Title
Understanding the Role of Top-Down Attentional Modulation in the Acquisition and Transfer of Perceptual Learning /

Permalink
https://escholarship.org/uc/item/9234w3f5

Author
Byers, Anna Catherine

Publication Date
2014

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Understanding the Role of Top-Down Attentional Modulation in the Acquisition and Transfer of Perceptual Learning

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

in

Psychology

by

Anna Catherine Byers

Committee in charge:

Professor John Serences, Chair
Professor Virgina de Sa
Professor Karen Dobkins
Professor Timothy Gentner
Professor Steven Hillyard
Professor Ayse Saygin

2014
The Dissertation of Anna Catherine Byers is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

Chair

University of California, San Diego

2014
TABLE OF CONTENTS

Signature Page................................................................. iii
Table of Contents.............................................................. iv
List of Figures........................................................................ v
List of Tables....................................................................... viii
Acknowledgements............................................................ ix
Vita....................................................................................... xi
Abstract of the Dissertation............................................... xii
Introduction......................................................................... 1
Chapter 1. Enhanced attentional gain as a mechanism for generalized perceptual in learning human visual cortex......................................................... 30
Chapter 2. Learning and Transfer of Feature-Based Attentional Modulation..... 64
Chapter 3. Enhanced Sensory Modulation in Human Early Visual Cortex Supports the Transfer of Perceptual Learning............................................. 77
General Discussion................................................................ 103
References........................................................................... 108
LIST OF FIGURES

Figure i.1: Reproduced with permission from Mukai et al., 2011 (their figure 5a). Regression slope analyses for accuracy in three attentional cue conditions:................................................................................................................. 25

Figure i.2: a) Reproduced with permission from Schoups et al., 2001 (their figure 3a). b) Reproduced with permission from Yang & Maunsell, 2004 (their figure 6).............................................................................................................. 26

Figure i.3: Reproduced with permission from Bao et al., 2010 (their figure 3). a) Time course of the VEPs for trained versus untrained orientation, before and after training. Normalized peak amplitude for the trained orientation versus untrained orientation at the trained location (b) and untrained location (c)........... 27

Figure i.4: Reproduced with permission from Gentner, 2007 (his figure 4b). Mean normalized response strength of single units in CMM to familiar versus unfamiliar songs for the two training regimes (2AFC and go-nogo)............... 28

Figure i.5: Reproduced with permission from Law & Gold, 2008 (their figures 3a and 4a). a) Average activity of MT neurons for various motion strengths as a function of viewing time. b) Average activity of LIP neurons for various motion strengths as a function of viewing time................................. 29

Figure 1.1: Schematic of the experimental paradigm. In both task types (attend orientation/attend-RSVP), subjects viewed a 2s presentation of a annular square-wave grating with a rapid serial visual presentation (RSVP) letter stream in the center, followed by a 400ms blank inter-stimulus-interval (ISI),.............. 57

Figure 1.2: Behavioral performance for training sessions outside of the scanner. a) Average orientation offset in degrees between the two gratings during each of the 10behavioral training sessions (at a fixed accuracy threshold, see panel B). Note that for one subject, the staircasing procedure did not converge,............. 58

Figure 1.3: Behavioral performance inside the scanner. a) Average orientation offset (at threshold) in degrees between the two gratings during the pre-training (dark gray) and post-training (light gray) scan sessions b) Accuracy in the attend-orientation condition was ~90% during both the pre- and post-training scan....... 59

Figure 1.4: Mean hemodynamic response functions (HRFs) for the attend-Orientation (black lines) and attend-RSVP (blue lines) conditions across training conditions (solid: pre-training, dashed: post-training). a) HRFs averaged across voxels in all three ROIs. b)-d) HRFs for each individual visual area (V1,............. 60

Figure 1.5: Orientation-selective tuning curves computed across V1-V3 and
Averaged across subjects, centered on the target orientation on a given trial and resulting parameter estimates from the best fitting von Mises function. a) Pre-training tuning curves during the attend-orientation (gray) and attend-RSVP.

Figure 1.6: Orientation-selective tuning curves for each individual visual area (V1, V2, V3) and averaged across subjects, centered on the orientation of the grating stimulus presented on each trial.

Figure 1.7: Parameter estimates from the best fitting von Mises function for each individual visual area (V1, V2, V3).

Figure 2.1: Schematic of the paradigm, displaying an example of each of the four cue types.

Figure 2.2: Fits of a Naka-Rushton function to the accuracy data from each of the 6 motion coherence levels in each cueing condition across training session.

Figure 2.3: Estimates of the C50 parameter across training session and cueing condition.

Figure 2.4: a) Learning for each cueing condition, quantified as the C50 threshold on day 1 minus C50 threshold on day 4.

Figure 3.1: Depiction of the stimulus design. Subjects fixated a central bull’s-eye and a red cue indicated which of the four gratings they should attend on that trial.

Figure 3.2: Orientation discrimination thresholds. a) Threshold offset for the Target quadrant (irrespective of trained quadrant) from the pre-test and post-test, respectively. b) Threshold offset for the trained quadrant when it was the target from the pre-test and post-test, respectively.

Figure 3.3: Mean hemodynamic response functions (HRFs) computed across V1-V3 for the target quadrant (blue) compared to non-target quadrants (green).

Figure 3.4: Mean hemodynamic response functions (HRFs) computed across V1-V3 for the trained quadrant when it was the target (purple) compared to non-target quadrants (green).

Figure 3.5: Orientation-selective response functions computed across V1-V3 and averaged across subjects, centered on the target orientation on a given trial. The left panel shows pre-training tuning for the target grating (blue) compared to the non-targets (green).

Figure 3.6: Orientation-selective response functions computed across V1-V3.
and averaged across subjects, centered on the target orientation on a given trial. The left panel shows pre-training tuning when the trained quadrant was the target (purple) compared to the non-targets (green).

Figure 3.7: Parameter estimates from the best fitting von Mises function to the orientation-selective response profiles for the target quadrant (blue bars), irrespective of which quadrant was trained, and the non-target quadrants (green bars), before (darker) and after (lighter) training.

Figure 3.8: Parameter estimates from the best fitting von Mises function to the orientation-selective response profiles for the trained quadrant as the target (purple bars) and the non-target quadrants (green bars), before (darker) and after (lighter) training.
LIST OF TABLES

Table 1.1: Statistical comparisons for each fit parameter across conditions and visual areas................................................................. 56
ACKNOWLEDGEMENTS

Completing a dissertation is no easy feat, and I could not have done it without the help and support of so many people. Most importantly, I would like to thank my advisor, John Serences, without whose unfaltering support I would never have made it this far. I have been incredibly lucky to work with such a brilliant, caring, and generally amazing PI. Not many people could handle my frequent outbursts of frustration with such poise and insight on how to manage what felt like insurmountable difficulties. Under Dr. Serences’ guidance and encouragement, I have learned more than I could ever have imagined in the past five years.

I would like to thank my other committee members, Tim Gentner, Karen Dobkins, Ayse Saygin, Virgina de Sa, and Steve Hillyard for their thoughtful comments on this dissertation.

I cannot forget to thank the raucous but incredibly intelligent members of the Perception and Cognition Lab, Miranda Scolari, Tiffany Ho, Tommy Sprague, Mary Smith, Sirawaj Itthipuripat, Eddie Ester, Javi Garcia, and Kim Kaye. Each and every one of them provided a unique style of support, without which I could never have completed a single project. Tiffany and Miranda, in particular, have been incredible role models since the moment I started in the lab. I could not have survived without them! I would also like to thank the innumerable undergraduate research assistants that are necessary to collect so much data for learning studies. I am forever in their debt.

Lastly, I would like to thank my amazing family and friends. My mom, dad, and sister have all helped me to get to this point and continue to motivate and inspire me.
Kristin Howell, Katherine Stavropoulos, and Stephanie Carmack are three of the best co-
graduate students anyone could ask for. Without their advice and moral support, these
past five years would have been a lot longer.

The Introduction, in part, is a reprint of the material as it appears in Exploring the
Relationship between Perceptual Learning and Top-Down Attentional Control in Vision
Research, 74, 30-39. Byers, A. & Serences, J.T. (2012). The dissertation author was the
primary author of this paper.

Chapter 1, in full, is a reprint of the material as it will appear in Enhanced
Attentional Gain as a Mechanism for Generalized Perceptual Learning in Human Visual
Cortex in Journal of Neurophysiology (in press). The dissertation author was the primary
author of this paper.
VITA

2009: Bachelor of Arts, The George Washington University, Washington, DC

2010: Master of Arts, University of California, San Diego; San Diego, CA

2014: Doctor of Philosophy, University of California, San Diego; San Diego, CA

PUBLICATIONS


ABSTRACT OF THE DISSERTATION

Understanding the Role of Top-Down Attentional Modulation in the Acquisition and Transfer of Perceptual Learning

by

Anna Catherine Byers

Doctor of Philosophy in Psychology

University of California, San Diego, 2014

Professor John Serences, Chair

Perceptual learning has classically been described as an improvement in sensitivity to a trained stimulus in a particular location in the visual field. However, the specificity of traditional perceptual learning makes it difficult to ascertain its relevance to real-world instances of perceptual expertise. Although more recent work has investigated the ability of perceptual learning to transfer across feature, task, and spatial location, the neural correlates this transfer still remain unclear. Furthermore, work regarding the nature of the relationship between perceptual learning and top-down attention suggests that there is substantial overlap between the mechanisms by which learning and attention modify perceptual sensitivity. Both perceptual learning and top-down attention have been shown
to improve the integrity of sensory responses through enhanced gain and/or noise reduction (*sensory modulation*) and improve the read-out of sensory signals by downstream decision mechanisms (*enhanced read-out*), and reduce intrinsic noise correlations in sensory areas (*noise reduction*). Given this overlap, I propose that more generalized perceptual learning results from the refinement of top-down attentional control, specifically feature-based attention, which subsequently promotes the transfer of learned perceptual improvements across spatial location. Here, I present both psychophysical (Chapter 2) and neuroimaging results (Chapters 1 and 3) supporting the role of feature-based attention in the acquisition and transfer of learning. Together, these studies suggest that learning to attend leads to an enhancement in the selectivity of feature-based attention via sensory modulation, which in turn supports improved behavioral performance at trained and untrained spatial locations.
1. Introduction

Top-down attention is a mechanism that enables the flexible modulation of neural activity so that behaviorally relevant stimuli can be processed more efficiently than competing distracters. In contrast, perceptual learning, or an improvement in sensitivity after repeated exposure to a stimulus (Gibson, 1963, 1969), typically induces a long-lasting and stable change that increases the efficiency of sensory processing for a highly specific stimulus.

Despite differences in their time-scale and flexibility, attention and perceptual learning (PL) are supported by similar neural mechanisms and appear to be intricately linked. For instance, both top-down attention and PL operate by increasing the signal-to-noise ratio (SNR) of sensory signals (Moran and Desimone, 1985; Zohary et al., 1994; Desimone and Duncan, 1995; Desimone, 1998; McAdams and Maunsell, 1999; Reynolds et al., 2000; Schoups et al., 2001; Furmanski et al., 2004; Martinez-Trujillo and Treue, 2004; Bao et al., 2010), by modulating intrinsic neural variability (Mitchell et al., 2007, 2009; Cohen and Maunsell, 2009, 2011; Adab and Vogels, 2011; Cohen and Kohn, 2011; Gu et al., 2011), and by optimizing the read-out of sensory information by decision mechanisms (Dosher and Lu, 1998, 1999, 2009; Palmer et al., 2000; Yu et al., 2004; Law and Gold, 2008, 2009; Palmer and Moore, 2009; Pestilli et al., 2011). In addition, attention and PL are known to be at least partially co-dependent: attention often acts as a modulator that enables the acquisition of learning-induced improvements in behavioral performance (Ahissar and Hochstein, 1993; Fahle, 2004, 2009) but does not always appear to be necessary for the expression of learning-related changes in neural activity.
after training (Furmanski et al., 2004; Bao et al., 2010; Hua et al., 2010; Adab and Vogels, 2011).

However, despite these links, many learning-induced changes in behavioral performance and in neural activity cannot be explained solely by top-down attentional factors (Watanabe et al., 2001; Yao and Dan, 2001; Seitz and Watanabe, 2003; Seitz et al., 2009; Frankó et al., 2010; Beste et al., 2011; McMahon and Leopold, 2012; see Seitz and Watanabe, 2009; Sasaki et al., 2010 for reviews). Furthermore, task demands and the characteristics of the trained stimulus set can influence both the generality of learning and the extent to which training can improve the efficiency of attentional control. The existing data present a complex picture that is by no means resolved; however, it now seems likely that are multiple routes by which learning can occur, and that these routes might operate in parallel during everyday perceptual experience to regulate the impact of potentially relevant stimuli on perception and decision-making.

2. Attention can Modulate Learning

Traditional studies of PL have focused on training-related improvements in the discrimination of basic visual features in a specific retinal location (Shiu and Pashler, 1992; Ahissar and Hochstein, 1993; Schoups et al., 1995; Yund and Efron, 1996; Fahle, 1997, 2004; Andersen et al., 2010). In these studies, attention is usually focused on the trained stimulus, and often interpreted as necessary in order to observe robust improvements in performance (Ahissar and Hochstein, 1993; Fahle, 1997, 2004). For example, when subjects attend to only one of two vernier stimuli during training, discrimination accuracy only improves for the attended stimulus but not the unattended
stimulus (Fahle, 2004). Similarly, accuracy on an orientation discrimination task improves more quickly when subjects attend to the location of the grating (Figure 1; Mukai et al., 2011), and changes in firing rates of sensory neurons that optimally discriminate a trained stimulus can rely on attention during training, and not only on repeated stimulus exposure (Figure 2; e.g., Schoups et al., 2001; Yang and Maunsell, 2004).

Additionally, top-down modulations have been shown to lead to learning even without the presence of a sensory signal (Dupuis-Roy and Gosselin, 2007; Tartaglia et al., 2009). For example, Tartaglia et al. (2009) demonstrated that when subjects were asked to repeatedly imagine performing a line bisection task in the absence of any visual input, their performance on a real line bisection task improved. This observation strongly suggests that top-down attentional influences are at least sufficient to observe PL.

Similarly, Dupuis-Roy and Gosselin (2007) used a “no-signal” task in which subjects chose which of two noise patches matched a target texture. Although there was no actual signal inserted into either noise patch, the experimenters defined the patch that happened (by chance) to correlate more with the target stimulus as the ‘target’. Despite the absence of strong bottom-up signals, subjects improved on this task with training, suggesting an influence of top-down feedback on the tuning of lower-level populations in a Hebbian-like fashion. Thus, these studies suggest that merely exerting top-down attentional control, even in the near or complete absence of a bottom-up signal, can induce learning-related modifications in behavior.

Learning-related gain modulations can also be context-dependent, such that responses to a trained stimulus can be enhanced or suppressed based on whether or not
the subject is performing a task with, and thus attending to, a given stimulus (Gilbert et al., 2000; Crist et al., 2001). Crist et al. (2001) had monkeys perform a task at fixation while previously trained stimuli were presented in flanking spatial positions. Cells that responded to the trained stimuli were suppressed, presumably because the trained stimuli were actively competing with the task-relevant stimuli being presented at fixation. However, responses in the same cells were enhanced in a second condition where the monkeys directed attention to the trained stimuli. These context-dependent effects indicate that the expression of learning-induced neural modulations can sometimes depend upon interactions between the bottom-up inputs and feedback connections from higher-level top-down control areas, which facilitate processing of a trained stimulus only when it is task-relevant (Gilbert et al., 2000). Furthermore, given that responses to trained stimuli were suppressed when they were presented as distracters in this paradigm, the context in which learning is evaluated also impacts the extent to which attention modulates the expression of learning.

Finally, the observation that learning can influence neural activity and connectivity across multiple levels of visual cortex and parietal and frontal areas (Gilbert et al., 2001; Li et al., 2004; Kourtzi and DiCarlo, 2006; Kourtzi, 2010) suggests that top-down attention plays a key role in coordinating task-related information from these areas during decision-making (Serences and Yantis, 2006; Kourtzi, 2010). This is particularly true in cases where learning is task-dependent (i.e., Crist, Li, & Gilbert, 2001; Li, Piëch, & Gilbert, 2004), as attention must be recruited to coordinate changes across sensory areas in a way that is sensitive to current task goals. Training with more complex objects may also require attentional coordination in order to efficiently integrate contour and
feature information into a size and position invariant representation of a whole object (Kourtzi & DiCarlo, 2006). Additionally, even though feedback is not strictly required to observe learning (McKee and Westheimer, 1978; Ball and Sekuler, 1987; Karni and Sagi, 1991; Shiu and Pashler, 1992; Fahle et al., 1995; Liu et al., 2012), it can play a facilitory role (Herzog and Fahle, 1997; Gilbert et al., 2001; Fahle, 2004, 2009; Seitz et al., 2006; Law and Gold, 2009; Shibata et al., 2009; Roelfsema et al., 2010; Sasaki et al., 2010), suggesting that top-down influences can be recruited throughout training to dynamically guide learning-induced changes in sensory processing or decision-making.

3. Learning Without Attention

In contrast to studies suggesting that PL depends on attentional processing, many studies have found that PL can be induced when a trained stimulus is ignored or even unseen (Watanabe et al., 2001, 2002; Yao and Dan, 2001; Seitz and Watanabe, 2003; Seitz et al., 2009; Frankó et al., 2010; Sasaki et al., 2010; Beste et al., 2011; McMahon and Leopold, 2012). Several of these studies induced changes in perception by manipulating the timing of stimulus presentations to exploit spike-timing dependent synaptic plasticity (Yao and Dan, 2001; Beste et al., 2011; McMahon and Leopold, 2012 see Seitz and Dinse, 2007 for a review). For example, Beste et al (2011) induced long-term potentiation (LTP) by having subjects passively view high frequency luminance changes. Following exposure to the high frequency stimulus, subjects were better at detecting luminance changes, demonstrating that improvements in a luminance change-detection task could occur even in the absence of a task during training. Moreover, low frequency luminance changes evoked long-term depression-like (LTD) effects and
performance was subsequently impaired. Similarly, when cats were repeatedly presented with a target grating that was always preceded by a task-irrelevant grating that was offset by 15°, the tuning curves of neurons that preferred the target were shifted in the direction of the distracter grating (Yao & Dan, 2001). This effect was dependent on the timing of the presentation of the preceding grating, such that changes were only induced in tuning when the two stimuli were presented within a 40 ms time window. This indicates that these changes depend on the relative timing of pre- and post-synaptic firing, such that pre-synaptic activity immediately preceding post-synaptic activity leads to synaptic potentiation. The authors also performed an analogous study in human subjects and found parallel timing-dependent changes in perceptual report following repeated exposure.

There is also evidence that PL can be observed for subthreshold stimuli, so long as they are temporally yoked to a task-relevant stimulus or to the delivery of a reward (Watanabe et al., 2001, 2002; Seitz and Watanabe, 2003; Seitz et al., 2009; Sasaki et al., 2010; Leclercq and Seitz, 2012). In one study, subjects were deprived of food and water before they passively viewed a sub-threshold grating. They then received a water droplet reward that was linked to the presentation of the sub-threshold grating rendered in a particular orientation. After training, subjects were better at detecting the orientation that was paired with the reward than orientations that were not paired with a reward, suggesting that the activation of subcortical reward pathways can automatically gate learning, even in the absence of overt attentional demands (Seitz and Watanabe, 2003; Seitz et al., 2009; Frankó et al., 2010; Roelfsema et al., 2010). However, other work indicates that this ‘task-irrelevant perceptual learning’ (TIPL) only occurs when there is no direct competition between target and distracter stimuli, presumably because attention
will suppress concurrently presented irrelevant stimuli if they actively interfere with target processing (Tsushima et al., 2008; Choi et al., 2009; Gál et al., 2009). Thus, it appears that TIPL may be a form of learning that acts as a complement to traditional attention-modulated learning, facilitating the processing of stimuli that are temporally associated with a behaviorally relevant target or the delivery of a reward (Seitz and Watanabe, 2005, 2009). However, even though TIPL does not depend on attention in many cases, it is still subject to attentional inhibition if the irrelevant item interferes with target processing (see Seitz and Wantanbe, 2009).

These interactions between TIPL and attention are further supported by more recent studies using ‘fast TIPL’, where subjects express learning effects for task irrelevant items within a single session (Leclercq and Seitz, 2011, 2012). In this paradigm, the subjects’ task was to identify rare targets embedded within a rapid serial visual presentation (RSVP) stream of distracters while they simultaneously viewed a series of images that they would need to recall at the end of each trial. Leclercq and Seitz found that subjects were better at remembering images that were presented at the same time as or immediately following the target, indicating that enhanced top-down attention due to target detection also led to better processing of the task-irrelevant image. In a second condition where the distracter stimuli were removed from the RSVP stream, the sudden onset of the rare targets exogenously captured attention. Under these conditions, memory for the task-irrelevant images suffered when they were presented at the same time as the target, presumably because the exogenous capture of attention by the target lead to a concurrent suppression of competing stimuli. Thus, these studies provide converging evidence that attentional factors can play a dominant regulatory role in
learning by enhancing relevant signals and suppressing signals associated with competing distracters.

4. Learning-Induced Neural Changes Can Persist without Attention

In many cases, top-down control appears to play a key role in the acquisition of learning, in part by increasing the impact of attentional control signals on sensory gain in early visual areas (Mukai et al., 2007). However, just as attention is not always required for the acquisition of learning (e.g., Beste et al., 2011; Seitz & Watanabe, 2003; Seitz, Kim, & Watanabe 2009), task-related attention is often not necessary to observe the expression of the learning-induced neural modifications after training, even for tasks where attention is recruited during training. Furmanski, Schluppeck, & Engel (2004) used functional magnetic resonance imaging (fMRI) to scan subjects before and after training on a task that required performing a discrimination on an obliquely oriented grating. Data from the pre-training scan session revealed an over-representation of cardinal compared to oblique angles in primary visual cortex. However, this ‘oblique effect’ was abolished after training, such that cardinal and oblique gratings evoked a similar response profile in primary visual cortex (V1). This response increase was limited to V1 and not seen in V2 or V3, which the authors interpreted as evidence against a role for attentional factors because the influence of top-down attention tends to be larger in later visual areas (Kastner et al., 1998). Additionally, if more top-down attentional gain was applied to oblique orientations compared to cardinal orientations then behavioral performance should differ; however, no such difference was observed. Bao and colleagues (2010) subsequently used EEG to demonstrate that training a specific orientation leads to an
increase in the amplitude of the temporally early C1 component of the visual evoked potential (onsetting ~50ms after the presentation of the trained stimulus). The increase in the C1 response was specific to both the location and the orientation of the trained grating, and occurred even when subjects were attending to a demanding RSVP task in the center of the screen (Figure 3; Bao et al., 2010). Thus, training-related changes in the amplitude of the C1 component could not be attributed to attentional factors because attention was focused on the center of the screen and not on the trained peripheral grating. Analogous findings have also been reported in single-unit physiology (Schoups et al., 2001; Raiguel et al., 2006; Vogels, 2010; Adab and Vogels, 2011). For example, Adab and Vogels (2011) found that training monkeys on an orientation discrimination task both increased the mean firing rate and decreased response variability of V4 neurons tuned to the trained orientation, an effect which was observed both during task performance and passive fixation.

More dramatically, perhaps, sensory neurons in animals often show learning-induced modulations of sensory responses even under anesthesia, strongly suggesting that top-down attention is not required for the expression of all types of learning-induced neural modulations. For instance, training cats to discriminate a particular spatial frequency with one eye led to a significant improvement in neural contrast sensitivity functions recorded in V1 under anesthesia (Hua et al., 2010). These training effects were highly specific: the contrast sensitivity of V1 neurons was not only greater in trained cats compared to untrained controls, but greater for neurons receiving input from the trained eye compared to the untrained eye. Likewise, after training on a shape discrimination task, a greater proportion of neurons in the inferotemporal cortex of anesthetized
monkeys responded to a set of trained shapes compared to control animals that had no prior experience with the same set of shapes (Kobatake et al., 1998). Finally, European starlings can be trained to recognize and discriminate many different motifs and this ability can be used to index plasticity in the auditory system (see Knudsen and Gentner, 2010, for a review). Akin to visual perceptual learning, starlings improve at discriminating complex songs with practice and these behavioral improvements correspond to changes in the firing rates of neurons in the caudomedial mesopalluim (CMM) and caudolateral mesopallium (CLM), two regions of secondary auditory cortex (Gentner and Margoliash, 2003; Jeanne et al., 2011). Under anesthesia, neurons respond more strongly to trained songs compared to untrained songs, and this effect is mediated by the training regime (Figure 4; Gentner and Margoliash, 2003; Gentner, 2007). In “go-nogo” trained birds, the responses to songs associated with positive reinforcement were higher than those to songs associated with no reinforcement. The selective learning of task-relevant songs is consistent with the notion that feedback driven attentional mechanisms may play a role in modulating the acquisition of learning, even though long lasting effects of learning can be instantiated in local circuits without top-down control (Crist et al., 2001; Gilbert et al., 2001; Gentner, 2007; Bao et al., 2010).

Collectively, these studies suggest that even though attention can facilitate the instantiation of PL during training, active attention is not strictly necessary for the expression of training-related changes in neural activity. However, it is important to note that even if top-down attention is not necessary to observe the effects of PL, attentional factors can still play an important role in modulating the expression of training-induced neural modulations, depending on specific task demands. Recall that in the study by
Crist, Li, and Gilbert (2001), responses in neurons tuned to a trained peripheral target were enhanced when the peripheral stimulus was attended and suppressed when the monkeys performed an alternate task at fixation. This suggests that, all else being equal, top-down attentional control can override or even reverse the effects of PL on neural activity depending on task demands. Thus, trained stimuli that are task-relevant or completely task-irrelevant can evoke enhanced responses after PL, even when they are not directly attended (Adab & Vogels, 2011; Bao et al., 2010; Schoups et al., 2001; Seitz, Kim, & Watanabe, 2009; Seitz & Watanabe, 2003; 2009). In contrast, top-down attention can override the effects of PL and suppress neural activity when trained stimuli actively compete with another relevant stimulus (as in Crist, Li, & Gilbert, 2001; Li, Piëch, & Gilbert, 2004).

5. Can Training Improve the Efficiency of Attentional Modulations?

Classic PL studies focus on training a specific stimulus feature in a specific retinal location. However, this specificity can limit the applicability of learning studies to real-world cases of perceptual expertise. For example, when a radiologist examines an x-ray, she cannot expect a fracture to always be in the same location and at the same orientation; she must learn a broad strategy for detecting fine discrepancies across the entire image. Therefore, learning often requires acquiring a general strategy so that the neural modulations that support more efficient perception can be flexibly applied to a large set of potentially relevant sensory features. Allocating attention in response to dynamically changing task demands can flexibly modulate neuronal responses by enhancing the gain in early sensory populations (McAdams and Maunsell, 1999; Treue and Martínez
Trujillo, 1999; Reynolds et al., 2000; Saenz et al., 2002; Treue, 2003; Martinez-Trujillo and Treue, 2004; Reynolds and Chelazzi, 2004; Serences and Yantis, 2006; Serences et al., 2009; Scolari et al., 2012), by improving the read-out of sensory information in later decision-making areas (Shaw, 1984; Eckstein et al., 2000; Palmer et al., 2000; Palmer and Moore, 2009), and by mediating intrinsic neuronal noise (Mitchell et al., 2007, 2009; Cohen and Maunsell, 2009, 2011; Cohen and Kohn, 2011). If the efficiency of these attentional mechanisms increases with training, then whole classes of potentially relevant stimuli could be more readily distinguished from irrelevant distracters.

Relatively few studies have examined the impact of training on the ability to flexibly deploy top-down attentional gain in the context of complex visual search tasks. Rettenbach and colleagues used a variety of visual search tasks that all recruited feature-based attention to a pre-specified target stimulus (Sireteanu and Rettenbach, 1995, 2000; Leonards et al., 2002). Even though the search displays varied in their exact composition and spatial distribution from trial-to-trial, practice led to a significant behavioral improvement, such that reaction times were eventually unaffected by increases in the size of the search array (Sireteanu & Rettenbach, 1995). The authors argued that training-related reductions in reaction time indicate a shift from inefficient to efficient search by eliminating the need to examine each object in the array serially until the target is detected (Sireteanu & Rettenbach, 2000). Training also transferred across task type, eyes, and spatial locations, which argues against a low-level locus of the training effects. Instead, the high degree of transfer suggests that subjects were learning a general attentional strategy that allowed them to more efficiently search the array for the target stimulus (Sireteanu & Rettenbach, 2000). In follow-up studies, these improvements were
linked to an enhanced ability to detect both brightness differences between the target and distracters, as well as a unique visual cue associated with the target (Leonards et al., 2002).

In a related line of work, Ellison and Walsh (1998) found that training on orientation, size, and color singleton search tasks transferred partially to conjunction tasks that combined these features, whereas training on a conjunction task asymmetrically transferred to several different singleton search tasks. In this case, the ability of a search strategy to transfer depends on the complexity of the training task: if subjects were able to more efficiently find a target defined by a conjunction of features, then this ability may more easily transfer to a condition in which there was only one target feature (as opposed to the other way around, see Ahissar and Hochstein, 1997, 2004; Ahissar et al., 2009). This relationship between the broadness of the trained task and the extent of transfer could account for the specificity of many traditional PL tasks, as they often employ a single stimulus feature. By extension, if subjects are trained on a more generalized task, then learning may influence the efficacy of flexible top-down attentional mechanisms and the benefits would extend beyond the specific stimulus exemplars that were used during training (Ahissar et al., 2009).

Video games have also been used as a form of generalized learning that improves perception and enhances the efficiency of attention (Green and Bavelier, 2003, 2012; Riesenhuber, 2004; Dye et al., 2009; Li et al., 2009; Bavelier et al., 2010; Green et al., 2010; Clark et al., 2011). Video game players are better at a variety of behavioral measures, including difficult flanker compatibility tasks, subitizing items in a briefly flashed display, distributing attention across wide eccentricities, overcoming the
attentional blink effect, and detecting small changes in a visual array (Green and Bavelier, 2003; Clark et al., 2011). Non-video game players who were trained on a video game also showed post-training improvements in contrast sensitivity (Green and Bavelier, 2003; Li et al., 2009). The improvements among video game players on a wide variety of tasks suggest that training can increase the efficacy of attentional modulations in a task-general manner, enabling improvements in many different task settings.

However, even though Li and colleagues (2009) found an improvement in contrast sensitivity with controlled training using video games – which is consistent with a relatively low-level change in sensory gain – it is unclear whether these improvements are tied to enhanced sensory representations, more efficient read-out of sensory responses during decision making, or more efficient distracter exclusion.

While the results of studies that employ complex stimuli support the notion that training increases the efficiency of attentional modulations in a fairly general manner that supports extensive transfer, recent data suggest that general improvements in task performance can also be observed even when training is carried out using a restricted stimulus set. For example, double-training paradigms reveal transfer to specific features or locations that are passively exposed while training is carried out with a separate task-relevant stimulus (Xiao et al., 2008; Zhang et al., 2010a, 2010b). For instance, the mere presence of a stimulus at the to-be-transferred location during training may enable the global transfer of feature-based learning to the exposed location (Xiao et al., 2008; see also Treue & Martinez-Trujillo, 1999; Martinez-Trujillo & Treue, 2004; Saenz et al., 2003; Serences & Boynton, 2007). Similarly, a brief pre-training exposure to the peripheral transfer location may prime this location, again enabling the transfer of
learning (Zhang et al., 2010c). Lu, Liu, and Dosher (2010) further demonstrated that learning in an orientation discrimination task also transfers across external noise conditions, but only from low external noise to high external noise. Guided by the Augmented Hebbian Reweighting Model (AHRM), the authors argued that read-out mechanisms are re-tuned to more selectively pool information about relevant input signals under low noise conditions. After optimization, the re-tuned read-out mechanisms can then be utilized to enhance the stimulus signal even when viewing a stimulus that is corrupted by a high level of external noise. In addition, training with a highly discriminable motion stimulus supports transfer across directions (Liu, 1999), possibly because low difficulty tasks can be carried out based on signals in higher cortical areas, which in turn leads to a higher degree of transfer after training (see Ahissar and Hochstein, 1997, 2004; Ahissar et al., 2009). Thus, the observation of extensive transfer even with tasks that use fairly basic visual features suggests that training may increase the efficiency of top-down attentional modulations in a manner that can generalize beyond specific low-level features, echoing the general types of learning observed with more complex stimuli, such as video games.

Collectively, the existing data suggest that, just as attention plays a modulatory role in the acquisition and expression of PL, training plays a complementary role in increasing the efficiency of top-down attentional control. In line with this notion, several studies have now demonstrated a reduction in the magnitude of modulations in areas of frontoparietal cortex that are commonly thought to mediate attentional control (Sigman et al., 2005; Mukai et al., 2007), even while activation in sensory areas simultaneously increases. This suggests that active attentional control is heavily recruited before training,
but that learning either improves the efficiency of top-down attentional gain or induces plasticity in local connections so that intervention by top-down attentional modulations can be minimized. In line with these general hypotheses, Sigman et al. (2005) demonstrated that the blood oxygenation level dependent (BOLD) response in posterior parietal cortex (PPC) and supplementary motor area (SMA) was significantly diminished after training on a shape discrimination task. Another study found that successful learners had strong BOLD responses in putative attention-control areas before training, including intraparietal sulcus (IPS), frontal eye fields (FEF), and supplementary eye fields (SEF). However, activation in all of these areas was reduced after training (Mukai et al., 2007). This decrease in activation was coupled with an increase in the functional connectivity between frontoparietal areas and areas of early visual cortex, which tentatively suggests that the efficacy of attentional control signals improved with training and that the associated behavioral improvements were not due solely to local plasticity in visual cortex. Thus, attentional control signals become more efficient with training and may have an equivalent impact on sensory neurons despite giving rise to a smaller metabolic trace and thus smaller BOLD responses. Since connections between putative control regions and sensory areas are known to mediate top-down attention signals (Moore and Armstrong, 2003; Moore and Fallah, 2004; Armstrong et al., 2006; Moore, 2006; Noudoost et al., 2010), the reinforcement of these pathways may allow faster and more efficient gain modulations after training. Given that the capacity to attend to multiple items is limited (Mullin and Egeth, 1989; Huang et al., 2004; Huang and Pashler, 2005; Scharff et al., 2011a, 2011b), improving the efficiency of attention through PL would
allow these limited attentional resources to be preferentially focused on behaviorally relevant stimuli.

6. Mechanisms of Perceptual Learning Mirror those of Attention

As we have discussed, several studies have shown that performance on complex visual search tasks can improve with training, but the mechanisms supporting these improvements are not always clear: training might improve the efficacy of attentional mechanisms by increasing the magnitude of sensory gain, by improving the read-out of sensory signals, or by more efficiently mediating internal and external sources of variability. Additionally, studies showing learning in the absence of top-down attention suggest that there may also be a separate low-level mechanism that supports learning in cases where no task-related attentional goals are in place (Watanabe et al., 2001; Yao and Dan, 2001; Seitz and Watanabe, 2003; Seitz et al., 2009; Frankó et al., 2010; Sasaki et al., 2010; Beste et al., 2011; Leclercq and Seitz, 2012; McMahon and Leopold, 2012).

However, in more typical settings, it may be advantageous for the mechanisms of attention to be recruited during learning in order to increase the efficiency of sensory processing and perceptual decision-making. Furthermore, parallels can be drawn between the perceptual improvements associated with learning and those observed with attention, which might guide future investigations of the two phenomena.

As with attention-driven gain enhancements, PL has been found to increase the magnitude of responses associated with the trained stimulus (Zohary et al., 1994; Schoups et al., 2001; Schwartz et al., 2002; Furmanski et al., 2004; Rainer et al., 2004; Yang and Maunsell, 2004; Sigman et al., 2005; Op de Beeck et al., 2006; Yotsumoto et
al., 2009, 2008; Lewis et al., 2009; Bao et al., 2010; Zhang et al., 2010a). Although the majority of these studies used a limited set of stimuli, it is possible that learning-induced response amplification for trained stimuli corresponds to an improved ability of attention to flexibly enhance the gain of sensory neurons that respond to a wide array of task-relevant stimuli, as in the case of visual search improvements and video gaming (Sireteanu and Rettenbach, 2000; Green and Bavelier, 2003).

Training-related improvements in the efficiency of attention may also be related to an improvement in the selective read-out of sensory signals, thereby increasing the speed and accuracy of decision-making without necessarily influencing sensory gain (Palmer et al., 2000; Palmer and Moore, 2009; Pestilli et al., 2011). Similar ideas have been advanced in the PL literature. For instance, Law and Gold (2008) found that there were no changes in the firing rates of MT neurons after training monkeys on a motion discrimination task that varied both motion direction and motion coherence (Figure 5a), but there were changes in the firing rates of the neurons in lateral intraparietal area (LIP), an area implicated in accumulating sensory evidence during decision making (Figure 5b; Law & Gold, 2008; 2009). The authors conclude that learning results from a change in read-out of the most informative sensory neurons and that these read-out changes are driven by feedback, which guides the selective enhancement of the connections between the most sensitive populations in MT and LIP in order to optimize performance (Law and Gold, 2009, 2010). This model is akin to similar proposals in the attention literature (Eckstein et al., 2000; Palmer et al., 2000; Palmer and Moore, 2009; Pestilli et al., 2011), in which responses from the most sensitive sensory neurons are pooled and responses from uninformative sensory neurons are filtered out, leading to overall improvements in
the ability to discriminate target features from distracters (Palmer and Moore, 2009; Gold et al., 2010; Pestilli et al., 2011).

Other models, such as the Perceptual Template Model (PTM) and the Augmented Hebbian Reweighting Model (AHRM), hold that learning is driven both by improved filtering of internal and external noise as well as by a selective enhancement of the most sensitive sensory inputs (Dosher and Lu, 1998, 1999, 2009; Petrov et al., 2005; Lu et al., 2010). As with Law and Gold’s model, learning results primarily from changing the weights associated with early sensory inputs being read-out during decision-making. However, in order for learning to occur in high noise environments, sensory systems must simultaneously enhance the representation of the stimulus as well as filter out stimulus noise (Petrov et al., 2005). In these cases, PL results from an improvement in the efficiency with which top-down control selectively modulates sensory gain enhancement and in the filtering of sensory signals.

General improvements in the efficiency of attentional modulations may also involve a reduction in intrinsic neural noise and the extent to which shared noise among sensory neurons is correlated (Dosher and Lu, 2000; Mitchell et al., 2007, 2009; Cohen and Maunsell, 2009, 2011; Cohen and Kohn, 2011). Reductions in variability at the single-unit level (Mitchell, Sundberg, & Reynolds, 2007) complement increases in gain by further improving the signal-to-noise ratio of cells that encode relevant stimuli. While more complex (Abbott and Dayan, 1999; Cohen and Kohn, 2011; Cohen and Maunsell, 2011), decorrelating shared variability across populations of sensory neurons is often advantageous, as removing correlated noise increases the precision of a response estimate generated via simple pooling operations (Shadlen et al., 1996; Cohen and Maunsell,
Although many of these empirical observations regarding the modulation of intrinsic noise have been documented in the realm of attention studies, recent PL studies have observed similar modulations. For example, Gu et al. (2011) found that training on a direction-heading task reduced noise correlations across sensory neurons in macaque dorsal medial superior temporal area (MSTd), compared to untrained control animals. Trained monkeys matched controls in terms of the time course of responses, tuning curve bandwidth and amplitude, and sensitivity to the stimulus, which led the authors to conclude that the observed noise decorrelations serve to improve the read-out of sensory inputs to decision-making mechanisms. Mathematical models of PL also indicate that internal noise suppression is a key factor in the behavioral improvements associated with learning (Dosher and Lu, 1998, 1999; Bejjanki et al., 2011; Lu et al., 2011). Thus, a reduction in neural (co)variability combined with enhanced sensory signals likely increases the amount of information carried by population responses in visual cortex, allowing for more efficient decision-making and improved behavioral performance (Bejjanki et al., 2011).

Given the broad similarities between documented behavioral and neural manifestations of attention and PL, training subjects on more general tasks may lead to correspondingly general and flexible improvements in sensory gain, read-out, and in the modulation of neural variability. In contrast, training on a highly specific stimulus may instead lead to relatively long lasting changes in local cortical circuits that, once instantiated, no longer need to be guided by top-down attentional control.

7. Conclusions
In many cases, top-down attention can mediate PL for task-relevant stimuli and can support very specific, as well as more general, learning effects that transfer widely across tasks and stimulus features (as in the case of video games, Green and Bavelier, 2003, 2012). On the other hand, PL can also occur for irrelevant stimuli that are presented outside the focus of attention by pairing an irrelevant stimulus with a reward, by pairing an irrelevant stimulus with another task-relevant stimulus, or even via specific types of bottom-up sensory input (e.g. spike-time dependent plasticity; Yao and Dan, 2001; Frankó et al., 2010; McMahon and Leopold, 2012). This observation that the acquisition of learning can proceed without top-down attention suggests that at a very basic level, there are multiple routes that can support the instantiation of long lasting changes in the efficiency of perceptual decision making. Moreover, many changes in processing that are induced by PL do not require active attention to be expressed, as learning-related modulations can be observed when subjects are attending elsewhere and even when animals are under anesthesia (Schoups et al., 2001; Gentner and Margoliash, 2003; Raiguel et al., 2006; Mukai et al., 2007; Bao et al., 2010; Adab and Vogels, 2011). However, attention and PL are not completely independent, as changes in attentional state can override the bottom-up instantiation of PL. This is particularly apparent when trained stimuli actively compete with other stimuli that are currently task relevant (Gilbert et al., 2000; Crist et al., 2001; Fahle, 2004, 2009). Thus, even though PL can occur in the absence of top-down attention, attention may still play a critical role as a ‘gatekeeper’ to determine how training-induced changes in processing are expressed based on current task demands and behavioral goals. However, developing a more complete characterization of the extent to which attention can override all forms of PL is a major
avenue for future research, particularly in situations that involve more complex and ecologically relevant stimulus sets.

This dissertation seeks to clarify the role of top-down attention in perceptual learning so as to contribute to a better understanding of the neuronal and behavioral changes that result from generalizeable improvements in perceptual sensitivity. First, Chapter 1 investigates whether top-down attention can be made more efficient by training multiple stimulus exemplars and whether learning on a more generalized task is supported by an enhancement in the sensitivity of orientation-selective responses in human visual cortex. While viewing the same visual stimulus, subjects performed alternating blocks of an orientation discrimination task and a rapid serial visual presentation (RSVP) in the fMRI scanner followed by 10 days of training and a final fMRI scan. Using an encoding model, we compared the impact of both attention and learning on the sensitivity of orientation-selective response profiles in early visual cortex. The results show that general perceptual learning is supported by an improvement in the efficiency of feature-based attention via a combined suppression of orientation signals when orientation is task-irrelevant and an enhancement of those signals when orientation is task-relevant. We conclude that feature-based attention can be enhanced during learning at that this enhancement is supported by more efficient allocation of sensory gain.

Since Chapter 1 showed that feature-based attention can be made more efficient through learning, Chapter 2 seeks to determine if the presentation of a feature-based attentional cue, compared to a spatial cue, would impact the rate of learning and enable transfer across spatial location. Using a psychophysical motion discrimination task,
subjects were given a feature, spatial, combined feature and spatial, or a neutral cue during four days of training. On the fifth day, they performed the same task in the opposite spatial position so as to assess the effects of the each cue type on the ability of learning to transfer. Since feature-based attention is spatially global, we expected that only learning on feature-cued trials would transfer. The results show that learning only occurs for feature-cued and combined cued trials, supporting a key role for feature-based attention, but not spatial attention, in generalized perceptual learning. Furthermore, only learning on the feature-cued trials transferred to the new spatial position, lending support to the hypothesis that generalized perceptual learning results from more efficient feature-based attentional selection that can be applied across the visual field.

Chapter 3 investigates the neural mechanisms supporting the transfer of learning across spatial position using fMRI and an encoding model. Here, subjects performed a pre-test on an orientation discrimination task at one of four possible locations, a scanning session with the same task, then 10 training sessions at a single location, and finally a post-test and post-training scanning session. The pre- and post-tests allowed us to compare performance at each of the four locations before and after specific training at only one location. Using the same encoding model from Chapter 1, we assessed the impact of attention and learning on the sensitivity of orientation-selective responses in early visual cortex. Behaviorally, we find that learning at one location does transfer to untrained locations. Learning and transfer on this task was supported by greater suppression of non-target orientation signals. This pattern of suppression was present both when the trained quadrant was the target and when the untrained quadrants were the
target, providing further evidence that learning to attend leads to more efficient sensory modulation.

The Introduction, in part, is a reprint of the material as it appears in Exploring the Relationship between Perceptual Learning and Top-Down Attentional Control in Vision Research, 74, 30-39. Byers, A. & Serences, J.T. (2012). The dissertation author was the primary author of this paper.
Regression slope analyses for accuracy in three attentional cue conditions: exogenous attention, endogenous attention with an arrow cue, and endogenous attention with a color cue, for attended (blue), divided-attended (green), and unattended (red) locations. Slopes of the learning curves for attended locations were significantly greater than divided-attended locations and unattended locations. Slopes of the learning curves for divided-attended locations were also significantly greater than those for unattended locations.

Figure i.1: Reproduced with permission from Mukai et al., 2011 (their figure 5a).
Figure 1.2: a) Reproduced with permission from Schoups et al., 2001 (their figure 3a). Slope of the tuning curve for neurons tuned to orientations 0-50 degrees offset from the trained orientation. The solid red line represents trained neurons and the dashed blue line represents untrained neurons. There is a significant increase in the slope of the tuning curve for trained neurons tuned 12-20 degrees away from the trained orientation, indicating that these neurons were the most informative for accurately discriminating the trained stimulus. b) Reproduced with permission from Yang & Maunsell, 2004 (their figure 6). Normalized population tuning curve for trained (black curve) and untrained (gray curve) populations, where 0 degrees represents the preferred orientation of a given neuron. There is a significant decrease in bandwidth and increase in amplitude for the trained populations compared to untrained populations, indicating a gain enhancement of those populations that optimally discriminate the trained stimulus.
Figure i.3: Reproduced with permission from Bao et al., 2010 (their figure 3). a) Time course of the VEPs for trained versus untrained orientation, before and after training. The gray bars mark the latency for the early segment of the C1 component. b, c) Normalized peak amplitude for the trained orientation (TO) versus untrained orientation (UO) at the trained location (b) and untrained location (c).
Figure i.4: Reproduced with permission from Gentner, 2007 (his figure 4b). Mean normalized response strength of single units in CMM to familiar versus unfamiliar songs for the two training regimes (2AFC and go-nogo).
Figure i.5: Reproduced with permission from Law & Gold, 2008 (their figures 3a and 4a). a) Average activity of MT neurons for various motion strengths as a function of viewing time. The solid lines represent each neuron’s preferred direction and dashed lines represent each neuron’s null direction. Each panel displays the different training periods. b) Average activity of LIP neurons for various motion strengths as a function of viewing time. The solid lines represent saccades into each neuron’s receptive field and dashed lines represent saccades out of each neuron’s receptive field.
CHAPTER 1: Enhanced attentional gain as a mechanism for generalized perceptual learning in human visual cortex


Abstract

Learning to better discriminate a specific visual feature (i.e., a specific orientation in a specific region of space) has been associated with plasticity in early visual areas (sensory modulation) and with improvements in the transmission of sensory information from early visual areas to downstream sensorimotor and decision regions (enhanced read-out). However, in many real-world scenarios that require perceptual expertise, observers need to efficiently process numerous exemplars from a broad stimulus class as opposed to just a single stimulus feature. Some previous data suggest that perceptual learning (PL) leads to highly specific neural modulations that support the discrimination of specific trained features. However, the extent to which perceptual learning acts to improve the discriminability of a broad class of stimuli via the modulation of sensory responses in human visual cortex remains largely unknown. Here, we used fMRI and a multivariate analysis method to reconstruct orientation-selective response profiles based on activation patterns in early visual cortex before and after subjects learned to discriminate small offsets in a set of grating stimuli that were rendered in one of nine possible orientations. Behavioral performance improved across 10 training sessions and there was a training-related increase in the amplitude of orientation-selective response profiles in V1, V2, and V3 when orientation was task-relevant compared to when it was task-irrelevant. These results suggest that generalized perceptual learning can lead to
modified responses in early visual cortex in a manner that is suitable for supporting improved discriminability of stimuli drawn from a large set of exemplars.
Introduction

Perceptual learning (PL) refers to an improved ability to detect or discriminate a sensory stimulus after repeated exposure (Gibson, 1963, 1969). This behavioral improvement is thought to be supported by at least two neural mechanisms: an increase in the fidelity of early sensory responses (sensory modulation account) and/or an increase in the efficiency with which early sensory information is relayed to – or ‘read out’ by – downstream decision mechanisms (enhanced read-out account). Consistent with the sensory modulation account, learning for a specific feature can enhance the gain of informative sensory signals (Zohary et al., 1994; Gilbert et al., 2001; Schoups et al., 2001; Furmanski et al., 2004; Yang and Maunsell, 2004; Bao et al., 2010; Jehee et al., 2012) and reduce correlated noise shared between visually responsive neurons (Adab and Vogels, 2011; Bejjanki et al., 2011; Gu et al., 2011). Conversely, read-out models can explain improved perceptual performance without the need to invoke learning-related changes in the fidelity of responses in early sensory areas (Dosher and Lu, 1999, 2007, 2009; Lu and Dosher, 1999; Petrov et al., 2005; Lu et al., 2010, 2011; Huang et al., 2012) and empirical data suggest that training can modify the degree to which informative sensory neurons are weighted during decision-making, particularly when a large set of training stimuli are employed (i.e. the trained stimuli change from session to session (Law and Gold, 2008, 2009, 2010; Gold et al., 2010).

However, most prior investigations that documented changes in early sensory modulation trained subjects to discriminate a single feature value (often in a single location), so the degree to which sensory modulation can support PL across multiple stimulus exemplars is largely unknown. This is an important question, as many real-
world instances of PL require identifying unique members from a broad class of possible features: a radiologist must find a small fracture irrespective of its precise location or its exact orientation, and a baggage screener at an airport must detect potential weapons of all shapes and sizes hidden in a piece of luggage. Intuitively, this more general form of PL might be supported primarily by changes in the read-out of unmodified sensory signals (e.g. Law and Gold, 2008), as any plasticity in sensory areas that facilitates the processing of one specific exemplar might interfere with the processing of other exemplars (Schoups et al., 2001; Fahle, 2004, 2009). On the other hand, it is possible that learning modulates early sensory responses in a more flexible manner that facilitates the processing of all stimuli from the trained set of exemplars, perhaps by enhancing the overall gain of feature-selective responses in early visual areas. Here, we tested this latter hypothesis in human subjects using functional magnetic resonance imaging (fMRI) and a multivariate analysis technique that assesses how PL modifies the characteristics of population-level orientation-selective tuning profiles in human visual cortex.

Methods

Subjects

Thirteen subjects (7 males and 6 females) between the ages of 18-28 (M=21.5, SD=2.8) were recruited from the University of California, San Diego (UCSD) to participate in the experiment, and all gave written informed consent in accordance with the Institutional Review Board at UCSD. Of these thirteen subjects, five were excluded from analysis: one did not complete both scanning sessions, two could not do the task for unknown reasons (with final day orientation offsets of 24° and 40°, which were
approximately 3-10 times larger than the other subjects), and two showed significant motion artifacts (i.e. multiple abrupt translations of 5mm or more in both scanning sessions). Subjects performed a brief 20-minute familiarization session before the initial scanning session, a 2-hour pre-training fMRI scan, ten 1-hour behavioral training sessions in the lab, and a final 2-hour post-training fMRI scan. Subjects were compensated $10/hour for the training sessions and $20/hour for the scan sessions.

Stimulus and task

All stimuli were created using MATLAB (v.7.8, Natick, MA) with the Psychophysics Toolbox (v.3, Brainard, 1997; Pelli, 1997). The physical properties of the stimulus display were identical across two attention conditions: (1) attend to the orientation of a grating stimulus (attend-orientation) and (2) attend to a central rapid serial visual presentation (RSVP) stream of letters (attend-RSVP). Each trial started with a 200ms fixation interval, followed by a peripheral circular square-wave grating and a centrally presented RSVP stream (Figure 1). The grating had a radius of 10° visual angle and a spatial frequency of 0.5 cycles per degree and flickered at a rate of 5Hz. The inner blank aperture around fixation extended 2° visual angle. The orientation of the grating was pseudo-randomly selected on each trial from a set of nine possible orientations that were evenly distributed across 180° of orientation space; ±3° of orientation jitter was randomly added to the first grating on each trial in order to ensure that subjects could not make their judgment with respect to a fixed orientation. The spatial phase of each of the gratings was also randomly determined on each trial to ensure that measured signals were related to orientation and not to specific luminance patterns on the display screen. After 2000ms, the grating was removed from the screen for 400ms, and then a second
flickering grating was presented for an additional 2000ms. On attend-orientation blocks, the subject’s task was to press one button if the second stimulus was rotated clockwise from the first stimulus, and another button if it was rotated counter-clockwise from the first stimulus (buttons were pressed with the first two fingers of the right hand). The RSVP stream appeared within the inner blank aperture of the grating, with each letter subtending 1.2° visual angle. The letter stream was presented during the same temporal epochs in which the oriented gratings were presented (i.e. 2000ms on, 400ms off, 2000ms on). Either an ‘X’ or a ‘Y’ was presented within each epoch of RSVP, and on attend-RSVP blocks the subject’s task was to press one button if the target letters matched (i.e. two ‘X’s’ or two ‘Y’s’) and another button if the two target letters did not match. Subjects had 2000ms to respond after the stimuli were removed from the screen, and this response window also served as the inter-trial-interval (ITI).

At the beginning of each training session, the clockwise or counter-clockwise rotational offset between the first and the second gratings was set to 30°, and performance was adjusted using the QUEST algorithm (Watson and Pelli, 1983). The orientation offset required to achieve 75% accuracy was taken as the threshold for each session and learning was operationalized as the change in the angular offset between the first and second gratings that was required to achieve threshold accuracy during each testing session. The exposure duration of each letter in the RSVP stream was manually adjusted to achieve ~75% accuracy to ensure that the task remained equally challenging across training sessions.

Training regimen
Before the initial scan session, subjects performed 4 blocks of 36 trials of the task (2 blocks each of the attend-orientation and attend-RSVP conditions), in order to familiarize them with the task. After the initial scan session, subjects completed ten 1-hour training sessions. During each training session, subjects performed 4 blocks of the attend-orientation condition and 4 blocks of the attend-RSVP condition, with each condition alternating on a block-by-block basis. Auditory feedback (short beeps, 100ms duration) was provided to indicate correct and incorrect responses during the familiarization and training phases, but not during either scan session. After the tenth training session, subjects completed the second scan session.

**fMRI sessions**

Subjects performed 8 blocks of the behavioral task in each of the two scan sessions, and each block lasted 5 min. Eight null trials were randomly interleaved with 36 task trials in each block. For the first scan session, the offset between the two presentations of the grating was based on the staircased offset from the familiarization session (M=8.3°, SEM=4.7°) and was manually adjusted at the end of each block based on subject performance to maintain consistent performance across runs (M=8.1°, SEM=0.95°). The letter exposure duration was also based on the exposure duration from the familiarization session (M=157.1ms, SEM=6.1ms), but this number was decreased slightly in the scanner (M=140.7ms, SEM=9.5ms), due to relatively high accuracy on this task during the pre-training session. For the second scan session, the orientation offset was based on the final training session (M=2.9°, SEM=0.64°), and again was manually adjusted at the end of each block to maintain consistent performance across runs (M=5.1°, SEM=0.33°). The letter exposure duration was also based on the last training
session (M=33.0ms, SEM=0.0ms), and this number was also manually adjusted to maintain consistent performance (M=85.4ms, SEM=5.0ms).

Functional localizer scans

Two 6.6-min functional localizer scans were run in each scan session to independently identify voxels in retinotopically organized early visual areas that responded to stimuli in the spatial location occupied by the visual stimulus in the main task. On each trial, a contrast-reversing checkerboard stimulus (spatial frequency = 1 cycle/°, temporal frequency = 8 cycles/second) was presented in the center of the screen for 10s. On half of the 16 trials, the radius of the stimulus was 5° visual angle and on the other half the radius of the stimulus was 10° visual angle (matching the size of the stimulus used in the main learning experiment). Subjects were instructed to press a button whenever the contrast of the flickering stimulus dimmed slightly. Contrast-dimming target events occurred 4 times per trial and each dimming lasted for 66.7ms. The occurrence of each target was determined pseudo-randomly, with the constraint that the target appeared at least 1s after stimulus onset and 1s before stimulus offset and each target was separated from the previous one by at least 1s. Trials were separated by a 10s ITI.

Retinotopic mapping

A standard meridian mapping procedure was used to identify visual areas V1, V2v/V2d, and V3v/V3d. Subjects passively viewed a contrast-reversing checkerboard stimulus flickering at 8Hz and covering 60° of polar angle and that alternated every 20s between the horizontal and vertical meridians (Engel et al., 1994; Sereno et al., 1995). A standard general linear model (GLM) with regressors corresponding to epochs of vertical
meridian and horizontal meridian stimulation was used to identify voxels in visual cortex that responded to each respective region of the visual field. Each regressor was generated by convolving a boxcar model of each stimulus timeseries (‘0’ if no stimulus, ‘1’ if stimulus present) with a standard difference-of-two gamma function model (time to peak of positive response: 5s, time to peak of negative response: 15s, ratio of positive and negative responses: 6, positive and negative response dispersion: 1). The retinotopic mapping data were then projected onto a computationally inflated rendering of each subjects’ gray/white matter boundary to facilitate the identification of the horizontal and vertical meridians in order to define each cortical area. Because we used a large central stimulus, voxels in ventral and dorsal aspects of V2 and V3 were combined.

**fMRI data acquisition and pre-processing**

Scanning was performed using a 3T GE MR750 MRI scanner with an 8-channel head coil at UCSD’s Keck Center for Functional MRI. Anatomical images were collected using a T1-weighted sequence that produced images with a 1mm³ resolution (TR/TE=11/3.3 ms, TI=1100ms, 172 slices, flip angle=18°). Functional images were collected using a gradient echo EPI pulse sequence with 35 slightly oblique slices, which covered the whole brain. The slices were collected in ascending interleaved order and were 3mm thick (TR = 2000 ms, TE = 30ms, flip angle = 90°, image matrix = 64 (AP) × 64 (RL), with FOV = 192mm (AP) × 192mm (RL), voxel size = 3mm × 3mm × 3mm). All EPI images were slice-time corrected, motion-corrected (both within and between scans), and high pass filtered (3 cycles/run). BrainVoyager QX (v2.3; Brain Innovation, Maastricht, The Netherlands) was used to perform the data analysis in conjunction with
custom analysis scripts written in MATLAB (version 7.11.0584; The Math Works, Natick, Massachusetts).

**Analysis of functional localizer data**

Data from the functional localizer scans were analyzed using a GLM that contained separate regressors marking the temporal interval in which each of the two stimulus types was presented (small checkerboard and large checkerboard). Each regressor was generated by convolving a boxcar model of the stimulus timeseries (‘0’ if no stimulus, ‘1’ if stimulus) with a standard difference-of-two gamma function model (time to peak of positive response: 5s, time to peak of negative response: 15s, ratio of positive and negative responses: 6, positive and negative response dispersion: 1). A contrast was then performed to identify voxels that responded more to the large 10° checkerboard compared to the small 5° checkerboard; this contrast was intended to exclude voxels that respond strongly to the location of the central RSVP stream. Voxels within each of the visual areas were included in subsequent analyses if they passed a threshold of p<0.05, after using BrainVoyager’s False Discovery Rate (FDR) algorithm to correct for multiple comparisons.

**Estimating trial-by-trial BOLD responses on attention scans**

Before estimating the trial-by-trial magnitude of the BOLD response during the main attention task, we first estimated the shape of the hemodynamic response function (HRF) separately for each visual area in each subject. This subject and area specific fitting was done because the shape of the HRF can vary substantially across subjects and brain regions (Zarahn et al., 1997; Gonzalez-Castillo et al., 2012). Data were averaged across all trials in the main scans and the mean stimulus-locked HRF was computed
separately for each visual area in each subject across a 24s window using a Finite Impulse Response model (Dale, 1999). We then used a gradient descent algorithm implemented in MATLAB to fit each HRF estimate using a standard difference-of-two gamma function model with time to peak, onset time, and the ratio of positive and negative components as free parameters (using the same seed parameters listed in previous section). The parameters that yielded the best fit were used to produce a custom HRF model for each visual area in each subject. We then created a design matrix with a boxcar model of the onset/offset time of each 4.4s stimulus epoch on each trial within a scan, yielding 36 regressors of interest (four instances of each of the nine orientations) along with 1 additional regressor as a constant term. Each boxcar regressor model was then convolved with the custom HRF estimate for that subject, and a GLM was used to estimate the relative amplitude of the BOLD response on each trial. This estimate of response amplitude one each trial in each voxel was then used as input to the orientation encoding model described below.

*Estimating orientation-selective BOLD tuning profiles using a forward encoding model*

Forward encoding models adopt a set of *a priori* assumptions about the important features or stimulus labels that can be distinguished using hemodynamic signals within an ROI (Brouwer & Heeger, 2009, 2011; Dumoulin & Wandell, 2008; Gourtzelidis et al., 2005; Kay & Gallant, 2009; Kay, Naselaris, Prenger, & Gallant, 2008; Mitchell et al., 2008; Naselaris, Prenger, Kay, Oliver, & Gallant, 2009; Schönwiesner & Zatorre, 2009; Scolari, et al., 2012; Thirion et al., 2006; reviewed in Naselaris, Kay, Nishimoto, & Gallant, 2011; Serences & Saproo, 2012). The features or labels in the model are then used to predict the pattern of BOLD responses. Unlike multivoxel pattern analysis
(MVPA) decoding techniques, this approach allows us to quantify changes in the shape of orientation population-response profiles that are related to perceptual learning. This technique thus complements MVPA approaches, as MVPA provides an index of how much overall information is encoded in a pattern of responses and how that level of information changes as a result of PL (or other experimental manipulations: see Serences and Saprio, 2012). However, a general increase in the information content of a response pattern might be supported by several different types of modulation (i.e. a decrease in the bandwidth of orientation-selective population tuning profiles, an increase in gain of orientation-selective population tuning profiles, etc.). Thus, by directly reconstructing population tuning profiles using an a priori encoding model, we can quantify the modulatory patterns that occurred as a result of generalized PL.

Here, we used a forward model adapted from Brouwer and Heeger (2009, 2011) to estimate the response across a series of orientation channels as a function of PL and the focus of attention. The model assumes that the BOLD response in a given voxel reflects the pooled activity across a large population of orientation-selective neurons, and that the distribution of neural tuning preference is biased within a given voxel either due to large-scale feature maps (Freeman et al., 2011) or to subvoxel anisotropies in cortical columns (Kamitani and Tong, 2005; Swisher et al., 2010). In either case, BOLD responses measured from many voxels in early visual cortex exhibit a modest but robust orientation preference (Haynes and Rees, 2005; Kamitani and Tong, 2005; Serences et al., 2009; Freeman et al., 2011).

For each subject and scan session, we first split the data into two sets (training and test sets), using a hold-one-out cross-validation method. Adopting the terminology and
formulations of Brouwer and Heeger (2009; 2011) for consistency, let $m$ be the number of voxels in a given visual area, $n_1$ be the number of observations (trials) in the training set, $n_2$ be the number of trials in the test set, and $k$ be the number of hypothetical orientation channels. Let $B_1$ ($m \times n_1$ matrix) be the training set, and $B_2$ ($m \times n_2$ matrix) be the test set. The training data in $B_1$ were then mapped onto the matrix of hypothetical channel outputs ($C_1, k \times n$) by the weight matrix ($W, m \times k$) that was estimated using a linear model of the form:

$$B_1 = WC_1. \quad (1)$$

where the ordinary least-squares estimate of $W$ is computed as:

$$\hat{W} = B_1 C_1^T (C_1 C_1^T)^{-1}. \quad (2)$$

The channel responses ($C_2, k \times n_2$) were then estimated for the test data ($B_2$) using the weights estimated in (2):

$$\hat{C}_2 = (\hat{W}^T \hat{W})^{-1} \hat{W}^T B_2. \quad (3)$$

Equations 1-2 are similar to a traditional univariate GLM in that each voxel gets a weight for each feature in the model (in this case, one weight for each orientation channel). Equation 3 then estimates channel responses on each trial based on the estimated weights.
assigned to each voxel and the vector of responses observed across all voxels on a given trial in the test set.

The construction of the basis set matrix C has a large impact on the resulting channel response estimates. In the present experiment, we used a half-sinusoidal function that was raised to the 5th power to emulate the approximate shape of single-unit tuning functions in V1, where the average 1/\sqrt{2} half-bandwidth of orientation tuned cells is approximately 20° (Schiller et al., 1976; Ringach et al., 2002; Gur et al., 2005). This function was then replicated 6 times, with the 6 copies evenly distributed across orientation space. The power of the sinusoidal function necessitated the use of at least 6 copies in order to adequately cover orientation space (Freeman & Adelson, 1991). While we selected the shape of the basis function based on existing physiology studies, all results that we report are robust to reasonable variations in this value (i.e. raising the half-sinusoid to the 5th, 6th, 7th or 8th power, all of which are reasonable choices based the large range of single-unit bandwidths observed in early visual areas). The analysis described above was then repeated with the center position of each of the 6 basis functions shifted across all 180° of orientation space in 1° increments to generate channel tuning functions with 1° resolution (which ultimately produced a 180° tuning function). Note that because of overlap between adjacent basis functions, each point along these 180-point tuning curves was not independent from neighboring points. However, this approach was adopted to maximize the smoothness of the orientation tuning functions, which in turn facilitates quantifying the amplitude, bandwidth, and baseline level using the model fitting approach described below. Last, the 180-point channel response function estimated on each trial was circularly shifted so that the channel matching the orientation of the
stimulus on that trial was positioned in the center of the tuning curve, thereby aligning all channel response profiles to a common stimulus-centered reference frame (which is plotted as 0° on the x-axis by convention, see Figures 5-6). The channel responses for each subject were then averaged across subjects within each scanning session so that group channel response functions for the different attention and training conditions could be compared.

Quantifying channel responses by fitting a von Mises function

After re-centering the channel responses to a common reference frame, data from each subject and each condition (attend-orientation/attend-RSVP, pre-training/post-training) were independently fit with a von Mises function:

\[ f(\theta) = ae^{\sigma(\cos(\mu - \theta) - 1)} + b \]  

(4)

with amplitude (\(a\)), mean (\(\mu\)), bandwidth (\(\sigma\)), and baseline (\(b\)) as independent free parameters that reflect distinct attributes of the function. The amplitude was restricted to a range from 0 and 2, the mean was restricted to a range from 60° and 120°, the bandwidth was restricted to a range between 0 and 8, and the baseline was restricted to a range from -3 and 3. The von Mises function (eq 4) was then fit to the data from each subject 150 times using randomly chosen initial seed values for each parameter on each iteration (across the range of allowable values) to ensure that the fitting algorithm did not settle in a local minimum. The set of parameters for each subject that yielded the lowest root mean squared error across the 150 iterations were then used in subsequent analyses.
Results

Behavior during training sessions outside of scanner

The staircasing procedure did not converge for one subject during the first of their 10 behavioral training sessions, rendering data from session 1 highly variable (that subject had an orientation offset of 22.2°, where the mean across all other subjects was 4.7°). Therefore statistical comparisons of the behavioral data were performed excluding this subject (although the size of the learning effect only increases if this subject is included, and their offsets during sessions 2-10 were stable and showed learning effects). The angular offset between the gratings needed to perform the task at criterion was steadily reduced across training sessions (see Figure 2a; one-way repeated measures ANOVA, F(9,54) = 2.18, p=0.038) and there was a significant decrease in the average orientation offset between session 1 (4.7°) and session 10 (3.0°; paired t-test, t(6) = 3.4150, p=0.014). Response time (RT) did not significantly change across sessions 1-10 in the attend-orientation task (one-way repeated measures ANOVA, F(9,54) = 0.48, p=0.88). Accuracy also did not significantly change across sessions 1-10 (see Figure 2b; one-way repeated measures ANOVA, F(9,54) = 0.8, p=0.62), suggesting that staircasing was successful in maintaining a criterion level of accuracy.

Observers also improved in the attend-RSVP condition, such that the exposure duration of each letter in the RSVP stream steadily decreased across the 10 training sessions (see Figure 2c; one-way repeated measures ANOVA, F(9,54) = 22.89, p=0.0002). The average letter exposure duration also significantly decreased from 128.57 ms/letter during session 1 to 33.33 ms/letter during session 10 (paired t-test, t(6) = 8.40, p=0.0002). However, accuracy also changed across the 10 sessions in the RSVP task (see
Figure 2d; one-way repeated measures ANOVA, $F(9,54) = 5.08, p=0.0001$), suggesting that the reduction in exposure duration across sessions is not a perfect indicator of the amount of learning that occurred in the RSVP task (although this was not an issue when comparing pre- and post-training behavioral performance in the scanner as all comparisons were carried out within session, see below). RTs for the attend-RSVP condition did not significantly change across training sessions (one-way repeated measures ANOVA, $F(9,54) = 0.82, p=0.6$).

**Behavior in the scanner**

In the scanner, the average orientation offset was $8.1^\circ$ (SEM: $0.95^\circ$) in the pre-training scan session and $5.1^\circ$ (SEM: $0.33^\circ$) in the post-training scan session (Figure 3a, paired t-test, $t(7) = 3.84, p=0.006$). For the RSVP task, the pre-training letter presentation speed was 140.63 ms/letter (SEM: 9.5ms/letter) and the post-training letter presentation speed was 85.42 ms/letter (SEM: 5.0ms/letter; Figure 3c, paired t-test, $t(7) = 6.33, p=0.0004$). We speculate that the slightly elevated orientation offsets and letter exposure durations in the scanner, compared to outside the scanner, were due to the lower-quality visual display in the scanner. Accuracy on attend-orientation trials was slightly lower during the post-training scan session compared to the pre-training session (see Figure 3b; orientation task: 91% pre-training SEM = 1%, 88% post-training SEM=1%, paired t-test, $t(7) = 2.78, p=0.03$). However, this small change in accuracy was not a major cause of the observed learning effects, as 4 subjects had balanced accuracy across pre- and post-training scan sessions and they all showed at least $4^\circ$ improvement with training. Accuracy in the attend-RSVP condition did not significantly differ across scan sessions
(see Figure 3d; letter task: 89% pre-training (SEM = 2%), 88% post-training (SEM = 2%, repeated measures t-test, t(7) = 0.8769, p=0.4).

fMRI Results

First, we computed the mean univariate HRF for each attention condition based on the z-scored BOLD time series averaged across all voxels in each ROI (see Methods, Figure 4). As all three areas showed a similar pattern of responses, we performed statistics on the data averaged across V1, V2, and V3 (compare panels in Figure 4 below). A three-way repeated measures ANOVA with time (0-24s), attention condition, and training (pre vs post training) revealed a main effect of time (F(12,84)=41.13, p<0.0001), a main effect of attention condition (F(1,7)=36.91, p=0.0005) and an interaction between time and attention (F(12,84)= 12.24, p<0.0001). However, there was no main effect of training (F(1,7)=0.11, p=0.74) and no interaction between training and time (F(12,84)=7.99, p=0.65) or training and attention condition (F(1,7)=2.91, p=0.133). Overall, this pattern suggests that the HRFs were significantly modulated by attention but not by learning.

We then used an encoding model (see Methods) to determine how learning modulated orientation-selective response profiles in V1, V2, and V3. Across both scan sessions and attention conditions, the averaged orientation-selective response profiles peaked at the target orientation and fell off in a Gaussian-like manner (see Figures 5, 6). We fit the tuning curve for each subject with a von Misses function and submitted the best fitting parameters to a three-way repeated-measures ANOVA with factors for attention condition (attend-orientation vs. attend-RSVP), training (scan 1 vs. scan 2), and visual area (V1, V2, V3). There was no main effect of visual area and no three-way
interaction between visual area, attention condition, and training condition for any of the fit parameters (amplitude, bandwidth, baseline, or mean, see Table 1 for all statistics). This suggests that all areas showed a similar pattern of modulation across task manipulations (see Figures 6 and 7 for data from individual visual areas). As a result, we hereafter focus our statistical analysis on the von Mises parameters fit to tuning functions that were computed after combining all voxels from V1-V3. Because the size of the orientation offset differed across pre- and post-training scan sessions, a comparison between pre- and post-training within each attention condition could not be directly assessed due to the sensory differences in the display. Thus, we focused on evaluating a two-way repeated measures ANOVA with attention condition and training as factors to compared to the magnitude of attention effects during pre- and post-training sessions as our main marker of PL. The ANOVA revealed a significant overall main-effect of attention condition on the amplitude of the orientation-selective tuning profiles, such that the amplitude in the attend-orientation condition was higher than the amplitude in the attend-RSVP condition (collapsed across scanning sessions; Figure 5a,b; F(1,7)=10.03, p=0.015). In addition, there was a significant interaction between attention condition and training on tuning function amplitude, indicating that the magnitude of the attentional modulation (i.e., attend-orientation vs. attend-RSVP) increased with training (Figure 5c; F(1,7)=7.98, p=0.025). This interaction was driven by a significant difference between response amplitudes in the attend-orientation and attend-RSVP conditions post-training, but not pre-training (t(7)=3.5, p<0.01 and t(7)=1.25, p=0.25, respectively). Finally, there was an overall main-effect of attention condition on the baseline parameter, such that baseline estimates were higher in the attend-RSVP condition than the attend-orientation
condition (Figure 5e; F(1,7)=7.84, p=0.026). No learning-related changes in any of the other parameters were observed.

Discussion

Learning to better discriminate a specific stimulus feature has been associated with enhanced sensory responses in early visual cortex (Gilbert et al., 2001; Schoups et al., 2001; Furmanski et al., 2004; Yang and Maunsell, 2004; Jehee et al., 2012) and with the improved read-out of sensory information by downstream sensorimotor and decision mechanisms (Dosher and Lu, 1998, 2009; Ghose et al., 2002; Law and Gold, 2008, 2009, 2010; Xiao et al., 2008; Zhang et al., 2010c, 2013; Gold et al., 2010; Lu et al., 2011; Huang et al., 2012). However, little is known about how these mechanisms support PL in situations where observers must learn to make difficult discriminations between stimuli that are drawn from a large set of exemplars. Indeed, modified early sensory responses that are highly selective for supporting the discrimination of a specific feature might be predicted to impair the discriminability of other stimuli (e.g. Schoups et al., 2001; Fahle, 2004, 2009). However, here we show that generalized PL leads to an enhancement of feature-selective attentional modulations in early visual cortex, supporting a version of the sensory modulation account of generalized PL.

Even though our data support a sensory modulation account of generalized PL, they do not speak to the possibility of additional contributions due to changes in the efficiency of sensory read-out. Indeed, sensory modulation and changes in the efficiency of read-out are not mutually exclusive. For example, Law and Gold (2008) trained subjects using a large stimulus set and their data provide support for enhanced read-out in the absence of changes in early sensory responses in motion selective area MT. Their task
involved discriminating stimuli moving in one of two directions, where the exact
direction and location of the motion stimulus was changed on a session-by-session basis
to optimize the response of cells in MT and the lateral intraparietal area (LIP). Law and
Gold did not observe training-related changes in MT, an area that represents the quality
of motion-evoked sensory signals. Instead, they found an increase in motion-driven
responses in LIP, were cells are thought to play a role in the ‘read out’ (or sensorimotor
transformation) of sensory signals that are generated in MT. However, there are several
differences between our study and theirs that might contribute to the differences that we
observe. First, we used fMRI to assess the response profile across orientation channels
tuned to all stimulus features (the entire range from 0°-180°) whereas Law and Gold just
assessed the activity of cells that were selectively tuned to the trained stimulus feature (or
to the anti-preferred feature). Thus, we mapped out the full tuning profile and they
mapped two points along this function. This could be a key difference, as our results
suggest that the overall amplitude increase in fMRI-based orientation tuning profiles is
due to a combination of enhanced responses in orientation channels tuned around the
relevant stimulus feature and to attenuated responses in orientation channels tuned
progressively farther away from the relevant stimulus (compare Figures 5a to 5b). Thus,
measuring the entire tuning profile might provide improved sensitivity to detect learning-
related changes in the amplitude of sensory responses. In addition, and perhaps more
importantly, our display contained multiple competing stimuli (a central RSVP stream
and a peripheral grating), and competition between multiple stimuli has been previously
shown to amplify the magnitude of attentional modulations in visual cortex (e.g. Moran
and Desimone, 1985; Motter, 1993; Desimone and Duncan, 1995; Kastner et al., 1998;
Thus, the degree of inter-stimulus competition might be a key factor in determining the magnitude of sensory modulation during PL. In either case, we believe that the current evidence – both from our study and from others – firmly suggests that PL enhances sensory modulation as well as enhanced read-out, and future studies will need to directly determine how much each mechanism contributes to behavioral indexes of PL, particularly as a function of the degree of competition between relevant and irrelevant stimuli in the visual field.

The present observation of a training-related increase in feature-selective attention effects is consistent with several previous studies that focused on PL for highly specific visual features (reviewed in Seitz and Dinse, 2007; Roelfsema et al., 2010; Sasaki et al., 2010; Byers and Serences, 2012). In particular, the present data are consistent with previous suggestions that top-down attentional modulations play a key role in both context-dependent effects on sensory processing, perceptual learning, and attenuating responses evoked by task-irrelevant features (Gilbert et al., 2000; Crist et al., 2001; Schoups et al., 2001; Li et al., 2004; Vidnyánszky and Sohn, 2005; Gál et al., 2009; Mukai et al., 2011; Gilbert and Li, 2012). For example, Gilbert and Li (2013) summarize evidence for top-down attentional control and recurrent feedback connections as a mechanism for changing early sensory responses based on task goals; a pattern which is present even in the pre-training orientation-selective response profiles described here. However, our data suggest a further key role of task context in gating post-training sensory responses, as enhanced responses to trained stimuli were primarily elicited when the orientation stimulus is task-relevant and not when the orientation stimulus is merely
present (i.e. in the attend RSVP condition; Gilbert et al., 2000; Li et al., 2004; Gilbert and Li, 2012). Furthermore, the nature of the neural changes associated with perceptual learning may be dependent on the context in which subjects were trained, suggesting that training with more irrelevant distractors could lead to even larger attentional modulations, both before and after learning (Luck et al., 1993; Motter, 1993; Kastner and Ungerleider, 2000; Li et al., 2004; Vidnyánszky and Sohn, 2005; Gál et al., 2009; Gilbert and Li, 2012).

Recently, Jehee et al. (2012) used fMRI to show that the ability of individual voxels in V1-V4 to discriminate between two similar orientations was improved when the oriented stimulus was attended compared to when it was unattended (Jehee et al., 2012). However, this improvement was specific to both the trained orientation and stimulus location, so the implication of their results to situations in which training is focused on improving the discriminability of a larger class of stimuli is unclear. In contrast, we addressed this complementary question by exploiting the joint information encoded by large-scale populations to show that generalized orientation learning improves the amplitude of feature-based modulations in early visual cortex. Thus, the current results demonstrate that PL modifies population-level orientation tuning profiles via enhanced attentional gain, and that these population-level modulations provide computational constraints on models of PL in more complex scenarios.

In the present study, we observed a qualitative attenuation of responses to the grating stimulus when subjects were performing the attend-RSVP task (see blue line in Figure 5b; although note that we cannot compare pre- and post-training response amplitude levels quantitatively in the attend-RSVP condition due to learning-related
changes in the nature of the sensory stimulus, see Results above). At first glance, this observation seems inconsistent with prior work showing PL even for task-irrelevant stimuli when they are paired with the presentation of a concurrent relevant or rewarding stimulus (Seitz and Watanabe, 2005; Seitz et al., 2009; Frankó et al., 2010). However, task-irrelevant PL seems to occur only when the irrelevant stimulus does not strongly compete or interfere with the processing of the relevant attended stimulus (Watanabe et al., 2001; Seitz et al., 2006, 2009; Tsushima et al., 2008; Choi et al., 2009; Seitz and Watanabe, 2009; Leclercq and Seitz, 2011, 2012; Huang and Watanabe, 2012). This criterion is not likely met in our study, as the grating was highly salient and it surrounded the relatively low-salience central RSVP stream. Thus, this stimulus configuration likely led to a high degree of inter-stimulus competition, and in turn this competition may have blocked task-irrelevant PL effects on the oriented grating when subjects were attending the central RSVP stream.

In previous reports (e.g., Regan and Beverley, 1985; Seung and Sompolinsky, 1993; Schoups et al., 2001), including work from our own lab (Scolari and Serences, 2009, 2010; Scolari et al., 2012), the performance of a fine discrimination task between similar orientations has been shown to mediate the gain of ‘off-channel’ neural populations, or those neurons that are tuned just away from the to-be-discriminated feature. From an information-theoretic viewpoint, this is optimal because these off-channel neurons have a steeper tuning function around the discriminated feature and should thus undergo a larger change in firing rate to the two stimulus alternatives (see Scolari and Serences, 2009 for more discussion). Thus, we might have expected such off-channel modulations in the present study to be enhanced with PL. However, in Experiment 1 of
Scolari et al. (2012), we demonstrated that off-channel gain – at least as assessed using fMRI and the same analysis approach and general paradigm used here – is not observed when the direction of the orientation offset is unknown in advance (i.e. the subject does not know whether to attend to and expect a clock-wise or a counter-clockwise rotation). In Experiment 2 in the same paper, we then demonstrated a robust off-channel gain effect when we pre-cued subjects about the rotational offset of the oriented stimuli in advance. Given that the present task closely mimics Experiment 1 from Scolari et al. (2012) in which no pre-cue was given, we did not expect to see off-channel gain in early visual areas in the present study and indeed no robust evidence for such an effect was observed.

Although we observed an increase in the overall amplitude of orientation-selective responses in early visual areas (Figure 5c), we did not observe any training-related changes in the mean BOLD response computed across all voxels in each ROI (Figure 4). Several previous reports also failed to observe changes in the mean BOLD response with PL (Yotsumoto et al., 2008; Zhang et al., 2010a; Jehee et al., 2012). Note however that the presence of enhanced feature-based attention effects (Figures 5, 6) can still be observed even if there is no net change in the magnitude of the mean HRF that is computed based on the average across all voxels. For example, we observed enhanced responses in channels tuned to the relevant stimulus and attenuated responses in channels tuned away from the relevant stimulus (Figure 5b, Figure 6). Thus, increases and decreases in the channel responses roughly cancel out and we would not predict a change in overall HRF amplitude. In addition, others have suggested that learning leads to improved coding for specific features even in the absence of a change in mean HRF amplitude, as PL might reduce response variability (which would effectively increase the
signal-to-noise ratio of the HRF without changing its overall magnitude; Jehee et al., 2012). Finally, it is possible that the amplitude of the mean BOLD response is modulated by PL, but that early learning-induced increases in BOLD amplitude return to baseline with further training (Yotsumoto et al., 2008). Thus, we may not have observed changes in HRF amplitude because our post-training fMRI session took place after the HRF returned to pre-training baseline levels.

In sum, we used a forward encoding model to make inferences about how PL influences the shape of population-level orientation tuning functions based on the multivariate response profiles in visual cortex (Serences and Saproo, 2012). We show that PL leads to a selective increase in the amplitude of orientation-selective tuning profiles in V1-V3 with training. This amplitude increase corresponds to an increase in the dynamic-range (or entropy) of orientation-selective response profiles within these regions. In turn, increasing the dynamic range of feature-selective response profiles should, in theory, increase the information content of these population-response profiles and thus the separability of responses evoked by stimuli rendered at different orientations (i.e. an amplitude increase should magnify the difference between mean responses evoked by two nearby orientations; Saproo and Serences, 2010, 2014).

Chapter 3, in full, is a reprint of the material as it will appear in Enhanced Attentional Gain as a Mechanism for Generalized Perceptual Learning in Human Visual Cortex in Journal of Neurophysiology (in press). The dissertation author was the primary author of this paper.
Table 1.1: Statistical comparisons for each fit parameter across conditions and visual areas. Significant effects shown in bold type.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Error</th>
<th>F-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual Area</td>
<td>(2, 14)</td>
<td>2.587</td>
<td>0.1107</td>
</tr>
<tr>
<td>Day</td>
<td>(1, 7 )</td>
<td>0.009</td>
<td>0.9273</td>
</tr>
<tr>
<td><strong>Attn Cond</strong></td>
<td>(1, 7 )</td>
<td><strong>10.025</strong></td>
<td><strong>0.0158</strong></td>
</tr>
<tr>
<td>Vis x Day</td>
<td>(2, 14)</td>
<td>0.711</td>
<td>0.5082</td>
</tr>
<tr>
<td>Vis x Attn</td>
<td>(2, 14)</td>
<td>0.241</td>
<td>0.7888</td>
</tr>
<tr>
<td><strong>Day x Attn</strong></td>
<td>(1, 7 )</td>
<td><strong>7.984</strong></td>
<td><strong>0.0256</strong></td>
</tr>
<tr>
<td>Vis x Attn</td>
<td>(2, 14)</td>
<td>0.348</td>
<td>0.7121</td>
</tr>
<tr>
<td><strong>Vis x Day</strong></td>
<td>(2, 14)</td>
<td><strong>0.338</strong></td>
<td><strong>0.086</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Area</td>
<td>(2, 14)</td>
<td>3.051</td>
<td>0.0795</td>
</tr>
<tr>
<td>Day</td>
<td>(1, 7 )</td>
<td>1.058</td>
<td>0.3378</td>
</tr>
<tr>
<td><strong>Attn Cond</strong></td>
<td>(1, 7 )</td>
<td><strong>0.009</strong></td>
<td><strong>0.9263</strong></td>
</tr>
<tr>
<td>Vis x Day</td>
<td>(2, 14)</td>
<td>0.17</td>
<td>0.8456</td>
</tr>
<tr>
<td>Vis x Attn</td>
<td>(2, 14)</td>
<td>1.044</td>
<td>0.378</td>
</tr>
<tr>
<td><strong>Day x Attn</strong></td>
<td>(1, 7 )</td>
<td><strong>3.988</strong></td>
<td><strong>0.086</strong></td>
</tr>
<tr>
<td>Vis x Attn</td>
<td>(2, 14)</td>
<td>1.816</td>
<td>0.199</td>
</tr>
<tr>
<td><strong>Vis x Day</strong></td>
<td>(2, 14)</td>
<td><strong>1.591</strong></td>
<td><strong>0.2384</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Baseline</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Area</td>
<td>(2, 14)</td>
<td>3.246</td>
<td>0.0695</td>
</tr>
<tr>
<td>Day</td>
<td>(1, 7 )</td>
<td>0.199</td>
<td>0.6687</td>
</tr>
<tr>
<td><strong>Attn Cond</strong></td>
<td>(1, 7 )</td>
<td><strong>7.84</strong></td>
<td><strong>0.0265</strong></td>
</tr>
<tr>
<td>Vis x Day</td>
<td>(2, 14)</td>
<td>0.43</td>
<td>0.6591</td>
</tr>
<tr>
<td>Vis x Attn</td>
<td>(2, 14)</td>
<td>3.11</td>
<td>0.0761</td>
</tr>
<tr>
<td><strong>Day x Attn</strong></td>
<td>(1, 7 )</td>
<td><strong>4.121</strong></td>
<td><strong>0.0819</strong></td>
</tr>
<tr>
<td>Vis x Attn</td>
<td>(2, 14)</td>
<td>0.969</td>
<td>0.4037</td>
</tr>
<tr>
<td><strong>Vis x Day</strong></td>
<td>(2, 14)</td>
<td><strong>1.554</strong></td>
<td><strong>0.2458</strong></td>
</tr>
<tr>
<td><strong>Vis x Attn</strong></td>
<td>(2, 14)</td>
<td><strong>0.962</strong></td>
<td><strong>0.3593</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Area</td>
<td>(2, 14)</td>
<td>0.348</td>
<td>0.7121</td>
</tr>
<tr>
<td>Day</td>
<td>(1, 7 )</td>
<td>0</td>
<td>0.9856</td>
</tr>
<tr>
<td><strong>Attn Cond</strong></td>
<td>(1, 7 )</td>
<td><strong>0.742</strong></td>
<td><strong>0.4175</strong></td>
</tr>
<tr>
<td>Vis x Day</td>
<td>(2, 14)</td>
<td>0.338</td>
<td>0.7191</td>
</tr>
<tr>
<td>Vis x Attn</td>
<td>(2, 14)</td>
<td>1.554</td>
<td>0.2458</td>
</tr>
<tr>
<td><strong>Day x Attn</strong></td>
<td>(1, 7 )</td>
<td><strong>0.962</strong></td>
<td><strong>0.3593</strong></td>
</tr>
<tr>
<td>Vis x Attn</td>
<td>(2, 14)</td>
<td>1.591</td>
<td>0.2384</td>
</tr>
</tbody>
</table>
Figure 1.1: Schematic of the experimental paradigm. In both task types (attend-orientation/attend-RSVP), subjects viewed a 2s presentation of an annular square-wave grating with a rapid serial visual presentation (RSVP) letter stream in the center, followed by a 400ms blank inter-stimulus-interval (ISI), and then another 2-s exposure to a grating and a central RSVP stream. While visible, the gratings flickered at 5Hz to drive a strong visual response (and the phase of the grating’s spatial frequency was randomly selected on each flicker). On attend-orientation blocks, subjects indicated whether the second grating was rotated clockwise or counter-clockwise from the first grating. On attend-RSVP blocks, subjects monitored the RSVP stream for a target letter (either X or Y) and indicated whether the target letter in the second stream matched (i.e. two ‘X’s’ or two ‘Y’s’) or mismatched (i.e. an ‘X’ followed by a ‘Y’ or vice versa) the target letter in the first stream. Subjects responded via a manual button press with either their index finger (to indicate a counter-clockwise rotation of the gratings or a match between target letters) or their middle finger (to indicate a clockwise, rotation of the gratings or a mismatch between target letters).
Figure 1.2: Behavioral performance for training sessions outside of the scanner. a) Average orientation offset in degrees between the two gratings during each of the 10 behavioral training sessions (at a fixed accuracy threshold, see panel B). Note that for one subject, the staircasing procedure did not converge, leading to an over-inflated estimate of their orientation offset on session 1; hence all statistical analyses were performed excluding this subject (and data from that subject are excluded from this plot as well, see main text for more details). b) Accuracy on the attend-orientation condition was maintained at a fixed rate across all behavioral testing sessions. c) Average letter exposure duration in ms/letter across each of the 10 behavioral training sessions. d) Accuracy on the attend-RSVP condition decreased across training sessions (see text for more details). All error bars ±1SEM. *indicates significance across sessions at p<0.05
Figure 1.3: Behavioral performance inside the scanner. 

a) Average orientation offset (at threshold) in degrees between the two gratings during the pre-training (dark gray) and post-training (light gray) scan sessions. 

b) Accuracy in the attend-orientation condition was ~90% during both the pre- and post-training scan sessions, although accuracy in the post-training scan decreased slightly (see main text). 

c) Average letter exposure duration in ms/letter during the pre-training (thick bars) and post-training (thin bars) scan sessions. 

d) Accuracy on the attend-RSVP condition was maintained at ~90% during both the pre- and post-training scan sessions. All error bars ±1SEM. * indicates significance at p<.05.
Figure 1.4: Mean hemodynamic response functions (HRFs) for the attend-orientation (black lines) and attend-RSVP (blue lines) conditions across training conditions (solid: pre-training, dashed: post-training). a) HRFs averaged across voxels in all three ROIs. b)-d) HRFs for each individual visual area (V1, V2, V3, respectively). *indicates a significant effect of attention condition. †indicates a significant main effect of time. ×indicates a significant attention by time-course interaction. Note that learning did not modulate the time-course or the magnitude of attention effects. All error bars ±1SEM. All p<0.05
Figure 1.5: Orientation-selective tuning curves computed across V1-V3 and averaged across subjects, centered on the target orientation on a given trial and resulting parameter estimates from the best fitting von Mises function. a) Pre-training tuning curves during the attend-orientation (gray) and attend-RSVP conditions (blue). b) Post-training tuning curves for the same two conditions. c) Mean amplitude of the best-fitting von Mises function the orientation tuning curves in the attend-orientation (gray) and attend-RSVP conditions (blue) during the pre-training (left) and post-training (right) scan sessions. Gray * indicates a significant effect of attention condition. × indicates a significant interaction between attention condition and training condition. p<0.05. d) Mean bandwidth of the best-fitting von Mises function in the attend-orientation (gray) and attend-RSVP (blue) tasks during the pre-training (left) and post-training (right) scan sessions. e) Mean baseline of the best-fitting von Mises function in the attend-orientation (gray) and attend-RSVP (blue) tasks during the pre-training (left) and post-training (right) scan sessions. Blue * indicates a significant effect of attention condition at p<0.05. All error bars are ±1SEM.
Figure 1.6: Orientation-selective tuning curves for each individual visual area (V1, V2, V3) and averaged across subjects, centered on the orientation of the grating stimulus presented on each trial. The left column shows the pre-training tuning curves for V1, V2, and V3 for the attend-orientation (gray) and attend-RSVP (blue) conditions. The right column shows the post-training orientation-selective tuning curves for V1, V2, and V3 for the attend-orientation (gray) and attend-RSVP (blue) conditions. All error bars are ±1SEM.
Figure 1.7: Parameter estimates from the best fitting von Mises function for each individual visual area (V1, V2, V3). The leftmost column shows the amplitude of best-fitting von Mises function for the attend-orientation (gray) and attend-RSVP (blue) tasks during the pre-training (left) and post-training (right) scan sessions. The center column shows the bandwidth of the best-fitting von Mises function in the attend-orientation (gray) and attend-RSVP (blue) tasks during the pre-training (left) and post-training (right) scan sessions. The rightmost column shows the baseline of the best-fitting von Mises function in the attend-orientation (gray) and attend-RSVP (blue) tasks during the pre-training (left) and post-training (right) scan sessions. All error bars are ±1SEM.
CHAPTER 2: Learning and Transfer of Feature-Based Attentional Modulation

Abstract

Perceptual learning has traditionally been considered to be specific to the training conditions, including the feature dimension and the spatial location of the stimulus. Although recent work on transfer of learning suggests that under certain attention-mediated training conditions learning can transfer across spatial location, the exact role of attentional cues during both learning and transfer have not been thoroughly fleshed out. Here, we employ both feature-based attentional cues and spatial attention cues and a broad stimulus set in order to examine the role of attention as a mediator for perceptual learning and its transfer. We find that both a feature cue and a combined feature and spatial cue support learning, but that only feature cues allow for the transfer of learning across spatial location. We take these results to mean that feature-based attention can be made more efficient through perceptual learning of broad stimulus sets. Furthermore, because feature-based attention is spatially global, the benefits associated with learning to attend to feature are transferrable across spatial location. Thus, generalized perceptual learning is supported by an enhancement in top-down feature-based attentional mechanisms that is not limited to the trained location.
Introduction

Perceptual learning has long been characterized as the improvement in one’s ability to detect or discriminate a stimulus after repeated exposure to that stimulus (Gibson, 1963, 1969). Recently, there has been debate about both the role of attention in mediating learning, as well as the specificity of these perceptual benefits (Shiu and Pashler, 1992; Herzog and Fahle, 1997; Ahissar and Hochstein, 2004; Fahle, 2004, 2009; Seitz and Watanabe, 2005, 2009; Xiao et al., 2008; Yotsumoto and Watanabe, 2008; Ahissar et al., 2009; Zhang et al., 2010c; Roelfsema et al., 2010; Sasaki et al., 2010; Pavlovskaya and Hochstein, 2011; Wang et al., 2012). For instance, one model of perceptual learning suggests that top-down attention is recruited during training in order to select the neural populations that most efficiently code the trained stimulus (Ahissar and Hochstein, 2004; Ahissar et al., 2009). According to this Reverse Hierarchy Theory, top-down attention starts the selection process at higher cortical levels and works back toward lower sensory inputs in order to enhance the populations with the best signal-to-noise ratio for the trained stimulus. This model also suggests that the degree of transfer of learning is dependent upon the level of visual processing at which learning takes place—the further down the visual stream the stimulus is processed, the more likely the learning is to transfer across retinotopic position, orientation, etc. (Pavlovskaya and Hochstein, 2011). Others have suggested that top-down attentional modulations gate error feedback and are necessary to shape learning, such that discrimination only improves, or at least improves more rapidly for attended stimuli (Gilbert et al., 2000; Schoups et al., 2001; Fahle, 2004, 2009; Mukai et al., 2011).
Recent work on transfer of learning also implicates top-down attentional mechanisms as necessary for the association of trained sensory inputs with the to-be-transferred location (Xiao et al., 2008; Zhang et al., 2010c, 2013; Wang et al., 2012). These studies emphasize task-irrelevant stimulation in the transfer location via pre-tests or double-training, suggesting that this allows top-down attention to pair the sensory inputs with higher level non-specific decision rules, thereby promoting transfer to other locations (Xiao et al., 2008; Zhang et al., 2010c).

Although there has been much discussion about general top-down attention gating perceptual learning, the impact of specific attentional cues on perceptual learning is unclear. The role of feature-based attention is particularly important, since most perceptual learning paradigms train attention to a particular sensory feature, such as orientation, motion direction, etc. We have previously suggested that training on multiple stimulus exemplars as opposed to a single feature would allow for the recruitment of and an improvement in the efficiency of top-down feature-based attention (Byers and Serences, 2012). This has important implications for transfer of learning because feature-based attention is spatially global and an improvement in the allocation of feature attention should transfer across the visual field (Treue and Martínez Trujillo, 1999; Saenz et al., 2002; Martinez-Trujillo and Treue, 2004; Maunsell and Treue, 2006). Conversely, space-based attention should not show the same transferability. Here, we compared the relationship between feature, spatial, and combined feature and spatial attention cues on perceptual learning and its ability to transfer across spatial position.

**Methods**
Thirty-one subjects (22 females, 9 males) were recruited from University of California, San Diego and were given written informed consent in accordance with UCSD’s Institutional Review Board. Of these, six subjects were discarded: 2 performed at chance, 4 could not be calibrated by the eye tracker, and 5 withdrew before the study was complete, leaving 20 subjects total (16 females). Subjects completed four training sessions and a transfer session, each comprised of 18 blocks of 32 trials. All subjects had their eyes tracked while performing the task in order to ensure that they were fixating on the central fixation ring. Trials with broken fixations were not included in the analysis. Subjects were compensated $10/hour plus bonus money for accurate responses.

The stimuli were created using MATLAB (v.7.8, Natick, MA) with the Psychophysics Toolbox (v.3, Brainard, 1997; Pelli, 1997). The stimulus was comprised of four random dot kinematograms (RDKs) either above or below a central fixation ring at an eccentricity of 8 degrees of visual angle (Figure 1). Each RDK had a radius of 2.5 degrees of visual angle. The target RDK could move in one of 9 equally spaced directions between 0-360° at one of 6 logarithmically spaced motion coherence levels between 0-80%. All distractor RDKs moved with 0% motion coherence. The task was a 2-interval forced choice (2IFC) task in which one of four possible cue types appeared for 1s, followed by the first stimulus interval for 1s, a 1-s inter-stimulus interval (ISI), and the second 1-s stimulus interval. Subjects were told to determine whether the target RDK with the most coherent motion occurred in the first interval or the second interval. A quarter of the trials contained a feature cue in the form of a red dot on the fixation ring that indicated the direction of the most coherent motion in the target RDK. Another quarter of the trials contained a spatial cue in which a small square corresponding to the
target RDK changed from black to blue. The third quarter of the trials had a combined
cue in which the feature cue and spatial cue were both presented. Finally, the remaining
trials had a neutral cue in which no additional information about the target was presented.
All cues (except neutral) were 100% valid. Auditory feedback beeps were given after
each trial. Subjects performed four sessions of training in one stimulus position and then
transferred to the opposite position on the fifth session.

*Fitting data with a psychometric function*

Each subject’s accuracy at the six motion coherence levels across the two
attention conditions and five training sessions were independently fit with a Naka-
Rushton function of the form:

\[
R(C) = a \cdot \left( \frac{C^q}{C^q + C_{50}^q} \right) + b
\]  

(1)

where \( R \) is the response accuracy at each coherence level \( (C) \), \( a \) is the amplitude
parameter, \( C50 \) is the coherence at the half maximum of the response, \( q \) is the exponent
that determines the slope of the function, and \( b \) is the baseline. For our analyses, we let \( a \)
and \( C50 \) vary freely and kept \( b \) fixed at chance performance. Fit values for \( a \)
were

constrained such that the peak response could not exceed 100% accuracy. We also kept
the exponent, \( q \), fixed at 2, based on previous studies (see Herrmann et al., 2010;
Carandini and Heeger, 2012; Itthipuripat et al., 2014). The estimated \( C50 \) parameter was
taken to be the threshold performance for each attention condition in each session.
Results

The Naka-Rushton function captured the majority of the variance in behavior across cueing conditions and training sessions (Figure 2; feature: $R^2=0.79$, spatial: $R^2=0.80$, combined: $R^2=0.80$, neutral $R^2=0.77$, session 1: $R^2=0.78$, session 4: $R^2=0.79$, transfer session: $R^2=0.77$). A repeated-measures ANOVA revealed an effect of attention and training session (session 1-5) on estimated C50 thresholds (attention: $F_{(3,19)}=20.72$, $p<0.0001$; session: $F_{(4,19)}=9.13$, $p<0.0001$). There was also an interaction between attention and training session ($F_{(12,228)}=1.8$, $p<0.05$). Independent ANOVAs for learning and transfer showed a main effect of learning ($F_{(1,19)}=17.79$, $p<0.0006$), but not transfer ($F_{(1,19)}=0.79$, $p>0.3$). Figure 3 shows the C50 thresholds for sessions 1, 4, and the transfer session across cueing conditions. Both the learning and transfer comparisons showed an effect of attention condition (learning: $F_{(3,19)}=8.45$, $p<0.0002$; transfer: $F_{(3,19)}=8.98$, $p<0.0002$) and there was a marginal interaction between learning and attention ($F_{(3,57)}=2.62$, $p<0.06$). Planned t-tests revealed a significant effect of learning for the feature cued trials (Figure 4a; $t_{(19)}=4.11$, $p<0.0006$) and a significant effect of transfer for the feature cued trials (Figure 4b; $t_{(19)}=2.6$, $p<0.03$). There was also a significant effect of learning for the combined cued trials (Figure 4a; $t_{(19)}=2.25$, $p<0.04$). No other effects of learning or transfer were observed.

Discussion

Here, we show that the efficiency of feature-based attention can be improved by learning to discriminate a broad set of visual features, in this case motion direction, suggesting that the mechanisms supporting feature-based attention are flexible. Our
paradigm gives us the advantage of being able to disentangle feature learning from spatial learning and assess the impact of training on feature and spatial attention independently. Given that performance only improved on the feature cued and combined cued trials, learning on this type of broad feature discrimination task must be driven primarily by attentional modulation and not pure sensory improvements. Furthermore, our results also suggest that perceptual learning may be limited to the feature domain and that space-based sensory processing cannot be improved. The dependence of learning on this task on attentional cues provides support for models of perceptual learning in which top-down attention acts as a gating mechanism for learning (Ahissar and Hochstein, 1993, 2004; Herzog and Fahle, 1997; Fahle, 2004, 2009; Seitz and Watanabe, 2005; Ahissar et al., 2009). In line with the Reverse Hierarchy Theory (Ahissar and Hochstein, 2004; Ahissar et al., 2009), it appears that feature-based attention is recruited during learning in order to enhance the relevant motion directions and improve task performance. We expound upon this theory by suggesting that in cases of broader stimulus sets, it is more efficient to hone feature-based attentional modifications, rather than selectively enhancing specific stimulus representations (see Byers and Serences, 2012).

Importantly, learning to attend to feature transferred across spatial location, indicating that this type of feature-based attentional training is more generalizable than traditional perceptual learning (Karni and Sagi, 1991; Shiu and Pashler, 1992; Fahle, 1997; Hua et al., 2010). We suggest that this ability of feature learning to transfer across location is driven by an improvement in the efficiency of allocating feature-based attention, which has been shown to be spatially global (Treue and Martínez Trujillo, 1999; Saenz et al., 2002; Martinez-Trujillo and Treue, 2004; Maunsell and Treue, 2006).
This has important implications for perceptual learning because it shows that in cases of training with broader stimulus sets, it is more efficient to recruit top-down feature-based attention, rather than rely on bottom-up sensory changes (Crist et al., 1997, 2001; Gilbert et al., 2001; Bao et al., 2010; Hua et al., 2010). Furthermore, the transfer of feature-based attentional learning across spatial position suggests that the broader the training set, the more generalizable the learning, again suggesting that the processing level at which learning takes place modulates the extent of transfer (Ahissar et al., 2009; Pavlovskaya and Hochstein, 2011).

Although we did not employ a double-training or pre-test paradigm, our results complement this type of study, which implicate top-down attention as the moderator between learning in one location and transfer to another (Xiao et al., 2008; Zhang et al., 2010c, 2013; Wang et al., 2012). This work implicates attention as necessary to pair the more specific feature learning with more general decision rules (Xiao et al., 2008; Zhang et al., 2010b). They highlight the importance of top-down attention in selectively associating the learned features other stimulated (but untrained) features or locations. In our case, we suggest that this association is the result of refining feature-based attention mechanisms, which are inherently applied across retinotopic location (Saenz et al., 2002).

In conclusion, learning to attend improves the efficiency of feature-based attention, but not spatial attention, in a spatially global manner. This provides support for the hypothesis that the mechanisms of feature-based are flexible and that traditional studies of perceptual learning may be limited by only training one specific feature in a specific location. By using a broader training set, it becomes more optimal to recruit top-
down attention instead of relying on bottom-up sensory signals to perform the task, thereby allowing for more generalized learning benefits.
Figure 2.1: Schematic of the paradigm, exaggerated to display an example of each of the four cue types. The feature cue (upper left panel) was a red dot on the fixation ring that indicated the direction of the most coherent motion. The spatial cue (upper right panel) was a blue square corresponding to the RDK patch that contained the most coherent motion. The combined cue (lower left panel) used both the feature and spatial cues. The neutral cue (lower right panel) showed no information about the target RDK. Across all cueing conditions, the task was to determine which interval contained the RDK with the most coherent motion (black arrow). Each stimulus interval lasted 1s and was separated by a 1-s ISI. Subjects trained for four days in either the top or bottom location and transferred to the opposite location on the fifth day.
Figure 2.2: Fits of a Naka-Rushton function to the accuracy data from each of the 6 motion coherence levels in each cueing condition across training session. Note that the plots here show curves fit to the group average for display purposes. All analyses were conducted using individual fits for each subject. Error bars represent ±1SEM.
Figure 2.3: Estimates of the C50 parameter across training session and cueing condition. a) C50 grouped by session (day 1, day 4, transfer day) for each cueing condition: neutral (black), feature (blue), spatial (green), and combined (purple). There was a main effect of day and a main effect of attention condition (see text for details). b) Same data from (a) replotted and grouped by attention condition. Error bars represent ±1SEM.
Figure 2.4: a) Learning for each cueing condition, quantified as the C50 threshold on day 1 minus C50 threshold on day 4. There was a significant effect of learning for feature-cued trials and combined cue trials, but not for neutral or spatially cued trials. b) Transfer for each cueing condition, quantified as the C50 threshold on day 1 minus the C50 threshold in the transfer day. There was only a significant effect of transfer for the feature-cued trials. * indicates P < 0.05. Error bars represent ±1SEM.
Chapter 3: Enhanced Sensory Modulation in Human Early Visual Cortex Supports the Transfer of Perceptual Learning

Abstract

Recent work regarding the specificity of perceptual learning has shown that under certain training conditions the benefits of learning can transfer across spatial location and that this transfer is gated by top-down attentional control. However, the neural mechanisms supporting this transfer are unclear. Here, we investigate whether training on an orientation discrimination task transfers across spatial location and whether this transfer is supported by an improvement in the efficiency of sensory modulations associated with top-down attention. Subjects performed a pre-test on an orientation discrimination task at four different spatial locations, followed by a pre-training fMRI session, 10 training sessions at only one of the four locations, and finally a post-test and post-training fMRI session. We found that subjects learned on this task and that this learning transferred to the untrained spatial locations. Using an encoding model to assess the impact of attention and learning on the sensitivity of orientation-selective response profiles in early visual cortex, we found that learning was supported by an improvement in the efficiency of sensory modulations via greater suppression of non-target orientation signals after learning. This effect was observed for both when the trained and untrained quadrants were the target. These results suggest that transfer of learning is supported by enhanced sensory modulation via attentional selection.
Introduction

Perceptual learning and the accompanying neuronal changes have often been shown to be specific to the trained spatial location (Shiu and Pashler, 1992; Schoups et al., 1995, 2001; Crist et al., 1997; Gilbert et al., 2001; Furmanski et al., 2004; Yotsumoto et al., 2008; Jehee et al., 2012). In particular, Schoups and colleagues (2001) presented macaque monkeys with two different oriented gratings and trained them to perform a fine orientation discrimination task with only one grating while the other was passively presented. Over the course of learning, the monkeys’ threshold orientation offsets consistently decreased, but learning was specific to both the trained orientation and stimulus location. This behavioral improvement corresponded to an increase in the sensitivity of the neurons that preferred the trained orientation. There was no change in the sensitivity of the neurons that corresponded to the passively presented grating, nor to any untrained orientations, leading the authors to conclude that behavioral specificity results from specific changes in the firing rates of early sensory populations. More recently, a study employed functional magnetic resonance imaging (fMRI) and an orientation discrimination task to examine the effects of learning on the ability of voxels in early visual areas to discriminate the trained orientation (Jehee et al., 2012). They found that training led to an improvement in orientation discrimination, but only at the trained orientation and in the trained location. Although there was no mean change in the blood oxygenation level dependent (BOLD) responses after learning, using a signal detection decoding model, they found that voxels in visual areas V1-V4 were better able to discriminate the presented orientation on a given trial. Again, this improvement was limited to the trained orientation in the trained location.
However, there is psychophysical evidence that learning is not always specific to the trained location and can transfer across spatial position under the right training conditions (Xiao et al., 2008; Jeter et al., 2009, 2010; Lu et al., 2010; Zhang et al., 2010c, 2013; Pavlovskaya and Hochstein, 2011; Wang et al., 2012). One study showed that having a brief pre-test at a peripheral to-be-transferred location facilitated the transfer of learning on a foveal orientation discrimination task to that untrained location, as well as across other peripheral locations (Zhang et al., 2010c). Other work by this group has shown that by employing a “double-training” or “training-plus-exposure (TPE)” paradigm also promoted transfer of learning across spatial locations by facilitating the functional association of higher-level decision rules with the stimulated early sensory regions (Xiao et al., 2008; Wang et al., 2012; Zhang et al., 2013). Using event-related potentials (ERPs) and a vernier acuity task, the authors suggest that transfer of learning is associated with an enhancement of top-down attentional signals via an amplitude increase in the attention-related N1 component (Zhang et al., 2013). Specificity, on the other hand, is related to a decrease in the amplitude of the N1 component, indicating that learning specificity may be the result of suppression of these top-down attentional connections with early sensory areas when a single stimulus at a single location is over trained.

Despite the psychophysical and theoretical evidence of transfer, there is little neuronal data regarding the mechanism supporting the transfer of enhanced visual sensitivity from one retinal location to another. The aforementioned fMRI and ERP studies (Jehee et al., 2012; Zhang et al., 2013) seem to be in opposition to each other, but since one does not employ the TPE procedure and the other cannot provide insight into changes in orientation sensitivity that may be associated with learning and transfer, it is
difficult to draw conclusions about the effects of learning on orientation sensitivity at both the trained location and stimulated, but untrained locations. Additionally, much of the neural evidence in favor of learning specificity comes either from single neurons or coarse changes in BOLD signal and therefore cannot provide insight into how learning and transfer impacts sensory sensitivity at the population level (Schoups et al., 2001; Furmanski et al., 2004; Yotsumoto et al., 2008). Here, we employ an orientation discrimination task in which subjects are pre- and post-tested at four possible locations and then trained at only one location while non-target gratings are presented in the three remaining locations. Using fMRI and an encoding model, we test whether learning on this task is supported by an increase in the sensitivity of orientation-selective populations in retinotopic regions of early visual cortex and whether this increase in sensitivity transfers across spatial location.

Methods

Fifteen right-handed male subjects (mean age: 22.9) from University of California, San Diego were recruited and were given written informed consent in accordance with UCSD’s Institutional Review Board. Of these, five were not included in the analyses: two dropped out before completion, one showed excessive motion in the fMRI scanner, and two could not be tracked by the eyetracker. Subjects completed a 30-minute pre-test, a 2-hour pre-training fMRI scan, 10 1-hour training sessions, a 30-minute post-test, and a final 2-hour post-training fMRI scan. Subjects’ eyes were tracked during all behavioral sessions and a gaze-contingent display was used, such that each trial would not begin until the subject was fixating a central bull’s-eye and the trial would abort and
restart if the subject broke fixation while the stimulus was up. Subjects were paid $10/hour for the behavioral sessions and $20/hour for the fMRI sessions.

All stimuli were created using MATLAB (v.7.8, Natick, MA) with the Psychophysics Toolbox (v.3, Brainard, 1997; Pelli, 1997). The stimulus was comprised of four full contrast sinusoidal gratings, each spanning 5 degrees of visual angle in diameter with a spatial frequency of 3 cycles per degree (Figure 1). Each grating was offset from a central fixation bull’s-eye by 8 degrees of visual angle. On each trial, a red dot at fixation indicated which of the four gratings the subject should attend on that trial. During the pre-test, post-test, and pre- and post-training scans, each of the four gratings was cued equally often. During the training sessions the same grating was cued on every trial; the trained location was varied between subjects. The task was to determine whether the attended grating was shifted clockwise or counterclockwise from an imagined reference angle at either 30° or 120°. The reference angle was varied between subjects. During the pre- and post-test sessions, the offset between the attended angle and the reference angle was staircased independently for each grating, in order to get a measure of the threshold offset before and after training at each position. The staircasing was a 2-up 1-down procedure that included an increasing staircase that began at 4° offset and a decreasing staircase that began at 15° offset. The size of the offset changed by 90% of the current offset for both increasing and decreasing trials. The same staircasing procedure was used for the 10 training sessions. During the fMRI sessions the offset between the attended grating and the reference angle varied randomly between 2°, 5°, or 8° on each trial, in order to match the stimulus display before and after training. Each trial began with a 500-msec fixation, followed by the presentation of the cue and the
gratings for 2000 msec. Subjects had an additional 2000 msec to respond by pressing buttons with both middle fingers simultaneously for a clockwise rotation or both index fingers for a counterclockwise rotation. During the pre-test, post-test, and training sessions, subjects were given feedback for 100 msec after each trial. There was a 900 msec inter-trial interval ITI before the next fixation. Subjects completed 320 trials (80 at each quadrant) during the pre-test and post-test sessions and eight blocks of 80 trials during the training sessions. During the fMRI sessions, three subjects performed five blocks of 57 trials (9 nulls) and the remaining seven subjects performed ten blocks of 57 trials (9 nulls).

Functional localizer scans

In addition to the main experimental runs, subjects also performed 1 or 2 4.8-min runs of a functional localizer during each scanning session. The localizer scans were designed to independently drive voxels in retinotopically organized early visual cortex that corresponded to the same spatial positions of each of the four gratings in the main experimental task. The stimulus was comprised of two contrast-reversing checkerboards (spatial frequency = 3 cycle/°, temporal frequency = 8 cycles/second) that occupied two of the four grating positions across the diagonal (i.e. upper left and lower right or upper right and lower left). The checkerboards appeared for 10 s and subjects were instructed to maintain fixation while monitoring for a brief, 66.7-ms contrast-dimming that occurred 4 times per trial. The occurrence of each target was determined pseudo-randomly, such that the target appeared at least 1s after stimulus onset and 1s before stimulus offset and was separated from the previous target by at least 1s. Subjects performed a total of 28 trials per run.
Retinotopic mapping

A standard meridian mapping procedure was used to identify visual areas V1, V2v/V2d, and V3v/V3d. Subjects passively viewed a contrast-reversing checkerboard stimulus flickering at 8Hz and covering 60° of polar angle and that alternated every 20s between the horizontal and vertical meridians (Engel et al., 1994; Sereno et al., 1995). A standard general linear model (GLM) with regressors corresponding to epochs of vertical meridian and horizontal meridian stimulation was used to identify voxels in visual cortex that responded to each respective region of the visual field. Each regressor was generated by convolving a boxcar model of each stimulus timeseries (‘0’ if no stimulus, ‘1’ if stimulus present) with a standard difference-of-two gamma function model (time to peak of positive response: 5s, time to peak of negative response: 15s, ratio of positive and negative responses: 6, positive and negative response dispersion: 1). The retinotopic mapping data were then projected onto a computationally inflated rendering of each subjects’ gray/white matter boundary to facilitate the identification of the horizontal and vertical meridians in order to define each cortical area.

fMRI data acquisition and pre-processing

Scanning was performed using a 3T GE MR750 MRI scanner with a 16-channel head coil at UCSD’s Keck Center for Functional MRI. Anatomical images were collected using a T1-weighted sequence that produced images with a 1mm³ resolution (TR/TE=11/3.3 ms, TI=1100ms, 172 slices, flip angle=18°). Functional images were collected using a gradient echo EPI pulse sequence with 35 horizontal slices, which covered visual cortex. The slices were collected in ascending interleaved order and were 3mm thick (TR = 2000 ms, TE = 30ms, flip angle = 90°, image matrix = 64 (AP) × 64
(RL), with FOV = 192mm (AP) × 192mm (RL), voxel size = 3mm × 3mm × 3mm). All EPI images were slice-time corrected, motion-corrected (both within and between scans), and high pass filtered (3 cycles/run). BrainVoyager QX (v2.3; Brain Innovation, Maastricht, The Netherlands) was used to perform the data analysis in conjunction with custom analysis scripts written in MATLAB (version 7.11.0584; The Math Works, Natick, Massachusetts).

Analysis of functional localizer data

Data from the functional localizer scans were analyzed using a GLM that contained separate regressors marking the temporal interval in which each of the two stimulus positions was presented (upper left/lower right and upper right/lower left). Each regressor was generated by convolving a boxcar model of the stimulus timeseries (‘0’ if no stimulus present, ‘1’ if stimulus present) with a standard difference-of-two gamma function model (time to peak of positive response: 5s, time to peak of negative response: 15s, ratio of positive and negative responses: 6, positive and negative response dispersion: 1). A contrast was then performed to identify voxels that responded more to the upper left-lower right than the upper right-lower left and vice versa. These contrasts were used to identify voxels in visual cortex corresponding to each of the four grating locations. The regions of interest (ROIs) from the localizer runs were then combined with the retinotopically defined visual areas (V1-V3v/d) in order to create separate ROIs for each quadrant in each visual area (upper left V1, V2, V3; lower left V1, V2, V3; upper right V1, V2, V3; lower right V1, V2, V3). Voxels within each of the visual areas were included in subsequent analyses if they passed a threshold of P < 0.05, after using
BrainVoyager’s False Discovery Rate (FDR) algorithm to correct for multiple comparisons.

*Estimating trial-by-trial BOLD responses*

Before estimating the trial-by-trial magnitude of the BOLD response during the main experimental task, we first estimated the shape of the hemodynamic response function (HRF) separately for each visual area quadrant in each subject. This subject and area specific fitting was done because the shape of the HRF can vary substantially across subjects and brain regions, as well as to allow for comparisons across quadrants (Zarahn et al., 1997; Gonzalez-Castillo et al., 2012). Data were averaged across all trials in the main scans and the mean stimulus-locked HRF was computed separately for each visual area in each subject across a 24s window using a Finite Impulse Response model (Dale, 1999). We then used a gradient descent algorithm implemented in MATLAB to fit each HRF estimate using a standard difference-of-two gamma function model with time to peak, onset time, and the ratio of positive and negative components as free parameters (using the same seed parameters listed in previous section). The parameters that yielded the best fit were used to produce a custom HRF model for each visual area in each subject. We then created a design matrix with a boxcar model of the onset/offset time of each 2s stimulus epoch on each trial within a scan, yielding 48 regressors of interest (12 instances of each quadrant as the target and 36 instances of that quadrant as a non-target) along with 1 additional regressor as a constant term. Each boxcar regressor model was then convolved with the custom HRF estimate for that subject, and a GLM was used to estimate the relative amplitude of the BOLD response on each trial. This estimate of
response amplitude one each trial in each voxel was then used as input to the orientation encoding model described below.

\textit{Estimating orientation-selective BOLD tuning profiles using a forward encoding model}

Forward encoding models adopt a set of \textit{a priori} assumptions about the important features or stimulus labels that can be distinguished using hemodynamic signals within an ROI (Brouwer & Heeger, 2009, 2011; Dumoulin & Wandell, 2008; Gourtzelidis et al., 2005; Kay & Gallant, 2009; Kay, Naselaris, Prenger, & Gallant, 2008; Mitchell et al., 2008; Naselaris, Prenger, Kay, Oliver, & Gallant, 2009; Schönwiesner & Zatorre, 2009; Scolari, et al., 2012; Thirion et al., 2006; reviewed in Naselaris, Kay, Nishimoto, & Gallant, 2011; Serences & Saproo, 2012). The features or labels in the model are then used to predict the pattern of BOLD responses. Because an increase in the information content of a response pattern might be supported by various types of modulation (i.e. a decrease in the bandwidth of orientation-selective population tuning profiles, an increase in gain of orientation-selective population tuning profiles, etc.), we directly reconstruct the population tuning profiles using an a priori encoding model. This approach allows us to quantify the modulatory patterns that occur as a result of learning and the extent to which these modulations are limited to the trained quadrant compared to the untrained quadrants.

We used a forward model adapted from Brouwer and Heeger (2009, 2011) to estimate the response across a series of orientation channels as a function of learning and the focus of attention. The model assumes that the BOLD response in a given voxel reflects the pooled activity across a large population of orientation-selective neurons, and that the distribution of neural tuning preference is biased within a given voxel either due
to large-scale feature maps (Freeman et al., 2011) or to subvoxel anisotropies in cortical columns (Kamitani and Tong, 2005; Swisher et al., 2010). In either case, BOLD responses measured from many voxels in early visual cortex exhibit a modest but robust orientation preference (Haynes and Rees, 2005; Kamitani and Tong, 2005; Serences et al., 2009; Freeman et al., 2011).

For each subject and scan session, we first split the data into two sets (training and test sets), using a hold-one-out cross-validation method. Adopting the terminology and formulations of Brouwer and Heeger (2009; 2011) for consistency, let $m$ be the number of voxels in a given visual area, $n_1$ be the number of observations (trials) in the training set, $n_2$ be the number of trials in the test set, and $k$ be the number of hypothetical orientation channels. Let $B_1$ ($m \times n_1$ matrix) be the training set, and $B_2$ ($m \times n_2$ matrix) be the test set. The training data in $B_1$ were then mapped onto the matrix of hypothetical channel outputs ($C_1, k \times n$) by the weight matrix ($W, m \times k$) that was estimated using a linear model of the form:

$$B_1 = WC_1. \quad (1)$$

where the ordinary least-squares estimate of $W$ is computed as:

$$W = B_1 C_1^T (C_1 C_1^T)^{-1}. \quad (2)$$

The channel responses ($C_2, k \times n_2$) were then estimated for the test data ($B_2$) using the weights estimated in (2):
\[ \tilde{c}_2 = (\tilde{W}^T \tilde{W})^{-1} \tilde{W}^T \tilde{b}_2. \] (3)

Equations 1-2 are similar to a traditional univariate GLM in that each voxel gets a weight for each feature in the model (in this case, one weight for each orientation channel). Equation 3 then estimates channel responses on each trial based on the estimated weights assigned to each voxel and the vector of responses observed across all voxels on a given trial in the test set.

The construction of the basis set matrix \( C \) has a large impact on the resulting channel response estimates. In the present experiment, we used a half-sinusoidal function that was raised to the 5\textsuperscript{th} power to emulate the approximate shape of single-unit tuning functions in V1, where the average 1/\sqrt{2} half-bandwidth of orientation tuned cells is approximately 20° (Schiller et al., 1976; Ringach et al., 2002; Gur et al., 2005). This function was then replicated 9 times, with the 9 copies evenly distributed across orientation space. The power of the sinusoidal function necessitated the use of at least 6 copies in order to adequately cover orientation space (Freeman & Adelson, 1991). We selected the shape of the basis function based on existing physiology studies, however, all results that we report are robust to reasonable variations in this value (i.e. raising the half-sinusoid to the 5\textsuperscript{th}, 6\textsuperscript{th}, 7\textsuperscript{th} or 8\textsuperscript{th} power, all of which are reasonable choices based the large range of single-unit bandwidths observed in early visual areas). The analysis described above was then repeated with the center position of each of the 9 basis functions shifted across all 180° of orientation space in 1° increments to generate channel tuning functions with 1° resolution (which ultimately produced a 180° tuning function). Note that because of overlap between adjacent basis functions, each point along these 180-point tuning
curves was not independent from neighboring points. However, this approach was adopted to maximize the smoothness of the orientation tuning functions, which in turn facilitates quantifying the amplitude, bandwidth, and baseline level using the model fitting approach described below. Lastly, the 180-point channel response function estimated on each trial was circularly shifted so that the channel matching the orientation of the stimulus in the quadrant of interest on that trial was positioned in the center of the tuning curve, thereby aligning all channel response profiles to a common stimulus-centered reference frame (which is plotted as 0° on the x-axis by convention). The channel responses for each subject were then averaged across subjects within each scanning session so that group channel response functions for the different locations, attention, and training conditions could be compared.

Quantifying channel responses by fitting a von Mises function

The re-centered channel responses for each condition (each quadrant as target vs non-targets, pre-training/post-training, and trained vs untrained quadrant as, pre-/post-training) were independently fit with a von Mises function of the form:

\[ f(\theta) = a e^{\sigma (\cos(\mu - x)) - 1} + b \]  

with amplitude \( a \), mean \( \mu \), bandwidth \( \sigma \), and baseline \( b \) as independent free parameters that reflect distinct attributes of the function. The mean was restricted to a range from 80° and 100° and the bandwidth was restricted to a range between 0.5 and 12. Because of the high between subject variability and the added noise of having four simultaneously presented stimuli, we used resampling with replacement across the 10
subjects’ data and fit the average with the von Mises function (eq 4); this process was repeated 1000 times. During the fitting, we iteratively stepped through each of the possible bandwidth values in steps of 0.5. On each iteration, the amplitude and baseline were estimated by solving the linear systems equation necessary to map the predicted function onto the data using a linear model. The set of parameters for each subject that yielded the lowest root mean squared error across the all iterations were then used in subsequent analyses. In order to make statistical comparisons for the estimated fit parameters, we evaluated the differences across attention and learning conditions for each resampling iteration, giving us a distribution of effects that we could compare against 0. We defined P as the probability of observing an effect in the opposite direction of the observed data (multiplied by 2, as we performed two-tailed tests).

Results

Behavior

Irrespective of training quadrant, there was a significant decrease in threshold offset across the four quadrants from pre-test (6.1°) to post-test (5.2°; t(1,9) = 6.36, P < 0.0002). The threshold decrease in the trained quadrant from pre-test (6.6°) to post-test (5.7°) was only marginal (t(1,9) = 2.11, P = 0.06); however, there was a significant decrease in threshold across the 10 training sessions (day 1: 4.7°, day 10: 3.8°; repeated-measures ANOVA: F(9,81) = 8.8, P < 0.000001). Because orientation offset was fixed during the scanning sessions, we compared accuracy before and after training as a measure of performance improvement. There was no change in accuracy when disregarding which quadrant was the trained quadrant (pre-training: 76%, post-training:
80%; \( t(1,9) = -1.49, P = 0.17 \). However, there was a significant increase in accuracy for the trained quadrant after training (pre-training: 73%, post-training: 80%; \( t(1,9) = -2.45, P < 0.04 \)).

**fMRI Results**

We first estimated the BOLD HRFs associated with each quadrant when that quadrant was a target, non-target, and when it was the trained quadrant (Figures 3-4; see Methods). There was a main effect of time point in the HRF (3-way repeated-measures ANOVA, \( F(12,345) = 10.38, P < 0.00001 \)), but no other effects of attention, learning, or trained quadrant (attention: \( F(1,9) = 0.02, P = 0.89 \); learning: \( F(1,9) = 1.11, P = 0.32 \); train/untrain as target: \( F(1,9) = 0.62, P = 0.45 \); learning*attention: \( F(1,9) = 0.83, P = 0.36 \); learning*train/untrain \( F(1,9) = 1.47, P = 0.22 \)).

We then used an encoding model (see Methods) to examine the relationship between attention, trained quadrant, and learning on the sensitivity of orientation-selective populations in early visual cortex. Across all conditions, the orientation-selective response profiles peaked at the target orientation and fell off in a Gaussian-like manner (see Figures 5-6). After fitting each tuning curve with a von Mises function (see Methods), we compared the best-fitting parameters for each attention and training condition. In early visual cortex (V1-V3v/d), there was a main effect of attention (target versus non target, irrespective of trained quadrant, see Figure 7) on the amplitude, bandwidth, baseline, and mean of the orientation-selective response profiles (all \( P < 0.001 \), resampling test, see Methods). There was an effect of learning on the mean of the target orientation response profiles (\( P < 0.04 \)), but none of the other parameters.

Importantly, there was an effect of learning on the amplitude and baseline of the non-
target response profiles, such that the overall non-target response profiles became less selective after training (both $P < 0.04$). When looking at the trained quadrant across attention and learning conditions, we found the same effect of attention on all four fit parameters (see Figure 8, $P < 0.001$). There was no effect of learning on any of the trained target orientation-selective responses, but there was an effect of learning on the baseline of the non-target responses, such that the overall non-target response profiles became less selective after training ($P < 0.05$).

**Discussion**

The present data show that learning on a fine orientation discrimination task transfers across spatial position, as characterized by a reduction in orientation discrimination thresholds across all four quadrants. This replicates previous findings showing that that a pre-test or TPE at the to-be-transferred location enable transfer of learning across spatial location (Xiao et al., 2008; Zhang et al., 2010b, 2010c; Wang et al., 2012; Zhang and Yang, 2013). We also show that learning is supported by an improvement in the efficiency of feature-based attention, such that after learning there is an overall decrease in the sensitivity of orientation-selective response profiles for non-target gratings. Interestingly, this effect of non-target suppression occurred both when the trained quadrant was the target and when the analysis was agnostic to training quadrant. We take this to mean that training refines top-down attention in a general manner that can be applied equally to trained and untrained targets. This finding, as well as the lack of change in the HRFs after learning, is consistent with the results of Byers and Serences (*in press*), suggesting that learning is at least in part supported by greater suppression of
task-irrelevant orientation information but not overall BOLD responses (see also, Yotsumoto et al., 2008; Zhang et al., 2010a; Jehee et al., 2012). Additionally, these data are in accord with previous studies showing the context-dependence of changes in sensory processing after learning (Crist et al., 2001; Li et al., 2004; Vidnyánszky and Sohn, 2005; Gál et al., 2009; Mukai et al., 2011; Piëch et al., 2013).

Furthermore, both theoretical and experimental studies on transfer implicate top-down attentional mechanisms in facilitating transfer across spatial location, stimulus feature, and task type (Ahissar and Hochstein, 1997, 2004; Yu et al., 2004; Dosher and Lu, 2007; Xiao et al., 2008; Zhang et al., 2008, 2010b; Ahissar et al., 2009; Jeter et al., 2009; Baeck and Op de Beeck, 2010; Pavlovskaya and Hochstein, 2011; Wang et al., 2013; McGovern et al., 2012; Wang et al., 2012; Zhang and Yang, 2013; Dosher et al., 2013). Specifically, the Reverse Hierarchy Theory of perceptual learning proposes that top-down attention serves as a gating mechanism that selects which sensory populations should be recruited during training in order to support efficient learning (Ahissar and Hochstein, 2004; Ahissar et al., 2009). An implication of this selection is that when task demands drive higher-level populations, learning should be more transferrable across common feature dimensions and less retinotopic (Pavlovskaya and Hochstein, 2011). Although our task does not necessarily recruit higher sensory areas, it appears that the most efficient learning mechanism under these conditions was to more selectively filter incoming sensory inputs, which is modulated by top-down attentional control. Therefore, when interpreting our data within the Reverse Hierarchy framework, it follows that the recruitment and refining of attentional filtering mechanisms should not be limited by spatial location (Ahissar et al., 2009; Byers and Serences, 2012). These theories, in
conjunction with studies showing that transfer results from an association of a learning rule to the sensory signals from untrained stimuli and that the strength of top-down attentional signals modulates the degree of transfer, the present results seem to indicate that some cases of perceptual learning improve upon attentional mechanisms and are not be limited to specific sensory changes (Xiao et al., 2008; Zhang et al., 2010c, 2013).

Finally, it is important to keep in mind that the current data are quite noisy and the impact of presenting four competing stimuli while collecting BOLD data is unclear. Thus, these effects should be examined further with the collection of additional subjects in order to thoroughly understand the extent to which enhanced attentional selectivity promotes the transfer of learning. However, as the data currently stand, it appears that learning-related increases in the efficiency of top-down attention via sensory modulation are spatially global.
Figure 3.1: Depiction of the stimulus design. Subjects fixated a central bull’s-eye and a red cue indicated which of the four gratings they should attend on that trial. During pre- and post-test sessions, each position was cued equally often. During the training sessions, the same position was cued on every trial. Target gratings were offset either clockwise or counter-clockwise from 30° or 120° (counterbalanced between subjects). Non-target gratings could be any orientation between 0-180°. See text for details.
Figure 3.2: Orientation discrimination thresholds. a) Threshold offset for the target quadrant (irrespective of trained quadrant) from the pre-test and post-test, respectively. b) Threshold offset for the trained quadrant when it was the target from the pre-test and post-test, respectively. c) Threshold offsets across the 10 training sessions. All error bars are ±1SEM.
Figure 3.3: Mean hemodynamic response functions (HRFs) computed across V1-V3 for the target quadrant (blue) compared to non-target quadrants (green). The left panel shows the HRFs before training and the right panel shows the HRFs after training. Error bars are ±1SEM.
Figure 3.4: Mean hemodynamic response functions (HRFs) computed across V1-V3 for the trained quadrant as the target (purple) compared to non-target quadrants (green). The left panel shows the HRFs before training and the right panel shows the HRFs after training. Error bars are ±1SEM.
Figure 3.5: Orientation-selective response functions computed across V1-V3 and averaged across subjects, centered on the target orientation on a given trial. The left panel shows pre-training tuning for the target grating (blue) compared to the non-targets (green). The right panel shows the post-training tuning curves for the same two conditions. Error bars are ±1SEM.
Figure 3.6: Orientation-selective response functions computed across V1-V3 and averaged across subjects, centered on the target orientation on a given trial. The left panel shows pre-training tuning when the trained quadrant was the target (purple) compared to non-targets (green). The right panel shows the post-training tuning curves for the same two conditions. Error bars are ±1SEM.
Figure 3.7: Parameter estimates from the best fitting von Mises function to the orientation-selective response profiles for the target quadrant (blue bars), irrespective of which quadrant was trained, and the non-target quadrants (green bars), before (darker) and after (lighter) training. The upper left panel shows the estimates for amplitude, the upper right panel shows the estimates for bandwidth, the lower left panel shows the estimates for baseline, and the lower right panel shows the estimates for mean. Black * indicate significant effect of attention. Green * indicate significant effect of training on non-target parameters. Error bars represent 95% confidence intervals on resampled data (see Methods).
Figure 3.8: Parameter estimates from the best fitting von Mises function to the orientation-selective response profiles for the trained quadrant as the target (purple bars) and the non-target quadrants (green bars), before (darker) and after (lighter) training. The upper left panel shows the estimates for amplitude, the upper right panel shows the estimates for bandwidth, the lower left panel shows the estimates for baseline, and the lower right panel shows the estimates for mean. Black * indicate significant effect of attention. Green * indicate significant effect of training on non-target parameters. Error bars represent 95% confidence intervals on resampled data (see Methods).
GENERAL DISCUSSION

We have proposed that the substantial overlap between the behavioral and neural effects of both top-down feature-based attention and perceptual learning implicate a possible common mechanism that is recruited for each of these types of perceptual enhancement (Byers and Serences, 2012). Both feature-based attention and perceptual learning have been shown to improve perceptual sensitivity via improved sensory modulation (Desimone and Duncan, 1995; Treue and Martínez Trujillo, 1999; Reynolds et al., 2000; Schoups et al., 2001; Ghose et al., 2002; Yang and Maunsell, 2004; Martinez-Trujillo and Treue, 2004; Mitchell et al., 2009, 2007; Cohen and Maunsell, 2009, 2011; Adab and Vogels, 2011; Cohen and Kohn, 2011; Gu et al., 2011; Jehee et al., 2012) and enhanced read-out at the level of decision-making (Dosher and Lu, 1999; Palmer et al., 2000; Law and Gold, 2008, 2009; Palmer and Moore, 2009; Pestilli et al., 2011; Dosher et al., 2013). These commonalities suggest that under the appropriate training conditions, feature-based attention could be recruited during perceptual learning in order to improve the efficiency of these top-down modulations and to support more generalized learning. Given that feature-based attention is spatially global (Treue and Martínez Trujillo, 1999; Saenz et al., 2002; Martinez-Trujillo and Treue, 2004; Maunsell and Treue, 2006), improvements in the allocation of this type of attention should transfer across spatial location. Here, we first investigate whether feature-based attention can be made more efficient with training and if this type of learning is supported by an improvement in sensory modulation (Chapter 1), then if training feature-based attention transfers across spatial location (Chapter 2), and finally, whether training on a task that recruits top-down attention can transfer across spatial location through improved sensory
modulation (Chapter 3).

In Chapter 1, we show that by training subjects on a more generalized perceptual learning task, they improve on their ability to make fine orientation discriminations with one of nine unique exemplars. This type of learning is supported by an enhancement in the efficiency of feature-based attention through improved sensory modulations. By amplifying the difference in the amplitude of orientation-selective response profiles when orientation was task-relevant compared to task-irrelevant, subjects were able to more selectively attend to orientation. These data corroborate sensory modulation accounts of perceptual learning, but do not rule out the possibility of enhanced read-out at a later decision-making stage (Schoups et al., 2001; Ghose et al., 2002; Furmanski et al., 2004; Law and Gold, 2008, 2009; Jehee et al., 2012; Dosher et al., 2013). Importantly, Chapter 1 also shows that mere changes in overall BOLD signal are not enough to account for the effects of learning and that orientation information coded by early sensory populations can be improved with training (e.g., Yotsumoto et al., 2008).

Furthermore, the results of Chapter 1 provide support for the role of attention as a modulator of perceptual learning, which adaptively select the level at which learning takes place (Ahissar and Hochstein, 1993, 2004; Ahissar et al., 2009). In conjunction with other work showing that perceptual learning refines attentional mechanisms by enhancing relevant information and suppressing irrelevant information, these data lay the foundation for Chapter 2 by demonstrating the plasticity of top-down attention (Crist et al., 2001; Gilbert et al., 2001; Li et al., 2004; Vidnyánszky and Sohn, 2005; Gilbert and Sigman, 2007; Gál et al., 2009). Because feature-based attention can be made more
efficient through learning, it follows that training feature-based attention may allow for more transferrable learning-related perceptual benefits.

Chapter 2 applied the findings of flexible feature-based attention by comparing the impact of feature cues and spatial cues on perceptual learning and the extent of its transfer. In this chapter, we show that feature-based attention, but not spatial attention, modulates perceptual learning. Taken in conjunction with the findings of Chapter 1, these results suggest that the mechanisms of feature-based attention can be improved with training, allowing for greater sensitivity to the attended feature. This type of learning implicates top-down attentional modulation as a gating mechanism for the acquisition of perceptual learning, supporting previous work suggesting that attention selects the level at which learning occurs (Ahissar and Hochstein, 1993, 2004; Crist et al., 2001; Gilbert et al., 2001; Li et al., 2004; Ahissar et al., 2009; Fahle, 2009). Given that no learning was observed for spatially cued or neutral trials, these data also support an account of learning that is context-dependent, such that improved perceptual sensitivity is yoked to the relevance of the feature signals to current task demands (Gilbert et al., 2000; Crist et al., 2001; Li et al., 2004; Piëch et al., 2013).

Moreover, the benefits of learning to attend to feature are not limited to the trained spatial location and transfer to untrained locations. Spatial transfer of feature learning provides support for the hypothesis that training feature-based attention leads to generalizeable perceptual sensitivity improvements because feature-based attention is spatially global (Treue and Martínez Trujillo, 1999; Saenz et al., 2002; Martinez-Trujillo and Treue, 2004; Maunsell and Treue, 2006). This type of learning paradigm, in which multiple features are trained concurrently rather than hyper-training a single feature in a
single location, promotes more general perceptual learning, which is less limited in its applications to real-world examples of perceptual expertise.

Chapter 3 replicates the finding of learning-induced improvements in sensory modulation from Chapter 1, adding that transfer is also supported by this type of gain modification. In this chapter, we show that perceptual learning is supported by an enhancement in the efficiency of top-down attentional modulations on orientation-selective responses in early visual cortex through greater suppression of task-irrelevant orientation signals after training. This improvement in non-target suppression occurred both for when the trained quadrant was the target and when any of the four quadrants was the target, suggesting that learning and transfer result from the refinement of top-down attentional control mechanisms that filter incoming sensory inputs. Chapter 3 also corroborates previous studies showing that both a pre-test and exposure at the to-be-transferred locations (or features) promote transfer of learning via top-down attentional modulations facilitating the association of learning rules with those sensory inputs (Xiao et al., 2008; Zhang et al., 2008, 2010b, 2010c, 2013; Wang et al., 2012). This study expands upon previous work by showing that the nature of the attentional modulations that support transfer is one of more efficient modulation of task-irrelevant sensory signals.

Taken together, Chapters 1-3 show that top-down attention can be sharpened through perceptual learning, primarily via an enhancement in the efficiency of sensory modulation administered by feature-based attentional mechanisms, and that this learning is non-specific. Although the both the behavioral effects and mechanisms of perceptual learning are diverse, this dissertation shows that some instances of learning can be
explained by an improvement in the allocation of top-down feature-based attention (Dosher and Lu, 1999; Schoups et al., 2001; Furmanski et al., 2004; Law and Gold, 2008; Ahissar et al., 2009; Seitz and Watanabe, 2009; Seitz et al., 2009; Fahle, 2009; Byers and Serences, 2012; Gilbert and Li, 2012; Leclercq and Seitz, 2012; Dosher et al., 2013). This work has important implications for the generalizeability of learning and its applications to real-world instances of enhanced perceptual and attentional capabilities. Learning to attend allows for both more efficient use of neural resources and less specific perceptual benefits, making this type of learning much more practical than traditional perceptual learning. Still, much work needs to be done to fully understand the extent of learning to attend and whether other attentional mechanisms, such as improved read-out, are also modifiable through training.


