# Neural Mechanisms of Feature Conjunction Learning: Enduring Changes in Occipital Cortex After A Week of Training

Sebastian M. Frank,[1,2][†] **Eric A. Reavis,**[2][†] **Peter U. Tse,**[2]
and Mark W. Greenlee[1]*

[1]*Institute for Experimental Psychology, University of Regensburg, Regensburg, Germany*
[2]*Department of Psychological and Brain Sciences, Dartmouth College, Hanover, New Hampshire*

◆━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━◆
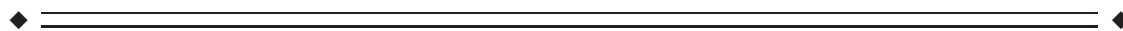
Abstract: Most visual activities, whether reading, driving, or playing video games, require rapid detection and identification of learned patterns defined by arbitrary conjunctions of visual features. Initially, such detection is slow and inefficient, but it can become fast and efficient with training. To determine how the brain learns to process conjunctions of visual features efficiently, we trained participants over eight consecutive days to search for a target defined by an arbitrary conjunction of color and location among distractors with a different conjunction of the same features. During each training session, we measured brain activity with functional magnetic resonance imaging (fMRI). The speed of visual search for feature conjunctions improved dramatically within just a few days. These behavioral improvements were correlated with increased neural responses to the stimuli in visual cortex. This suggests that changes in neural processing in visual cortex contribute to the speeding up of visual feature conjunction search. We find evidence that this effect is driven by an increase in the signal-to-noise ratio (SNR) of the BOLD signal for search targets over distractors. In a control condition where target and distractor identities were exchanged after training, learned search efficiency was abolished, suggesting that the primary improvement was perceptual learning for the search stimuli, not task-learning. Moreover, when participants were retested on the original task after nine months without further training, the acquired changes in behavior and brain activity were still present, showing that this can be an enduring form of learning and neural reorganization. *Hum Brain Mapp 35:1201–1211, 2014.* © 2013 **Wiley Periodicals, Inc.**

Key words: attention; fMRI; perceptual learning; visual cortex; visual search; neural plasticity

◆━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━◆

INTRODUCTION

To quickly and accurately extract biologically important information from patterns of light on the retina, the human visual system must be able to learn relevant patterns of features. Such "perceptual learning" typically occurs with repeated exposure to a particular stimulus pattern, and leads to improved speed and accuracy in processing the stimulus [Gibson, 1963; Goldstone, 1998; Seitz and Watanabe, 2005].

Perceptual learning is well documented for single visual features [e.g., Vernier offset, motion direction, texture; see Ball and Sekuler, 1982; Karni and Sagi, 1991; Poggio et al., 1992]. However, in ecological vision, single visual features may prove insufficient to guide behavior effectively. For example, foraging solely for things that are red might turn up both poisonous and edible berries; foraging for red objects of a particular shape would be more adaptive. Similarly, there may be many yellow things and many moving things in the environment, but identifying a lion might require the ability to process a particular conjunction of yellow and motion. A capacity for perceptual learning of conjunctions of visual features (e.g., color and shape, or color and motion), particularly in the context of visual search, would confer a survival advantage because it would afford the rapid recognition of items of high biological relevance.

In visual search, the relationship between the number of items in a search array and the average time to find a search target provides an index of the efficiency with which the search stimuli are processed [Wolfe, 1998]. Targets that differ from distractors along a single feature dimension (e.g., a red target among green distractors) tend to be found efficiently: the target "pops out" and the search time remains relatively constant regardless of how many distractors are present [Treisman and Gelade, 1980]. In contrast, when the search target is defined by a particular combination of two features (e.g., color and location) and distractors are defined by different combinations of the same features, search tends to be inefficient: search time increases as the number of distractors increases [e.g., Treisman and Gelade, 1980].

However, feature-conjunction search can become more efficient with practice [Carrasco et al., 1998; Heathcote and Mewhort, 1993; Sireteanu and Rettenbach, 1995; Wang et al., 1994]. The neural mechanisms underlying perceptual learning of feature conjunctions are largely unknown. It has been hypothesized that such learning requires the development of more precise neural representations of feature conjunctions, allowing for improved bottom-up processing of the stimuli and a reduction in the amount of top-down attentional processing required [Gilbert et al., 2001; Walsh et al., 1998; Zhaoping, 2009]. But it is an open question whether this hypothesis is correct, and, if so, where exactly such representational changes might occur. Studies of perceptual learning for single visual features suggest that changes as early as primary visual cortex are associated with behavioral improvements [Kourtzi et al.,

2005; Lee et al., 2002; Sigman et al., 2005; Mukai et al, 2007; Yotsumoto et al., 2008]. However, learning of feature conjunctions might take place at a higher stage of processing where neurons integrate a larger set of visual features over a larger area.

To determine what visual areas are involved in learning visual feature conjunctions, we performed a longitudinal functional magnetic resonance imaging (fMRI) experiment in which participants learned to detect a conjunction of color and location during visual search [Kleffner and Ramachandran, 1992; Logan, 1995; Wolfe et al., 1990]. We chose a longitudinal design because the behavioral increases in search efficiency which take place over the course of such learning are known to have a nonlinear relationship to the time course of training. Moreover, some investigations of single-feature perceptual learning have observed nonlinearities in the relationship between brain activity and training time [Yotsumoto et al., 2008]. Measuring brain activity over the entire learning period makes it possible to observe any potential temporally nonlinear effects.

We simultaneously scanned and measured behavioral performance while subjects carried out a visual search task for a conjunction-defined target on eight consecutive days. We also tested the same participants in several variations of the search task immediately post-learning, and in the original task nine months after the initial training. With these post-tests, we could assess the stimulus-specificity and durability of learned changes in visual processing.

Based on what has been observed when learning a single visual feature, where the degree of search-related activity changes throughout visual cortex [Kourtzi et al., 2005; Lee et al., 2002; Sigman et al., 2005; Yotsumoto et al., 2008], we hypothesized that feature conjunction learning would be accompanied by changes in visual cortex activity level. However, because search targets defined by visual feature conjunctions require more integrated visual information to detect, we hypothesized that the changes in activation accompanying feature conjunction learning would be in higher visual areas than those found in single-feature learning experiments, where changes as early as V1 are commonly reported [Kourtzi et al., 2005; Lee et al., 2002; Sigman et al., 2005; Yotsumoto et al., 2008].

To test this, we defined retinotopic regions of interest (ROIs) where processing might change for each participant. We also functionally localized the object-sensitive lateral occipital complex (LOC) in each participant, since we expected that changes to processing of visual-feature conjunctions might occur at a level representing highly integrated information about objects [Grill-Spector et al., 2001]. In addition, human motion-sensitive area MT+ (V5) was localized to investigate whether the processing of an irrelevant feature (motion jitter of target and distractor stimuli) would also change with learning.

Previous experiments using transcranial magnetic stimulation to investigate processing changes in visual learning of feature conjunctions have suggested that the involvement of attentional areas declines as feature conjunctions are

learned [Walsh et al., 1998]. To investigate whether activity levels in attentional areas might also change as feature conjunctions are learned, in each participant we functionally localized the frontal eye fields (FEF) and the nearby supplementary eye fields (SEF), which are known to be involved in attentional processing [Corbetta et al., 1998]. We also anatomically localized the superior colliculi (another attention-related area) in each participant's brain. In line with previous reports, we expected that activity in these attentional ROIs would decrease as learning progressed.

METHODS

Participants

Three of the authors (23-, 24-, and 48-years old, all male) participated in the study. All were experienced in psychophysical and fMRI experiments and had normal vision. None had been trained on the task before.

Stimuli and Task

Each participant performed 14 experimental sessions on 14 separate days. Each experimental session took place in a 3-Tesla MRI scanner during functional scanning, and every experimental session consisted of eight scanning runs (~6.5 min) with 39 trials each.

There were four different conditions, explained in detail below. Each condition contained different stimuli, but aside from the stimulus differences, other experimental parameters were identical across sessions. In all cases, 32 circular stimuli were arranged in four concentric rings around a fixation point. A fixed number of stimuli were used because this simplifies interpretation of the neuroimaging data and maximizes experimental power. In preliminary behavioral experiments with a variable number of stimuli, we established that search speed for a single set size is an effective proxy for search efficiency (i.e., search slope) (see Supporting Information). In other words, the relationship between set size and search time is highly linear. The size of the stimuli and their distance from fixation were scaled according to the human cortical magnification factor [Duncan and Boynton, 2003]. The innermost ring was 0.67 degrees of visual angle (deg.) away from fixation and the entire display spanned 22.5 deg. Accordingly, visibility of the stimuli was constant over all locations when the participant fixated a central fixation mark, therefore allowing covert search without volitional eye movements.

On each of the 32 target-present trials per run, the target appeared in just one of the 32 disk positions. The participants' task was to locate this target stimulus as quickly and accurately as possible, reporting which ring contained the target (Key 1: inner ring to Key 4: outer ring). They were required to fixate and search without eye movements. Target location was counterbalanced within each run, thereby covering each possible stimulus location with equal frequency. On catch trials (seven per run) no target stimulus appeared and the search display contained 32 distractors. The participants reported the perceived absence of the target with another key (Key 5), pressed using their thumb. The stimulus array was presented for 4 sec, and stimuli were spatially jittered every 100 msec within stationary dark-gray circular placeholders to avoid perceptual fading. Participants could respond at any time from the onset of one trial to the onset of the next trial. After each response, the fixation point changed color, indicating whether the response was correct or incorrect.

We did not monitor eye movements during scanning. All participants had extensive experience maintaining fixation in the scanner and past eyetracking in the scanner had established that these participants could maintain central fixation. Nevertheless, some fixational eye-movements such as microsaccades are unavoidable. Since such eye-movements are known to influence brain activity [Tse et al., 2010], a change in the number of fixational eye-movements over the course of learning could confound interpretation of changes in brain activity associated with learning. To address this concern, we performed eyetracking outside the scanner in three different participants as they learned the search task (see Supporting Information).

Primary search task

The target stimulus was a red-green bisected disk presented amongst green-red bisected disk distractors (see Fig. 1A). The color intensities were adjusted to be subjectively isoluminant with one another, and the same colors were used in all subsequent tasks. The participants performed one session daily over a period of eight successive days. After nine months of no additional training, the participants performed one additional session to test the stability of the initial learning.

Flip control task

After completion of the initial learning, target and distractor stimuli were reversed so that what had been a target became a distractor and *vice versa* (see Supporting Information Fig. 1A). A total of three sessions—one per day—were completed in this condition.

Pop-out control task

These control stimuli elicit efficient visual search without learning [Wolfe et al. 1994]. Here, the participants searched for a red-green colored bull's-eye disk among green-red colored bull's-eye disks (Supporting Information Fig. 1B). One session was performed.

Conjunction control task

This condition elicited slow visual search after training on the primary task. The target was a red horizontal bar

on a green disk presented among two types of distractors: a green horizontal bar on a red disk and a vertical red bar on a green disk (Supporting Information Fig. 1C). One session was performed.

Figure 1.

Stimulus Generation and Presentation

Stimuli were programmed in MATLAB version 2007b (The MathWorks, Natick, MA), using the Psychophysics Toolbox [version 3.0.8; Brainard, 1997; Pelli, 1997], and back-projected onto a translucent circular screen (∼30 deg. diameter), located at the back of the scanner bore. The participants viewed the screen via a head coil-mounted mirror (viewing distance: 63 cm).

fMRI Parameters

Data were acquired with a fast event-related design. Event durations were constant (4 sec), and the inter-stimulus interval was jittered between 4, 6, and 8 sec (balanced within each run). MRI scanning was performed with a 3-Tesla Allegra head scanner (Siemens, Erlangen, Germany) and a one-channel head coil. Functional whole-brain images were acquired interleaved with a T2*-weighted gradient echo planar imaging (EPI) sequence (time-to-repeat, TR = 2 sec.; time-to-echo, TE = 30 msec.; flip angle, FA = 90°) consisting of 34 transverse slices (voxel-size = $3 \times 3 \times 3$ mm^3; inter-slice gap = 0.5 mm; field of view, FOV = 192 mm × 192 mm). In addition, we collected three high-resolution structural scans (160 sagittal slices each) with a T1-weighted, magnetization prepared rapid gradient echo (MP-RAGE) sequence (TR = 2.25 sec, TE = 2.6 msec, FA = 9°, voxel size = $1 \times 1 \times 1$ mm^3, no inter-slice gap, FOV = 240 mm × 256 mm). The sequence was optimized for the differentiation of gray and white matter by using parameters from the Alzheimer's Disease Neuroimaging Initiative project (http://adni.loni.ucla.edu/).

MRI Data Analysis

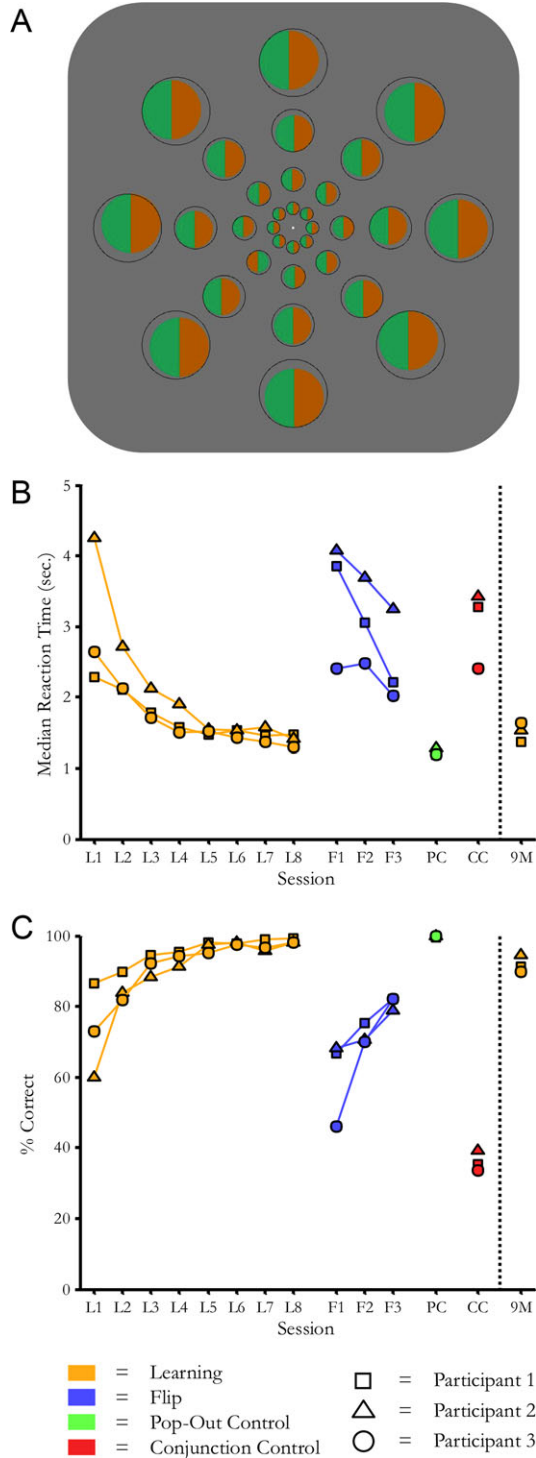MRI data analysis was performed with Freesurfer version 4.1 and the FSFAST toolbox (Martinos Center for

Figure 1.

Behavioral task. **A**: Depiction of example search stimuli. The correct response for this trial would be "Target in ring 2." To counteract perceptual fading, stimuli were spatially jittered every 100 ms within the stationary dark gray bounding rings. Each search display contained 32 stimuli in this arrangement and was presented for 4 sec. (corresponding to two durations of the MR volume acquisition or time-to-repeat, TR). The fixation spot changed color at the end of each trial to indicate a correct or incorrect response. **B**: Reaction times as a function of experimental session, in the original sequence of data collection. In orange: initial training sessions and nine-month retest (8+1 days). In blue: reversed target/distractor mapping (3 days). In green: control condition of search for red/green among green/red "bull's-eye" patterns, which does not require learning for pop-out. In red: untrained search for color/aspect ratio conjunction, containing the same area of red and green as the other 3 stimulus sets. **C**: Accuracy rates. Conventions as above.

Biomedical Imaging, Charlestown, MA). The anatomical scans of each participant's brain were averaged, reconstructed, and inflated [Dale et al., 1999; Fischl et al., 1999]. Functional images of each scanning day were motion-corrected [Cox and Jesmanowicz, 1999], co-registered (via the first functional image of each session) to the reconstructed individual brain, smoothed with a three-dimensional Gaussian kernel (full-width at half-maximum = 5 mm), and intensity normalized [Sled et al., 1998]. fMRI data were analyzed using a general linear model (GLM) approach with an event-related design. The blood-oxygenation-level-dependent (BOLD) response was modeled using a gamma-function (delta = 2.25, tau = 1.25, alpha = 2).

We calculated two GLMs. The primary GLM analysis used the individual trial-by-trial reaction times (RTs) as event durations. Therefore, for each trial the BOLD-response, taking into account the hemodynamic delay and typical response shape, was accumulated from trial onset until button press and ended thereafter with the typical fall time. This time-window covered the period expected to correspond to active search, leading up to the response. Faster response times therefore yielded an earlier and smaller peak of the gamma-function relative to slower RTs. A drawback of this approach is that decreasing event durations (i.e., faster RTs) across sessions cause variations in the shape of the gamma-function in the design matrices for each session. To control for this possible limitation in the RT-model, we conducted a different GLM analysis using constant event durations of 4 sec (i.e., stimulus presentation time) for all scanning days (see Supporting Information). In each of the statistical models, a single predictor for all events (i.e., target and catch trials) was used. However, for our primary MRI-model with RT as event duration we conducted an additional analysis with separate predictors for target and catch trials to check for any signal-differences between the two conditions (see Supporting Information). We included motion-correction parameters as regressors of no interest and a linear scanner drift predictor in all GLMs.

Regions of Interest

All fMRI analyses were focused on activations in independently localized ROIs, identified via additional scans performed on non-experimental days. We used phase-encoded retinotopy [DeYoe et al., 1996; Engel et al., 1994, 1997; Sereno et al., 1995] to define V1, V2, V3, V3ab, and V4v in occipital cortex. Accordingly, we presented a flickering (4 Hz) checkerboard bow-tie shaped wedge that rotated across 16 different screen positions (2 sec each position) for five cycles. In half of the runs, rotation was clockwise, in the other half it was counterclockwise. At least six runs were conducted in total (~2.7 min each). To functionally localize the FEF and SEF, the participants followed with their gaze a small fixation dot that jumped

unpredictably to one of six different locations across the horizontal screen axis [Kimmig et al., 2001]. During 12 sec long blocks with saccades the fixation dot covered each possible location twice (for one second per location). This stimulus period was contrasted with periods of central fixation. Two runs (9.6 min each) were performed that yielded reliable estimates of the eye fields. The LOC was identified by contrasting objects and scrambled objects presented for 14-sec blocks [Malach et al., 1995]. Human motion-sensitive area hMT+ (V5) was defined by contrasting visual motion against static stimuli [Tootell et al., 1995]. Specifically, we presented 12 sec long blocks during which 200 white dots moved coherently in 12 successive translational directions (1 sec each direction, dot speed: 20 pixels/~0.6° visual angle per image flip, random dot lifetime: 5–10 image flips, image flip rate: 30 Hz). This was contrasted with blocks of static dots. One run (9.6 min) was conducted. In all localizer experiments, ROIs were defined on fMRI data thresholded with a false discovery rate (FDR) of $P < 0.001$. In two retinotopic datasets, the threshold was reduced below this level to facilitate region delineation. Finally, the superior colliculi were localized anatomically on each individual's high-resolution MRI scan. No ROI overlapped spatially with any other ROI.

ROI results were first computed separately for left and right hemispheres and then averaged. This compensates for the effect of left and right ROIs having slightly different sizes on the mean activity of the area. Activation in each ROI during each session was computed as BOLD percent signal change relative to implicit baseline (i.e., the average activity across events and blanks in the ROI).

The relationship between changes in behavior and neural activity was assessed by correlating individual search times (median RT) and BOLD percent signal changes in each ROI across all conditions and participants.

Finally, we were interested in whether there might be increasing response-gain in the retinotopic area representing target location relative to retinotopic areas representing distractor locations. Following Tootell et al. [1998], we functionally localized the retinotopic location corresponding to each of the four concentric ring areas that contained the stimuli (see Fig. 1A). To do so, we stimulated each of the four rings with an annulus comprised of stripes flickering in different colors (flicker rate: 30 Hz) for 12-sec blocks, and contrasted these with blank blocks. Participants maintained fixation throughout the localizer. Next, we performed an MRI-analysis (RT-model) on target-trials using four signal-predictors, coding target location within one of the four rings. The BOLD percent signal change in each of the four conditions was then computed for the ring ROIs. The resulting values were averaged separately for trials where the target was in the represented part of the visual field and trials where the target was in a different part. Results presented are average values across the four ring ROIs. We also performed this analysis on sub-ROIs defined by the intersection of primary visual cortex and the ring-ROIs. We were able to define three sub-ROIs

within V1, one for the representation of the innermost ring, one for the outermost, and one for rings 2 and 3 together.

RESULTS

Behavioral Results

Over the eight days of training, participants' speed and accuracy on the search task improved (see Fig. 1b,c). Reaction times decreased and accuracy increased, both of which are indicative of learning. During retest on the original search task nine months later, participants' search times and accuracies were comparable to those recorded in the final training session. By contrast, when target and distractor identities were exchanged immediately after training, search times rebounded to pre-training levels. Search times were also slow for the conjunction-control condition, but not the pop-out control condition.

Because the position of the search target within each run was random without replacement and the participants knew it, we conducted a control analysis to assess the possibility that part of what participants learned was a memory strategy. If participants were able to remember which positions had already contained a search target within a run, then by the end of the run, the number of positions the participant would need to search should be reduced. If that were the case, then RTs near the end of each run should be faster than RTs near the beginning of each run. We compared the RT for the first nine trials in each run to the RT for the last nine trials in that run using a paired-samples t-test. There was no significant difference between RTs early in each run and late in the same run: $t(189) = 1.63$, ns (two runs were excluded due to missing RTs in one run and an outlier in the other).

Preliminary data collected outside the scanner indicated that in our task, reaction time for a set number of stimuli serves as an extremely effective proxy for search slope (i.e., the amount of processing time required per item), which is the standard measure of search efficiency (see Supporting Information) [Wolfe, 1998]. In the preliminary data, 98.6% of the variance in search slope (msec/item) was accounted for by simple RT in the condition with the maximal number of distractors, because the relationship between number of distractors and search time remains extremely linear throughout training (see Supporting Information).

Some previous studies of learning in visual search have found a change in the intercept of the search function (i.e., a reduction in general task-processing time), as well as changes to the slope of the search function [Scialfa et al., 2004]. Such an intercept change could reflect task-learning not specific to the trained stimuli. In the preliminary dataset, the intercept of the search function did decline somewhat along with the slope of the function (see Supporting Information). However, the correlation between RT for the maximum number of search stimuli and the intercept of the search function, though significant, was much lower than the near-perfect correlation between RT and search slope (see Supporting Information). Thus, the RT measure used in the MRI experiment is a better proxy of search slope than search intercept.

Analysis of the eyetracking data collected in the preliminary dataset suggests that there was no change in the rate of fixational eye movements as learning progressed (see Supporting Information).

fMRI Results

As learning of the visual feature conjunctions progressed, the amount of activity in retinotopic visual cortex increased (Fig. 2). The pattern of activity in those areas on the days when participants changed to the flipped-search condition reflected a sudden departure from the downward trend of training. Activity levels throughout visual cortex (V1–V4v, LOC, and hMT+ ROIs) during the pop-out control condition task were similar in magnitude to the levels observed on the last day of feature conjunction search training. The amount of activity in those areas during the conjunction-control task was similar to that observed on the first day of training. Importantly, the signal in retinotopic cortex was still as high upon retesting in the initial search task nine months later as it was on the last day of feature conjunction search training (see Fig. 2).

Across all participants and experimental conditions, behavioral improvements in search performance were significantly correlated with increased brain activity in visual cortex (V1: $r = -0.43$, $P = 0.005$; V2: $r = -0.5$, $P < 0.001$; V3: $r = -0.56$, $P < 0.001$; V3ab: $r = -0.44$, $P = 0.004$; V4v: $r = -0.64$, $P < 0.001$; LOC: $r = -0.46$, $P = 0.003$; hMT+: $r = -0.6$, $P < 0.001$) (see Fig. 3). This pattern of correlations was also significant for V1, V2, V3, V4v, hMT+, and right LOC in the control analysis where event durations were kept constant (see Supporting Information).

It is known that target-present responses are associated with increased visual cortex activity even when no target is present (i.e., for both hits and false alarms) [Ress and Heeger, 2003]. Over the course of learning, participants' accuracy improved, primarily due to an increase in hit-rate. To address the possibility that the increases in visual cortex activity could be attributable to a change in hit rate, rather than a change in search efficiency (as measured via the proxy of RT), we attempted to correlate hit rate with the amount of activity in each ROI. In contrast to the significant correlations between activity and RT, all but two of the correlations with hit-rate were not significant. There was a significant association in V4v ($r = 0.36$, $P = 0.02$), and hMT+ ($r = 0.44$, $P = 0.004$), but in both cases the r-values were lower than the RT-activity correlations. Furthermore, the correlation between frequency of target-present responses (i.e., hits + false alarms) and BOLD activity was not significant in any area.
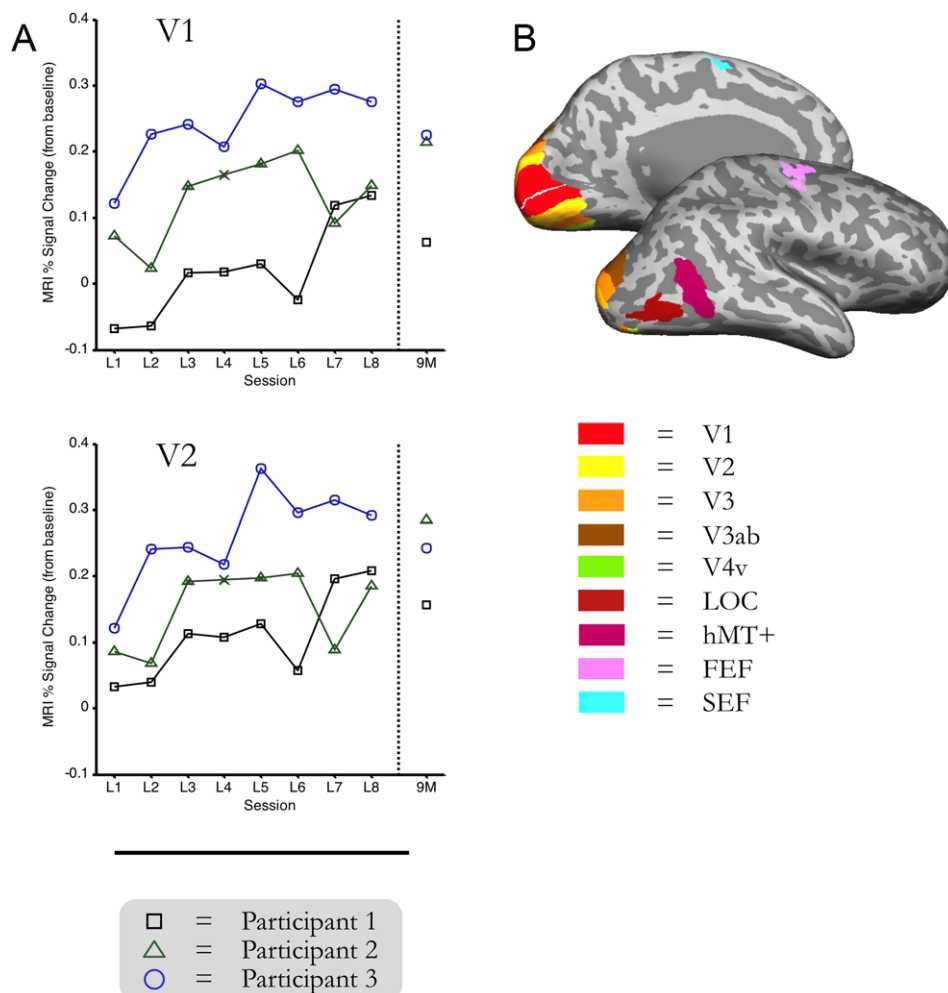
Figure 2.

Functional ROIs and fMRI data over time. Inflated brains depict individually mapped ROIs (V1–V4v), LOC, hMT+, FEF, and SEF in an example participant. Plots show percent-signal change values (RT-model) from learning Day 1 to learning Day 8 and from the follow-up scan after 9 months, relative to fixation-baseline for V1 and V2 in all participants (each line represents one participant). For illustrative purposes, missing data from Participant 2 on day 4 (due to a scanner problem) is interpolated, and demarcated with an "x."

We also correlated BOLD percent signal change that was computed separately for target and catch trials with median reaction times in target-present and target-absent trials, respectively. Correlations for both target and catch trials were highly similar to those reported in the combined analysis (see Supporting Information).

The analysis of the ring-ROIs revealed that the BOLD signal level in all retinotopic areas representing the target ring correlated significantly with RT ($r = -0.58$, $P < 0.001$). The correlation between activity in all retinotopic areas corresponding to the rings containing only distractors was also significant, though smaller ($r = -0.46$, $P = 0.003$). Similar results were obtained for the sub-ROIs covering the retinotopic locations of the rings within V1 (target ring: $r = -0.42$, $P = 0.006$; distractor rings: $r = -0.40$,

$P = 0.01$). Furthermore, we calculated the difference in activity between the target and distractor rings, and correlated this difference score with RT. Increasing target/distractor activity differences were highly correlated with decreasing RTs ($r = -0.70$, $P < 0.001$). This correlation between target/distractor activity differences and RT was also present in the V1 sub-ROIs ($r = -0.36$, $P = 0.02$).

In the primary analysis where RTs were used to model the hemodynamic response, no clear pattern of change in activity over time was apparent in the attentional ROIs. Behavioral improvements were not significantly correlated with activity in any attentional areas as estimated by the RT-model (FEF: $r = 0.037$, $P = 0.8$; SEF: $r = 0.005$, $P = 1$; superior colliculus (SC): $r = 0.089$, $P = 0.6$). By contrast, in the control analysis with constant event durations, decreasing activity
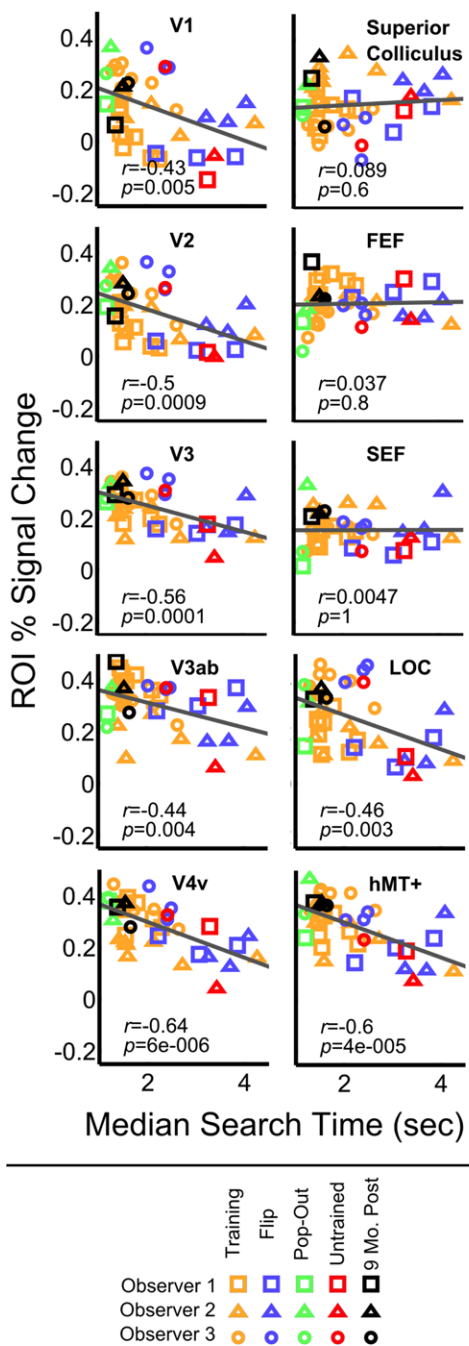
Figure 3.

Correlation between brain activity and behavior. Plots show the relationship between behavioral median reaction time and percent signal change relative to fixation baseline (RT-model), for all ROIs. Each point represents the data from one participant on one scanning day. The regression line is depicted in the color of the corresponding ROI. Rapid search is significantly correlated with high activity in retinotopic visual cortex, LOC, and hMT+. There is no significant correlation between activity in attentional ROIs (FEF, SEF, SC) and search speed.

in FEF, SEF, and SC was significantly correlated with faster search times (see Supporting Information).

DISCUSSION

Over the course of eight days, participants in the present experiment improved markedly in their ability to search for a conjunction of visual features. This is consistent with previous behavioral reports [Carrasco et al., 1998; Heathcote and Mewhort, 1993; Sireteanu and Rettenbach, 1995; Wang et al., 1994]. However, this study is the first to perform the entire training with concurrent MRI-scanning, enabling us to relate session-based changes in behavior to underlying alterations of neural activity.

We find evidence that learning-induced changes in performance are correlated with increasing activity in retinotopic cortex (V1–V4v), LOC, and hMT+. These learning-induced changes in performance and altered brain activity appear to last for at least nine months.

The significant association between search-speed for a conjunction of visual features and increases in the amount of activity at the earliest levels of cortical visual processing, even V1, contradicts our hypothesis that learned changes in processing of visual feature conjunctions would occur at a higher stage of processing than changes reported for simple visual features. Instead, even activity in the earliest levels of the cortical visual processing hierarchy correlates strongly with search speed, a behavioral indicator of stimulus processing efficiency.

We had expected that the amount of activity in attentional ROIs (specifically FEF, SEF, and SC) would decline as learning occurred, since previous studies have found a reduction in the involvement of attentional areas in conjunction search with learning [Walsh et al., 1998]. However, in the primary analysis, which modeled MRI event durations based on RT, the amount of activity in those ROIs remained relatively constant. By contrast, in the control analysis where event durations were fixed at 4 sec, the amount of activity in those areas appeared to decrease (see Supporting Information). This is an instructive difference, for it suggests that these areas are involved in visual search for feature conjunctions both before and after learning, but only during the active phase of search. Because the active phase of search decreases in duration as training progresses, the amount of activity in these areas appears to decrease when the event duration is held constant at 4 sec in the MRI analysis, but not when the duration of the MRI event is tied to the duration of active search (i.e., RT). The activity in those areas during active search might also be influenced by fixational eye-movements such as microsaccades. In either case, the data do not support the hypothesis that there would be a decline in the amount of activity in these areas reflecting a reduction in attentional allocation.

The significant correlation between visual cortex activity and search time does not appear to be an artifact of other

variables which might change alongside search time over the course of learning and alter neural activity, specifically, changes in target detection rate or eye movements.

Target detection has been linked to increased activity in visual cortex even for false alarm responses [Ress and Heeger, 2003]. In our data, participants' hit-rates increased slightly over the course of training. However, these changes in hit rate were not significantly associated with changes to the BOLD response in our data, except in V4v and hMT+. Even in those areas the correlation between activity and hit rate was lower than the correlation between activity and RT. There was no significant correlation between target-present response frequency (i.e., hits + false alarms) and BOLD response in any ROI. Therefore, the increasing activity in visual cortex over the course of training is not attributable to a mere increase in target detection rate.

Changes in participants' pattern of eye movements could introduce another possible confound to the interpretation of the results. Small fixational eye movements such as microsaccades have been associated with increases in activity in visual cortex [Tse et al., 2010]. However, analysis of eyetracking data from different participants performing the same task outside the scanner demonstrates that the number of microsaccades and saccades made during the pre-response window within which the MRI analysis was performed decreases over the course of training (since the rate remains constant and the window shortens). If anything, a decrease in the number of fixational eye movements over the course of training would be expected to produce a concomitant decrease in signal in visual cortex, given previous findings [Tse et al., 2010]. We find the opposite. Therefore, changes in fixational eye-movement patterns appear an unlikely explanation of our results.

Instead, the results of the ring-localizer analyses suggest that the observed increases in visual cortex activity with learning result from improved BOLD signal-to-noise ratio (SNR) for targets relative to distractors. Although the amount of activity in retinotopic areas corresponding to distractors increased as training progressed, the amount of activity in retinotopic areas corresponding to targets increased more. This divergence of the amount of target versus distractor activity would have produced an enhancement of SNR for the target. The difference between activity levels for targets and distractors was strongly predictive of RT, both in the target/distractor ROIs at large, and in the sub-ROIs representing the intersection of target/distractor rings and V1.

Another way in which SNR for search targets could be improved would be if distractor-related activity were reduced. Though such a reduction in distractor activity is not apparent in our data, it could take place at a later stage of learning. Yotsumoto and coworkers [2008] found that activity in visual cortex first increased, then decreased as participants learned a texture discrimination task. It could be that our training regimen, which was shorter than the one in that study, captured only the first phase of learning. If we had continued training for longer, it is possible that the amount of distractor-related activity might have later declined, giving the appearance of a region-wide decrease in activity, but further improving SNR for the target. Alternatively, it might be that the mechanisms which support learning of texture discrimination follow a different course of development than the mechanisms supporting visual feature conjunction learning. Further research will be required to elucidate the reasons for the difference between the present results and those of Yotsumoto and coworkers.

What type of learning do the observed changes in visual cortex reflect? Learning a complex perceptual task like feature-conjunction search might involve a synthesis of multiple different types of learning: perceptual learning of the stimuli, cognitive learning of task strategies, even motoric learning of response patterns.

Behaviorally, it is clear that the majority of the learning that took place in the experiment was stimulus-specific (i.e., perceptual). This is most evident from the pattern of behavior in the flip-control condition. In that condition, after 8 days of training, the target and distractor identities were reversed. If the learning were primarily task-related, or cognitive, then this exchange should not have impaired performance: the task remained the same, and participants were already highly experienced with the stimuli. In fact, since distractors greatly outnumbered targets during training, participants had over 30 times more exposure to the new targets than they had to the original targets at the time of the switch. Yet, search times returned to pre-training levels when we switched the stimuli. This suggests that a change in automatic, bottom-up, processing of the target stimuli was the primary factor driving the improvements in search performance. In other words, perceptual learning of the target appears to have formed the primary basis for the changes.

It is possible that other forms of learning may have contributed to the changes we report, but there is little evidence that they played a major role in the present effects. First, the lack of learning transfer to the flip-control condition precludes a major involvement of task-learning. Similarly, data from the participants in the preliminary, behavioral version of the task suggests the amount of task-learning that occurs during training is relatively modest. Because that version of the task included a variable number of stimuli from trial to trial, it was possible to estimate the intercept of the search function (i.e., the amount of time that would be required to make a response in the task if the number of stimuli were zero). Changes to the intercept of the search function over time would thus primarily reflect task-learning. Although there is a slight decrease in intercept RTs, the effects are far less dramatic than the changes in the search slope (i.e., the amount of time required to process each stimulus), a measure of perceptual learning (see Supporting Information).

Also, while the RT measure collected during scanning appears to be an extremely good proxy for search slope

(i.e., perceptual learning), it is a relatively poor proxy for search intercept (i.e., task-learning). The preliminary data collected with a variable number of stimuli indicates that the RT measure is a near-perfect stand-in for search slope: the two values have a correlation of $r = 0.993$ (see Supporting Information). RT is also correlated with search intercept, but to a much lower degree ($r = 0.640$). Given the magnitudes of the correlations between RT and slope versus intercept obtained in the preliminary data, it appears unlikely that the observed changes in visual cortex can be attributed solely to task-learning.

There is also little evidence that learning of explicit cognitive strategies contributed much if anything to the effects of learning in the present task. For example, target position was counterbalanced within each run and participants were aware of it. In theory, participants could have developed a memory strategy to keep track of which positions had already contained a target over a given run, reducing the number of positions needing to be searched as the run progressed. Yet there was no significant difference between RTs early and late in each run, suggesting that participants did not employ such a strategy.

In sum, while it remains an open question exactly what proportion of learning is attributable to non-perceptual learning processes, it appears reasonable to conclude that perceptual learning of visual feature conjunctions is the primary factor driving the relationship between behavior and brain activity seen in the present experiment.

Exactly what neural mechanism enables this learning remains an open question. The increasing activity measured for search targets in visual cortex could be driven by at least two possible neural mechanisms. One possibility is that neurons in early visual cortex become better tuned to the search stimuli, such that target stimuli evoke a larger response in early visual cortex that propagates up the visual hierarchy in a feedforward manner [Gilbert et al., 2001; Walsh et al., 1998; Zhaoping, 2009]. Another possibility is that the SNR improvements in early visual cortex do not involve changes to representations in visual cortex *per se*, but rather reflect feedback from enhanced allocation of exogenous visual attention to the targets [Bartolucci and Smith, 2011]. Or, perhaps the improved SNR results from some combination of both mechanisms. Future experiments will be required to disentangle these possibilities.

In conclusion, we find that changes to the way the brain processes stimuli during visual conjunction search are strongly correlated with behavioral improvements. Given the ecological prevalence of visual search in daily life, whether searching for food or a friend in a crowd, it seems likely that such changes to neural processing occur frequently, if not constantly, outside the laboratory. As we encounter the "blooming, buzzing confusion" [James, 1890] of the visual world, our brains continually adapt to apprehend meaningful perceptual patterns more quickly and accurately. We show that in the case of visual conjunction search learning, there are lasting changes at even the earliest levels of visual processing, reshaping our capacity to recognize complex patterns of visual features in a matter of days.

REFERENCES

Ball K, Sekuler R (1982): A specific and enduring improvement in visual motion discrimination. Science 218:697–698.

Bartolucci M, Smith AT (2011): Attentional modulation in visual cortex is modified during perceptual learning. Neuropsychologia 49:3898–3907.

Brainard DH (1997): The psychophysics toolbox. Spat Vis 10:433–436.

Carrasco M, Ponte D, Rechea C, Sampedro MJ (1998): "Transient structures": The effects of practice and distractor grouping on within-dimension conjunction searches. Percept Psychophys 60:1243.

Corbetta M, Akbudak E, Conturo TE, Snyder AZ, Ollinger JM, Drury HA, Linenweber MR, Petersen SE, Raichle ME, Van Essen DC, Shulman GL (1998): A common network of functional areas for attention and eye movements. Neuron 21:761–73.

Cox RW, Jesmanowicz A (1999): Real-time 3D image registration for functional MRI. Magn Reson Med 42:1014–1018.

Dale AM, Fischl B, Sereno MI (1999): Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179–194.

DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J (1996): Mapping striate and extrastriate visual areas in human cerebral cortex. Proc Natl Acad Sci USA 93:2382.

Duncan R, Boynton GM (2003): Cortical magnification within human primary visual cortex correlates with acuity thresholds. Neuron 38:659–671.

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994): fMRI of human visual cortex. Nature 369:525.

Engel SA, Glover GH, Wandell BA (1997): Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex 7:181–192.

Fischl B, Sereno MI, Dale AM (1999): Cortical surface-based analysis. II. Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195–207.

Gibson EJ (1963): Perceptual learning. Ann Rev Psychol 14:29–56.

Goldstone RL (1998): Perceptual learning. Ann Rev Psychol 49:585–612.

Gilbert CD, Sigman M, Crist RE (2001): The neural basis of perceptual learning. Neuron 31:681–697.

Grill-Spector K, Kourtzi Z, Kanwisher N (2001): The lateral occipital complex and its role in object recognition. Vis Res 41:1409–1422.

Heathcote A, Mewhort DJ (1993): Representation and selection of relative position. J Exp Psychol 19:488–516.

James W. 1890. The Principles of Psychology. New York: Holt.

Karni A, Sagi D (1991): Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. Proc Natl Acad Sci USA 88:4966–4970.

Kimmig H, Greenlee MW, Gondan M, Schira M, Kassubek J, Mergner T (2001): Relationship between saccadic eye movements and cortical activity as measured by fMRI: Quantitative and qualitative aspects. Exp Brain Res 141:184–194.

Kleffner DA, Ramachandran VS (1992): On the perception of shape from shading. Percept Psychophys 52:18–36.

Kourtzi Z, Betts LR, Sarkheil P, Welchman AE (2005): Distributed neural plasticity for shape learning in the human visual cortex. PLoS Biol 3:e204.

Lee TS, Yang CF, Romero RD, Mumford D (2002): Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. Nat Neurosci 5:589–597.

Logan GD (1995): Linguistic and conceptual control of visual spatial attention. Cogn Psychol 28:103–174.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995): Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci USA 92:8135–8139.

Mukai I, Kim D, Fukunaga M, Japee S, Marrett S, Ungerleider LG (2007): Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. J Neurosci 27:11401–11411.

Pelli DG (1997): The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442.

Poggio T, Fahle M, Edelman S (1992): Fast perceptual learning in visual hyperacuity. Science 256:1018–1021.

Ress D, Heeger DJ (2003): Neuronal correlates of perception in early visual cortex. Nat Neurosci 6:414–420.

Scialfa CT, McPhee L, Ho G (2004): Varied-mapping conjunction search: Evidence for rule-based learning. Spat Vis 17:543–557.

Seitz A, Watanabe T (2005): A unified model for perceptual learning. Trends Cogn Sci 9:329–334.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995): Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893.

Sigman M, Pan H, Yang Y, Stern E, Silbersweig D, Gilbert CD (2005): Top-down reorganization of activity in the visual pathway after learning a shape identification task. Neuron 46:823–835.

Sireteanu R, Rettenbach R (1995): Perceptual learning in visual search: fast, enduring, but non-specific. Vis Res 35:2037–2043.

Sled JG, Zijdenbos AP, Evans AC (1998): A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17:87–97.

Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW (1995): Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J Neurosci 15:3215–3230.

Tootell RB, Hadjikhani N, Hall EK, Marrett S, Vanduffel W, Vaughan JT, Dale AM (1998): The retinotopy of visual spatial attention. Neuron 21:1409–1422.

Treisman AM, Gelade G (1980): A feature-integration theory of attention. Cogn Psychol 12:97–136.

Tse PU, Baumgartner FJ, Greenlee MW (2010): Event-related functional MRI of cortical activity evoked by microsaccades, small visually-guided saccades, and eyeblinks in human visual cortex. Neuroimage 49:805–816.

Walsh V, Ashbridge E, Cowey A (1998): Cortical plasticity in perceptual learning demonstrated by transcranial magnetic stimulation. Neuropsychologia 36:45–49.

Wang Q, Cavanagh P, Green M (1994): Familiarity and pop-out in visual search. Percept Psychophys 56:495–500.

Wolfe JM (1998): Visual search. In: Pashler H, editor. Attention. Hove: Psychology Press. pp 13–74.

Wolfe JM, Yu KP, Stewart MI, Shorter a D, Friedman-Hill SR, Cave KR (1990): Limitations on the parallel guidance of visual search: Color × color and orientation × orientation conjunctions. J Exp Psychol 16:879–892.

Wolfe JM, Friedman-Hill SR, Bilsky AB (1994): Parallel processing of part-whole information in visual search tasks. Percept Psychophys 55:537–550.

Yotsumoto Y, Watanabe T, Sasaki Y (2008): Different dynamics of performance and brain activation in the time course of perceptual learning. Neuron 57:827–833.

Zhaoping L (2009): Perceptual learning of pop-out and the primary visual cortex. Learn Percept 1:135–146.