results in our VSTM recall data. First, it was observed that recall precision was worse for congruent trials (opposite to what we observed for the discrimination task). This difference is due to memory masking. In the terms of our model, the attentional gate is never closed completely. In this way, perceptual bleed allows information from externally driven visual inputs to trickle into the memory bank. For incongruent stimuli, perceptual bleed generates weak, broad activity in units orthogonally tuned to the Sample Gabor already in memory. The weak bleed-driven activity does not directly interfere with contents of VSTM. During recall, both sources of activity likely contribute to the recall process, but presumably the contribution from the bleed-activated units is weaker. For congruent stimuli, perceptual bleed directly influences units already engaged in the encoding of the Sample Gabor in VSTM. As a result, memory contents are contaminated by visual input (i.e., memory masking occurs). Unlike in the discrimination results, the differences in recall precision due to congruency effects are not reduced by practice. The reason is that the effect in our discrimination results is driven by differences and changes in the weights-based contributions of representational units, which can be augmented with practice. By contrast, the difference in recall precision across congruency conditions is driven by a fixed property of the visual system—an imperfect gating mechanism allows visual input to bleed into the memory bank. Although this activity can be ignored with practice when it is not directly relevant for decision
making, the bleed-based activity itself, and its destructive effects on VSTM cannot be overcome.

In conclusion, our results add to the growing literature indicating the presence of interactions between perception and visual short-term memory. We have found evidence that although stimulus congruency can exacerbate such interference, practice can completely abolish this difference. By contrast, although recall precision can improve practice, the debilitating effects of memory masking are a persistent phenomenon unaffected by training. We propose the SR-Memory model as an elegant and parsimonious explanation for the results observed in this study, as well as similar research implicating interference between externally-driven visual input and the contents of visual short-term memory. Our framework builds upon an established model of perceptual learning, adding relatively few assumptions about the connectivity of visual processing areas with short-term memory mechanisms. Furthermore, our memory mechanism allows us to remain agnostic about the locus of visual working memory in the brain while still accounting for the pattern results observed in our experimental data.
CHAPTER 6

General discussion

Despite the long reign of the “standard model” of working memory, it has become increasingly clear from recent advances in monkey and human physiology that working memory is more distributed across the brain than previously believed. Aside from its shortcomings, the standard model has been a tool of theoretical convenience and inspiration that has enabled researchers to ask the right questions to move the field of working memory forward. Nonetheless, the continued adherence to the conceptually-convenient buffer metaphor (such as the visuospatial sketchpad) will provide diminishing returns given its tacit assumption of isolated and protected memory stores. The challenges described in this work require that we either update or replace the standard model with a modern framework that will allow us to continue to refine the investigative direction that visual short-term research takes into the future. The need for change is most evident when we break free from the virtual confines of intentionally narrow studies of VSTM (i.e., studies designed to inspect working memory in-and-of itself, in isolation with respect to other perceptual and cognitive processes), and instead look at the interplay between VSTM and the perceptual processes that make use of it.

Taken together, the collection of studies presented here provide a glimpse into one example of this interplay—between visual short-term memory and the perception of external stimuli—and provides new insights into the relationship. These results
demand that we reconsider how working memory is encoded, stored, and retrieved in the brain. The work here is consistent with the “sensory recruitment” model (also referred to as the “emergent-property” hypothesis), which posits that working memory is subserved by the same areas involved in sensory representation. That is, the same neural hardware responsible for encoding information selected for storage in working memory is subsequently used for retaining that very same information. It is important to note that our results can also be reconciled with the framework of the standard model. To accomplish this, our understanding of the standard model must be updated to reflect the bi-directional interactions of memory masking and VSTM-perception contamination. In fact, the apparent weakness of the standard model may actually reflect the age of its metaphors more so than an irreconcilable flaw in its framework. Implied notions of “moving” information into working memory may be a misnomer completely. Rather, by virtue of having been selected and sustained by attention, persistent activity in sensory areas “enters” working memory by virtue of this activity alone. The movement of information may likely be a consequence of the conceptual framing of the standard model—a terminological convenience that does not reflect the underlying neural processes of encoding, storage, and retrieval.

If sensory areas are managing memory representations while simultaneously encoding external inputs, one can imagine the potential for conditions of cross-talk between the two sources of activity. Indeed, it is difficult to interpret the disruptive effects of memory masking (Magnussen & Greenlee, 1999) as anything but evidence of this. Prior to this volume, it was unclear whether working memory could have similar, systematic influences on the perception of visual input.

Several previous studies had found evidence consistent with working memory contamination (cf. Saad & Silvanto, 2013b; Scocchia et al., 2013; Kang et al., 2011). However, a common methodological shortcoming with the few existing studies left
open a simpler, alternative explanation. Namely, an experimental condition requiring observers to complete tasks under working memory load was compared to a control condition with no memory component. During analysis, only response accuracy (or equivalent) on a second task (completed during the memory retention interval) was considered. However, response times for concurrent tasks are known to increase under VSTM load (Woodman et al., 2001). Behavioral response measures collected in these studies were highly susceptible to systematic deviations arising from changes in response time profiles. Critically, these deviations predict the same changes in response behavior as VSTM contamination, rendering the results inconclusive at best.

In Chapter 4 we addressed the ambiguity of previous research with the aid of a novel design and analysis capable of differentiating the effects of VSTM contamination from those due to changes in response time characteristics. This was accomplished primarily with the help of two methodological strategies.

First, we utilized the tilt aftereffect (see Schwartz, Hsu, & Dayan, 2007, for a review) as tool to probe the underlying representations involve in the encoding of our stimuli. Visual aftereffects can be an effective method to indirectly measure underlying visual processes. We first demonstrated the utility of aftereffects as experimental tools in Chapter 2 by using several varieties of the motion aftereffect to help locate the neural locus of practice-induced improvements in two separate perceptual learning experiments. Participants were exposed to two types of motion aftereffects (static and dynamic), and the effect of adaptation was measured both before and after extended training on a motion direction discrimination task. In both experiments their was no significant change in motion aftereffect duration—the dependent measure of interest. Critically, the stimuli for the motion aftereffect tests and the discrimination training were identical. Because the motion aftereffect is known to arise from adaptation in early sensory areas involved in encoding stimulus motion information, we were able
to determine that those sensory representations were not changed by practice, and that learning was likely facilitated by different neural mechanisms. Following on this success, we again used aftereffects—this time the tilt aftereffect—as a methodological device to gain insights into the possible existence of VSTM-based interference.

Second, our design allowed for the collection and analysis of a more comprehensive response profile. For each trial, we measured response bias directly (although bias was the behavioral measure of interest in previous studies, it could only previously be estimated indirectly) and collected response times. Furthermore, we jointly analyzed bias and the entire response time distribution simultaneously by using a novel application of the drift diffusion model. This allowed us to disambiguate the known effect of VSTM on response times from the previously unknown effect of VSTM on the perception of sensory inputs. By dealing with speed-accuracy trade-offs explicitly in our analysis we were able to demonstrate for the first time unequivocal evidence of VSTM-based contamination of perception.

Our success with the diffusion model analysis is further supported by the work presented in Chapter 3. In this collection of experiments, we pioneered the use of response time distribution analysis for studying perceptual learning, demonstrating the efficacy of diffusion modeling as a methodological tool for diagnosing the underlying mechanisms driving behavioral responses. Prior to this work, improvement associated with perceptual learning was treated as a monolithic phenomenon. Typically, discriminability, captured by $d'$, is the metric of interest. However, $d'$ alone only captures changes in discriminability broadly speaking, with no regard for what changed. Thus, by monitoring only $d'$ (or by similar logic, accuracy or response time) across training, it is unclear if improvements are actually due to changes in underlying perceptual processing. Indeed, by jointly considering response accuracy
and the entire response time distribution, we were able to demonstrate that practice-based improvements in motion direction discrimination were due to a combination of factors—increased quality of sensory input to decision processes, as well as better timing of internal gating between these systems. The diffusion model provides a means for isolating the contribution from different mechanisms, whereas univariate data mask this heterogeneity. We were able to leverage this advantage in two experiments to identify previously unknown perceptual learning components, and in an additional experiment to confirm the existence of VSTM-based interactions with visual perception.

Given the confirmation that VSTM can contaminate the perception of visual input, we endeavored to answer a series of related followup questions. First, a reasonable assumption is that VSTM-contamination and memory masking are two manifestations of the same neural mechanisms. Second, it is unclear whether these forms of interactions can be reduced through training, or if they represent a fixed limitation of the underlying neural “hardware.” In Chapter 5 we described the results of an experimental protocol designed to answer both questions. Observers were trained for an extended period of time on a dual-task paradigm which involved completing a perceptual judgment during a VSTM retention interval. We discovered two striking asymmetries. First, the effect of similarity between stimulus properties in VSTM and visual input had an opposite effect for VSTM-contamination versus memory masking (increased similarity led to more disruption under memory masking, but less disruption for VSTM-contamination of perception). Second, differences in the degree of interference due to similarity were completely eliminated for VSTM-contamination after practice, but remained at pre-training levels for memory masking.

Taken together, our results strongly indicated that at least two separate mechanisms were contributing to the observed interactions. A novel model was proposed
based upon an established framework from perceptual learning known as the selective reweighting model. In the model, learning is assumed to occur through the incremental adjustment of read-out connections from sensory areas to decision processes downstream. Although sensory representations are fixed throughout training, decision areas learn to emphasize task-relevant information channels and ignore non-diagnostic ones. Stimulus information selected for VSTM is stored in a secondary layer of representational units with similar tuning properties to their complementary perceptual units. Because of the fixed effects of memory masking, we assume that the attentional gating mechanism responsible for controlling what makes it into VSTM is imperfect, allowing a small, but constant amount of perceptual input to bleed into VSTM stores. Identical to the perceptual units in the selective reweighting model, memory units also have direct connections to decision processes. Similarly, these connection weights can be strengthened or weakened through practice, conditional on each connection’s relevance to task completion. With these simple, and empirically-justifiable assumptions, our model is able to account for both asymmetries without the need for unreasonable complexity.

The results of the present work lend several methodological and theoretical contributions to the fields of perceptual learning and visual short-term memory. Novel experimental protocols were developed and used across 5 experiments to generate multivariate data capable of providing new insights into existing theoretical problems, as well as first evidence toward several previously unknown properties of the visual system. The results of this present work are consistent with two competing possibilities—the sensory recruitment hypothesis, as well as an updated version of the standard model of working memory in which bi-directional activity leaks (perhaps through an imperfect gating mechanism) between the memory buffers of the PFC and connected sensory areas. This is an added strength of our proposed model,
as it does not rely on a full acceptance of either alternative hypothesis. Despite this theoretical flexibility, the model makes several strong predictions regarding the architecture of VSTM. In this regard, the work described here is also generative in nature, and will provide testable hypotheses for additional followup work in the future.

6.1 Future Directions

Presently, the current work could benefit from a number of advances in the understanding of the underlying mechanisms—both in terms of the interactions observed, and also the fundamental properties of working memory. Followup work planned for the immediate future will focus on filling remaining theoretical gaps and fleshing out the implementational details of the proposed selective reweighting model. For example, it is not entirely clear at this point how the working memory bank should be implemented. In its current specification, the memory bank is conceptualized as a separate population of units from the perceptual bank. This formulation serves both practical and theoretical purposes. First, from a modeling perspective, treating the memory bank as a separate component offers many conveniences. For example, both the perceptual and the memory bank are freely manipulable independent from the assumptions of their inter-connectivity in the model’s current form. This means that different hypotheses about the nature of each layer can be tested under fixed assumptions regarding the connectivity. Second, the current instantiation—in particular, the aforementioned independence of separate layers relative to their interactions—provides a useful theoretical flexibility. At this stage in our understanding of the interactions of perception and working memory, our model as proposed can provide a parsimonious account for the observed behavioral data without becoming entrenched in the debate over the locus of working memory. In this way, adherents to the standard
model and the emergent-property view can both benefit from any predictions generated by our framework and we are not inhibited from advancing our understanding of this relatively new phenomenon.

Despite this flexibility, there is self-evident merit in identifying the precise mechanisms behind working memory, and future work will focus on ultimately abandoning theoretical flexibility for neurophysiological specificity. Specifically, this will initially take the form of additional behavioral experiments. At the time of this writing, work is already underway on an experiment designed to identify the point of equivalence between external stimulus-driven noise and internal memory delay-driven noise. In the study, observers complete a mixture of two different tasks (only one task is completed on a given trial)—a same-different orientation discrimination or an orientation recall task. Each task is crossed with an additional condition of memory retention duration. An adaptive procedure adjusts external noise in the memory stimulus to match performance to a noise-free condition in which the memory retention delay is longer. Results from this study will be used to better inform the specification of our selective reweighting model, with the intended goal of developing a computational instantiation of the framework outlined in the present work. The final product will be able to accept image files as input, converting the stimuli into population coded representations in the perceptual bank. By way of the selective tuning of connection weights to the decision making unit, and by incorporating our collective findings into an explicit formulation of the working memory functionality, the model will be designed to emulate performance on the learning of orientation discrimination, as well as the bidirectional effects of VSTM-perception interactions. Finally, simulations will be conducted to generate further predictions regarding the role of working memory in perceptual learning. Ideally, this will lead to an iterative series of projects in which behavioral studies will be used to refine the model and its predictive power.
References


mechanism underlying the motion aftereffect. Vision Research, 32, 561–568.


Law, C.-T., & Gold, J. I. (2009). Reinforcement learning can account for associative


Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.


137


Seriès, P., Stocker, A. A., & Simoncelli, E. P. (2009). Is the homunculus “aware” of


Appendix A: Supplementary information for chapter 3

A.1 Additional details on the model fits

As stated in chapter 3, the diffusion model (DM) was fitted individually for each observer and each block. There were 297 fits (= 27 observers x 11 blocks, including the “mini-block” after the motion-aftereffect session). The parameter-search algorithm minimized the $\chi^2$ discrepancy between predicted and observed quartile RTs (Ratcliff & Tuerlinckx, 2002). The distribution of the 297 individual $\chi^2$ values had a mean of 23.6 (median = 20.2, standard deviation = 17.5, upper quartile = 27.9, max = 147). More than 69% of the values were less than the critical $\chi^2 = 25.0$ ($df = 15, p < .05$). That is, the model did not deviate significantly from the data in two thirds of the cases. The deviation in the remaining cases, though statistically significant, were small and unsystematic as evident from Figures 3.4 and 3.5. The deviating cases were spread approximately uniformly across blocks and participants. It should also be noted that for sample sizes such as ours, “$\chi^2$ often indicates significant deviations from a model, when in fact, none exist” (Van Zandt, 2000, p. 452). In conclusion, the diffusion model seems to track our data to a sufficiently good approximation so that the model parameters are valid measures of the underlying cognitive processes.

Figure A.1 plots the learning curves for two DM parameters: across-trial standard deviation $\eta$ in the mean drift rates (top panel) and the across-trial variability $s_z$ in starting point (bottom panel). The upward linear trend in $\eta$ was statistically
Figure A.1: Learning profiles for two diffusion-model parameters: across-trial variability in drift rate (A) and in starting point (B). The data format is as in Figure 3.3 in the main text. The observers practiced motion-direction discrimination for 8 blocks (black symbols) and then were tested on the same task at the orthogonal direction (open symbols). The error bars are 90% within-subject confidence intervals. Shaded areas mark two additional sessions of motion-aftereffect measurements.
significant but the effect size was small (Table 3.1: F(1,182)= 6.0, p < .02, \(\eta^2_p = .08\)). This increase in \(\eta\) probably stems from the large increase in the mean drift rates themselves (see Figure 3.3C in the main text).

There were no significant learning trends for the starting-point variability \(s_z\) (F(1,182) = 0.37 for the linear and 0.20 for the quadratic contrasts, respectively). This is consistent with the synchronization hypothesis and suggests that the diffusion process was rarely too early even during the very early blocks. In other words, the large variability in nondecision times \(s_t\) in Figure 3.3F apparently reflects instances in which the diffusion process started too late rather than too early. When the diffusion process starts too late, we expect a decrease in drift rates because useful sensory evidence is missed. By contrast, when the diffusion process starts too early, we expect an increase in starting-point variability because noise is accumulated in the absence of evidence. In other words, the synchronization hypothesis includes two distinct scenarios. In the first scenario, the initial blocks contain many trials on which the diffusion process starts too late. As these long nondecision times are shortened with practice, the drift rates improve, the nondecision variability \(s_t\) decreases, and the starting-point variability \(S_z\) stays the same. The available data seem to show this pattern. In the second scenario, the initial blocks contain many trials on which the diffusion process starts too early. If such short nondecision times were prolonged with practice, \(s_z\) would have increased. The apparent lack of such increase makes the second scenario unlikely.

A.2 Motion aftereffect

The experiment contained two additional sessions that measured the motion after-effect (MAE; Mather et al., 1998) in the trained, test, and two control conditions. The MAE data address a research question that is independent of our current focus
and are the subject of Chapter 2. The presence of these MAE sessions does not affect the interpretation of the discrimination data reported in the main text. This appendix outlines the MAE procedure and explains why the two data sets can be treated separately.

A.2.1 MAE Procedure

The MAE sessions occurred on two separate days marked by the shaded areas in Figure 3.3. There were 81 trials per session. Each trial began with a beep and consisted of three phases: adaptation, MAE measurement, and reset. The adaptation duration was 10 s, the MAE duration was the dependent variable, and the reset duration was adjusted to make the trial total exactly 30 s. The adapting stimuli were the same as in the discrimination sessions, except that they moved for 10 s rather than 397 ms. Then a (freshly generated) static\(^1\) texture filled the aperture. It appeared to move due to the aftereffect. The participants were asked to press a key when the apparent motion stopped. The reset phase presented random dynamic noise within the aperture for the remainder of the trial. It minimized any residual carry-over to the next trial (von Grunau, 2002). A red fixation dot was visible in the aperture center at all times. A simple attentional task controlled for fixation during the adaptation phase. The direction of motion changed by \(\pm 3^\circ\) at random intervals and the participants pressed a key whenever they detected a change. Several such changes occurred on a trial, and bonus points (visible onscreen) were scored for detecting them.

\(^1\)This measured the static MAE (Verstraten et al., 1999). For 11 of the 27 observers, the test stimulus on half the trials was dynamic noise that measured the dynamic MAE.
A.2.2 Discussion

It is extremely unlikely that the motion-aftereffect task can affect the interpretation of the discrimination data because the two tasks are very different. Moreover, our main finding—dissociable learning mechanisms—hinges on the concept of stimulus specificity, which in turn is defined with respect to two directions: training and test. All MAE blocks contained a counterbalanced, randomized mixture of both\(^2\). The MAE protocol was not contingent on what direction had been practiced during the discrimination sessions. Furthermore, the MAE data showed no statistically significant effects of the direction of the adapting stimulus.

The last experimental session began with a “mini-block” in the trained direction. It is shown in Figure 3.3 in the main text as an isolated point with the same symbol as the points on the main learning curve (blocks 1–8). The purpose of this mini-block was to absorb the disruptive influence of the MAE postest and transition back to the discrimination task. The small drop in all performance measures in Figure 3.3 is typical of the switch cost that occurs in perceptual discrimination upon any change of the experimental situation (A. A. Petrov et al., 2005). In this case, the switch cost reflects the task transition. Critically, the motion-direction switch in block 9 causes a further drop (particularly in the drift rates in Figure 3.3C). This further drop cannot be attributed to the MAE.

\(^2\)For 16 observers, the MAE was also measured in the two opposite directions.
Appendix B: Supplementary information for chapter 4

B.1 Evaluation of alternative mixed-effects regression models

In addition to the mixed-effects logistic regression model used in our analysis, three additional competing models were fit to the same data. Our simplest model had the same fixed effects ($\beta_0, \beta_1, \beta_2, \beta_3$) but assumed only individual differences for stimulus-independent response bias $b_0$. That is, the simple model was a random intercept model that assumed that observers were affected equivalently by our experimental stimuli, and that differences arose solely from differences in overall response bias. A likelihood ratio test found our winning model to be significant over the simple model ($\chi^2(3) = 640.29, p = 2.2 \times 10^{-16}$), suggesting that observers were differentially affected by our experimental manipulations. We chose the fuller model due to this significance, and because simulations by Barr, Levy, Scheepers, and Tilly (2013) and others have shown that models without random slopes have relatively high Type I error rates.

A second competing model included the same terms as the best-fitting model, but was estimated assuming an unstructured covariance matrix (as opposed to a diagonal covariance matrix, as used in the winning model), which additionally allows off-diagonal variances and covariances to be estimated, providing a more complete fit of the data (at the expense of more degrees of freedom). A comparison of this
model with our best fitting specification found that the added complexity did not significantly explain more of the variance in the data ($\chi^2(6) = 5.87, p = .44$), so the model was rejected.

A maximally complex model was also tested that included all parameters in the winning model plus an additional random effect of subject $b_{i3}$ for the interaction between VSTM and Sample orientation. A likelihood ratio comparison revealed this model to be marginally better than our winning model ($\chi^2(1) = 4.17, p = .04$). This result suggests that VSTM may interact with perceptual codes to varying extents across individuals. Given known individual differences in VSTM capacity and precision, it is tempting to speculate on theoretical and neurophysiological associations between VSTM ability and interference with online perception. However, given the marginal significance of the finding and the additional degree of freedom (9 versus 8 for the less complex model), we are reluctant to follow this line of reasoning. As such, we have elected to retain our best-fitting model over this competitor.

It is important to note that this decision does not affect our overall findings. Coefficients from both the winning and maximally complex models are identical to the third decimal place, and only $\beta_1$ and $\beta_3$ were significant in each model. In fact, an identical pattern of coefficient significance was observed for the entire family of models tested, indicating that our results are not an artifact of model selection.