

Neural Correlates of Context-Dependent Feature Conjunction Learning in Visual Search Tasks

Eric A. Reavis,^{1*} Sebastian M. Frank,^{1*} Mark W. Greenlee,² and Peter U. Tse¹

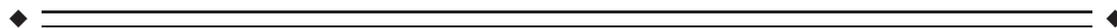
¹*Department of Psychological and Brain Sciences, Dartmouth College, Hanover, New Hampshire, USA*

²*Institute of Experimental Psychology, University of Regensburg, Regensburg, Germany*



Abstract: Many perceptual learning experiments show that repeated exposure to a basic visual feature such as a specific orientation or spatial frequency can modify perception of that feature, and that those perceptual changes are associated with changes in neural tuning early in visual processing. Such perceptual learning effects thus exert a bottom-up influence on subsequent stimulus processing, independent of task-demands or endogenous influences (e.g., volitional attention). However, it is unclear whether such bottom-up changes in perception can occur as more complex stimuli such as conjunctions of visual features are learned. It is not known whether changes in the efficiency with which people learn to process feature conjunctions in a task (e.g., visual search) reflect true bottom-up perceptual learning versus top-down, task-related learning (e.g., learning better control of endogenous attention). Here we show that feature conjunction learning in visual search leads to bottom-up changes in stimulus processing. First, using fMRI, we demonstrate that conjunction learning in visual search has a distinct neural signature: an increase in target-evoked activity relative to distractor-evoked activity (i.e., a relative increase in target salience). Second, we demonstrate that after learning, this neural signature is still evident even when participants passively view learned stimuli while performing an unrelated, attention-demanding task. This suggests that conjunction learning results in altered bottom-up perceptual processing of the learned conjunction stimuli (i.e., a perceptual change independent of the task). We further show that the acquired change in target-evoked activity is contextually dependent on the presence of distractors, suggesting that search array Gestalts are learned. *Hum Brain Mapp* 00:000–000, 2016. © 2016 Wiley Periodicals, Inc.

Key words: neuronal plasticity; learning; visual perception; visual cortex; cognition



INTRODUCTION

A large body of research has characterized perceptual learning, which can be defined as a change in bottom-up

processing of sensory stimuli independent of endogenous cognitive states (e.g., attention) [Gibson, 1963; Goldstone, 1998; Sasaki et al., 2010]. In vision, extensive evidence demonstrates that repeated exposure to a particular visual

*The first two authors contributed equally to this work.

Eric A. Reavis is currently at the Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, and Desert Pacific Mental Illness Research, Education, and Clinical Center, Greater Los Angeles Veterans Affairs Healthcare System, Los Angeles, CA

Contract grant sponsor: German Research Council; Contract grant number: DFG GR988_20-2 (to M.W.G.); Contract grant sponsor: John Templeton Foundation; Contract grant numbers: 14316 & 23437 (to P.U.T.)

*Correspondence to: Eric A. Reavis, Room 27-440, 760 Westwood Plaza, Los Angeles, CA 90024. E-mail: ereavis@ucla.edu
Sebastian M. Frank, 6207 Moore Hall, Hanover, NH 30755. E-mail: sebastian.m.frank.gr@dartmouth.edu

Received for publication 10 July 2015; Revised 17 February 2016; Accepted 24 February 2016.

DOI: 10.1002/hbm.23176

Published online 00 Month 2016 in Wiley Online Library (wileyonlinelibrary.com).

feature (e.g., orientation or spatial frequency) can alter a person’s perception of that feature [Ball and Sekuler, 1987; Crist et al., 1997; Karni and Sagi, 1991; Shiu and Pashler, 1992]. Neuroimaging and neurophysiological results suggest that such perceptual changes are attributable to changes in neural tuning at early levels of visual processing [Furmanski et al., 2004; Schoups et al., 2001; Schwartz et al., 2002; Shibata et al., 2012; Yotsumoto et al., 2008].

Learning can also influence the processing of more complex visual stimuli that ordinarily require substantial cognitive resources. For example, using a visual search task, several studies have shown that training can improve the efficiency with which conjunctions of visual features (e.g., combinations of color and shape) are processed by the visual system [Heathcote and Mewhort, 1993; Sireteanu and Rettenbach, 1995; Carrasco et al., 1998; Sigman and Gilbert, 2000; Frank et al., 2014, 2016; Reavis et al., 2015]. However, it remains unclear whether the changes that accompany conjunction learning reflect enhancement of bottom-up processing, as occurs for single-feature perceptual learning, or whether conjunction learning depends on top-down effects such as learning better control of endogenous attention for particular combinations of features. In this paper, we test the hypothesis that conjunction learning leads to bottom-up changes in stimulus processing.

Previously, we found a neural signature of feature conjunction learning in a visual search task [Frank et al., 2014]. As learning progressed, blood-oxygen level-dependent (BOLD) activity in the retinotopic location of target stimuli increased relative to BOLD activity in the retinotopic locations of simultaneously presented distractors, even though the physical intensity of the two types of stimuli was identical and did not change over the course of the experiment. However, that result was based on a very small experimental sample with a longitudinal design. In the original report, three study authors completed 14 days of functional MRI (fMRI) scans; such a small, nonrepresentative experimental sample raises questions about the generalizability of the results. Furthermore, the experiment tested only one type of visual feature conjunction (color-location conjunctions), raising questions as to whether such an increase in target-evoked activity relative to distractor-evoked activity might be peculiar to only certain types of visual-feature conjunction learning. Also, critically, the design of the experiment did not permit dissociation of task-driven top-down activity changes from bottom-up perceptual changes because all activity was measured while participants performed the visual search task.

In this paper, we report the results of two experiments that address these limitations of our original study. Experiment 1 replicates the finding that feature conjunction learning leads to increased target-evoked activity relative to distractor-evoked activity using a much larger and more representative sample and a more common experimental design. Furthermore, Experiment 1 extends the original results by showing that this neural signature of feature conjunction learning in visual search generalizes to a totally different type of conjunction stimulus: complex motion trajectories, which involve spatiotemporal conjunctions of features. Experiment 2 controls for top-down influences on activity by measuring target- and distractor-evoked activity before and after learning under passive-viewing conditions, while participants were engaged in an unrelated, attentionally demanding fixation task. Moreover, Experiment 2 used two groups of participants who trained on different types of feature conjunctions (color-location conjunctions like those used in the original report, or positional conjunctions: an ‘L’ among ‘Ts’). Experiment 2 reveals, first, that such a neural signature of conjunction learning is evident for a third class of feature conjunctions, and, second, that changes in bottom-up processing are sufficient to elicit this signature pattern of neural activity after learning, ruling out the possibility that the learned increase in target salience reflects only top-down, cognitively-driven changes in activity.

EXPERIMENT 1: MOTION CONJUNCTIONS

Methods

Participants

Twenty-seven participants from the University of Regensburg community were recruited for Experiment 1. Twenty participants, including one author (SMF), completed pre- and post-training fMRI scans and were therefore included in the current analysis (mean age: 26 ± 5 years, 10 females). Participants gave informed written consent prior to participation. The study was approved by the local ethics committee at the University of Regensburg.

Methods summary

Experiment 1 investigated learning of complex motion trajectories (spatiotemporal feature conjunctions) in a visual search task. The experiment included a behavioral task, fMRI, and structural MRI. Participants completed 12 sessions of visual search training, learning to find a dot with a V-shaped motion trajectory among distractor dots with inverted-V-shaped trajectories (see Fig. 1a). We collected fMRI measurements of brain activity during this difficult search task during the first and last training sessions. In those fMRI sessions, participants also completed a simple search task for a dot moving with an upward slope among distractor dots moving with a downward

Abbreviations

BOLD	Blood-oxygen level-dependent
FDR	False-discovery rate
fMRI	Functional MRI
IES	Inverse efficiency score
IPS	Intraparietal sulcus

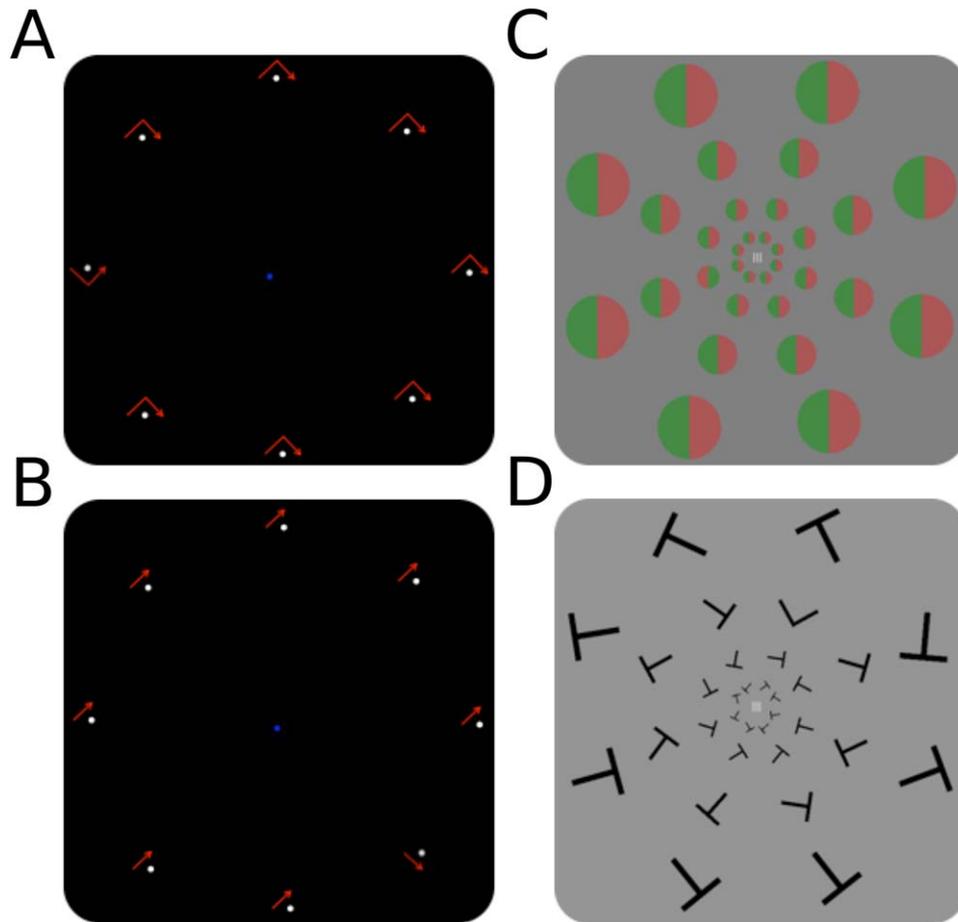


Figure 1.

Visual search tasks in Experiments 1 (a–b) and 2 (c–d). (a) Experiment 1: Motion-conjunction search task: Participants searched the array for a target dot, cycling through a v-shaped motion trajectory. Distractor dots cycled through an inverted v-shaped trajectory. In the example shown, the target is located at 9 o'clock. The red lines and arrows signify the motion trajectory of the individual dots, and were not presented during the experiments. (b) Experiment 1: Control search task: Participants searched the array for a target moving downward and to the right. Distractors moved upward and to the right. In the example shown the target is located at half-past 4 o'clock. This task was designed to elicit efficient search without training, unlike

the demanding task in (a). Again, the red lines and arrows only signify the motion trajectories and were not presented during the experiments. (c) Experiment 2: Color-location conjunction search task. Participants in the R/G training group searched the array for a target disk with red on the left and green on the right among distractor disks with green on the left and red on the right. In the example shown the target is located in ring 2 at 8 o'clock. (d) Experiment 2: Positional conjunction search task. Participants in the T/L training group searched the array for an L-target among T-distractors. In the example shown the target is located in ring 3 at 1 o'clock.

slope (see Fig. 1b). We also collected structural MRI scans and functional localizer scans during those MRI sessions.

We recently published analyses of the Experiment 1 data showing that preexisting individual differences in the cortical thickness of MT+ and posterior parietal cortex predicted participants' rate of learning on the complex motion trajectory learning task [Frank et al., 2016]. The data collection methods of Experiment 1 are described in detail in that paper. Analyses reported there show that

participants' performance on the complex motion trajectory search task improved significantly with training, that there was a significant overall increase in activation of MT+ and visual cortex with training, and that the amount of activity change in MT+ was correlated with behavioral learning rate. In this paper, we report a notably different analysis of the Experiment 1 data, focusing on simultaneous differences in the retinotopically specific activity evoked by search targets and distractors, and how those

retinotopically specific patterns of activity change with learning. These analyses are novel and the results have not been reported previously.

Data Analysis

Behavioral data analysis

In both the learning and control search tasks, behavioral data were analyzed by computing an inverse efficiency score (IES) combining reaction time and accuracy [Townsend and Ashby, 1978]: for each session and each task, each participant's median reaction time was divided by fractional accuracy, then log-transformed. Learning was quantified by determining the slope of the linear function best describing IES scores across sessions (learning rate). Steeper (i.e., more negative) slopes indicated faster learning. Such slopes can be related to participants' initial performance, such that participants who perform poorly in early sessions improve more rapidly than participants who perform well in early sessions.

Imaging data analysis

General. Anatomical and functional MRI data were analyzed with FreeSurfer and the FSLFAST toolbox [Dale et al., 1999; Fischl et al., 1999; Martinos Center for Biomedical Imaging, Charlestown, MA]. Preprocessing steps are described in Frank et al., [2016].

Three predictors were constructed for each of the two visual search tasks (complex and simple motion trajectories): two predictors modelled activity from correct detection (i.e. 'hit') trials, one for trials when the target appeared on a meridian, one for off-meridian target trials. A third regressor of no interest modelled incorrect target present trials (misses) and all target absent trials (correct rejections and false alarms). For each predictor and each visual search task, only the period of active search was modelled. In other words, in each trial, the BOLD response was only modelled until the participant responded. Since the predictor of interest in both fMRI sessions is restricted to the period of active search for 'hit' trials only, changes in the BOLD response cannot be attributed to changes in participant reaction times or accuracy after training. Therefore, changes in BOLD activity between the sessions can be interpreted as reflecting improvements from learning [see Frank et al., 2014, 2016].

However, one could argue that if more correct trials enter the analysis after learning than before, it would reduce the amount of noise in estimates of target gain after training, leading to an apparent increase in target-evoked activity after training. Therefore, in a control analysis, we randomly selected an equal number of correct trials in each participant's post-learning fMRI scan as in his or her pre-learning scan and performed our analysis on this matched set of trials in both sessions. In the control analysis, unselected correct trials from the post-learning scan

were included in the regressor of no interest. All models also contained motion-correction parameters as regressors of no interest and a linear scanner drift predictor.

ROI definitions. Localizer blocks containing flickering checkerboard stimuli in the search locations on the vertical and horizontal meridians were contrasted with off-meridian-stimulation blocks at a false-discovery rate (FDR) threshold of $P < 0.001$. ROIs representing dot locations on and off the visual meridians were defined on each participant's inflated cortical surfaces (see Fig. 2a,b). Left- and right-hemisphere ROIs were defined separately for the on- and off-meridian locations. The on- and off-meridian ROIs did not overlap in either hemisphere. The extent of the ROIs was also spatially limited to exclude areas within the intraparietal sulcus (IPS).

Retinotopic gain for targets among distractors. For each ROI (i.e., on- and off-meridian representations in left and right visual cortex), each task (learning and control), and each fMRI session (pre- and post-training), BOLD percent signal change was computed for five different stimulus conditions: (1) target on left horizontal meridian, (2) target on right horizontal meridian, (3) target on vertical meridian, (4) target off-meridians left, and (5) target off-meridians right. In each case activity was computed relative to implicit baseline (predominantly blank inter-trial intervals). We aimed to contrast the BOLD response for target-present trials when a target was presented in an ROI with target-present trials when only distractors were presented in that location. Therefore we averaged across hemispheres BOLD activity in the meridian ROIs when the target appeared on-meridian with BOLD activity in the off-meridian ROIs when the target appeared off-meridian, yielding a combined 'target signal' measure. Likewise, we averaged activity for distractor stimuli, across hemispheres, in the on- and off-meridian ROIs, yielding a combined 'distractor signal' measure. Next, we subtracted the distractor signal from the target signal. This contrast quantifies the retinotopically specific increase in signal for a target plus one/two distractors (off-/on-meridian ROIs), relative to two/three distractors (off-/on-meridian ROIs), within a scan session. Learning-related changes in this contrast score were assessed with a 2x2 repeated-measures ANOVA with factors 'stimulus type' (trained or untrained) and 'time' (before or after training).

RESULTS

Behavior

All participants improved in the learning task with training. One outlier participant showed a learning index slope more than two standard deviations greater than the group mean and was excluded from all further analyses. Participants' IES improved, on average, from 1.77 (SEM = 0.05) in the first training session to 1.21 (0.05) in the last training

session. The slope of the function describing participants' learning IES across sessions was significantly different from zero, indicating a significant change occurred [$t(18) = -12.3$, $P < 0.001$, Cohen's $d = 2.9$]. Accuracy and median reaction times both improved [session 1: 62% (SEM= 2%) & 3.7s (0.16s); session 2: 88% (2%) & 3.0s (0.13s)].

In the untrained task, performance improved marginally between the two fMRI sessions (session 1 IES: 0.94 (SEM= 0.05); session 2 IES: 0.88 (0.04), $t(18) = -1.94$, $P = 0.07$, $d = 0.45$). This change was driven by a small improvement in reaction time [session 1: 0.92s (0.05s), session 2: 0.86s (0.04s)]. Accuracy was stable and near ceiling [session 1: 98% (0.5%), session 2: 98% (0.2%)], as expected for efficient visual search.

Functional Imaging

Figure 3a shows the difference between target- and distractor-evoked signals for the trained and untrained stimuli, before and after training. A 2×2 repeated measures ANOVA with factors 'stimulus type' (trained or untrained stimuli) and 'time' (before or after training) revealed a significant interaction between stimulus type and time [$F(1,18) = 5.80$, $P = 0.03$, $\eta^2 = 0.24$]: the gain for target vs. distractors increased for the trained stimuli and remained constant for the untrained stimuli. There were no significant main effects of either stimulus type [$F(1,18) = 1.04$, $P = 0.32$, $\eta^2 = 0.06$] or time [$F(1,18) = 1.26$, $P = 0.28$, $\eta^2 = 0.07$]. The control analysis with matched trials for the pre- and post-learning scans yielded similar results: there was a significant interaction between stimulus type and time [$F(1,18) = 7.73$, $P = 0.01$, $\eta^2 = 0.30$] and no significant main effect of either stimulus type [$F(1,18) = 0.26$, $P = 0.62$, $\eta^2 = 0.01$] or time [$F(1,18) = 1.92$, $P = 0.18$, $\eta^2 = 0.10$].

EXPERIMENT 1 DISCUSSION

The results of Experiment 1 replicate and extend our previous findings [Frank et al., 2014] using a similar task but very different stimuli: learning of complex motion trajectory conjunctions is associated with an increase in target-evoked activity relative to distractor-evoked activity in retinotopic visual cortex, like the increases previously found using color-position conjunctions [Frank et al., 2014]. In other words, the salience of a learned target relative to distractors increases with training, in terms of neural signal-to-noise. A control analysis shows that this effect is not attributable to an increase in the number of correct trials after training. Extending our original report, Experiment 1 shows the change in retinotopic signal-to-noise for conjunction-defined visual search targets over time in a much larger and more representative sample of participants and with very different stimuli.

However, Experiment 1 had two important limitations. First, the experimental and control conditions were consistent across participants (i.e., the learning condition con-

tained complex trajectories and the control condition contained simple trajectories for all participants), and there was marginal improvement of participants' performance on the easier control task. Second, Experiment 1 did not control for top-down effects of attention and other cognitive processes on activity measured in the fMRI sessions. As in previous studies, participants actively searched for the target in the fMRI sessions, presumably by covertly shifting their attention across the search array. Therefore, it is possible that increases in target-evoked activity could have been caused by greater top-down deployment of attention to the target conjunction after training; spatial attention is known to increase activity in the retinotopic area of the attended location [Tootell et al., 1998; Kastner et al., 1999; Bahrami, Lavie & Rees, 2007]. Alternatively, training might have led to an automatized, bottom-up increase in retinotopic gain for the target, independent of active search processes.

Experiment 2 addresses these competing hypotheses and uses an experimental design that avoids consistent use of the same stimuli for the learning and control conditions across participants. Experiment 2 measured activity related to two types of visual feature conjunctions that elicit robust learning effects—color-location conjunctions and positional conjunctions—using a crossover design. It controls for the possible influence of attention or other cognitive processes by measuring target- and distractor-evoked activity before and after training while participants performed an orthogonal, attention-demanding visual task at fixation.

EXPERIMENT 2: COLOR-LOCATION AND POSITIONAL CONJUNCTIONS

Methods

Participants

Twenty participants from the Dartmouth College community (mean age: 20 ± 3 years, 15 females) were recruited for Experiment 2. Participants gave informed, written consent prior to participation. The study was approved by the Dartmouth College institutional review board.

Design and procedure

As in Experiment 1, participants completed two sessions of fMRI, one before and one after training on a visual conjunction search task. During each scan, participants viewed two different sets of conjunction stimuli. Although the presentation parameters were similar to those of a typical visual search task, participants were required to complete an orthogonal task that demanded focused visual attention at fixation, preventing them from attending to or searching the conjunction stimuli.

Between the two fMRI sessions, each participant completed seven sessions of training on one of the two sets of

conjunction stimuli (counterbalanced across participants) on separate days. Retinotopically specific target- and distractor-evoked activity was compared before and after learning between the learned and unlearned conjunction stimuli using an approach similar to that of Experiment 1.

Stimuli

Participants viewed two types of conjunction stimuli: color-location conjunctions and positional conjunctions. Specifically, for the color-location conjunction condition, participants saw red and green bisected disks [see Frank et al., 2014; Reavis et al., 2015] (Fig. 1c). The target stimulus was red on the left and green on the right and distractors were green on the left and red on the right. Colors were standard red and green values and were kept constant across participants. The positional conjunction stimuli were 'T' and 'L' shapes, presented at random orientations (Fig. 1d). The black line segments forming the stems and arms of the T and L stimuli were of equal lengths and differed only in the position of their attachment: end-to-end or end-to-center. The L-stimulus was the target and T-stimuli served as distractors.

On each trial during training and scanning 32 stimuli appeared, arranged in four concentric rings. Stimulus size in each ring was scaled according to cortical magnification in order to engage a similar sized area of retinotopic cortex for each stimulus [Duncan and Boynton, 2003]. The position of each stimulus jittered slightly every 100ms during each trial in order to counteract the possibility of perceptual (Troxler) fading. The total diameter of the stimulus display subtended 41.6°. The innermost edges of the innermost stimuli were 0.8° from a central fixation point, and those innermost stimuli each subtended 0.6°. Each concentric ring of stimuli was separated from the next by 0.8°. The outermost stimuli each subtended 3.7°. Stimuli were presented on a gray background.

Behavioral training

Between the two fMRI sessions (see below), each participant completed seven behavioral training sessions on separate days, practicing visual search with one of the two sets of conjunction stimuli (counterbalanced across participants). There were 120 trials per training session, including 20 target-absent trials. Participants self-initiated each trial and covertly searched the stimulus array for the presence or absence of the target. Participants indicated the location of the target (in ring 1, 2, 3, or 4) or target absence (five-alternative forced choice task). Participants were asked to respond as quickly and accurately as possible. Stimuli remained on the screen until participants responded. After each response, participants received accuracy feedback (green or red fixation point for correct or incorrect, respectively). A typical training session lasted 5-15 minutes.

fMRI task

During the pre- and post-training fMRI sessions, participants viewed both types of feature conjunction stimuli (learned and unlearned). Each of the two stimulus sets was shown under three different conditions: (1) one target among 31 distractors (equivalent to a target-present trial during training), (2) 32 target stimuli only, and (3) 32 distractor stimuli only. Accordingly, there were six different conditions in total during each fMRI session.

Stimuli were presented in 12s blocks, and each stimulus block was followed by a 12s blank interval. Each block contained four 2s presentations of the stimuli, separated by 1s inter-stimulus intervals. For blocks where all stimuli were the same (i.e. all-target or all-distractor blocks), all four presentations were identical. For target-among-distractor blocks, the position of the target changed between each of the four presentations, but always remained at the same ring eccentricity within a block. All stimulus locations and eccentricities contained a target with equal frequency; this was counterbalanced across runs.

Throughout the scan, participants completed an attentionally demanding task to ensure good fixation and prevent them from covertly searching the stimulus arrays. Throughout the entire duration of each run, a small vertically oriented grating spanning 0.27° appeared continuously at the center of the screen. The grating alternated every 200ms between a vertical orientation and randomly chosen 5° clockwise or counterclockwise rotations. Participants' task was to count the number of times the grating rotated clockwise or counterclockwise throughout the 12s stimulus block and the first nine seconds of the subsequent baseline block, after which a question mark appeared in the center of the screen for three seconds. Then, the participant pressed one of two buttons to indicate whether the grating had rotated clockwise or counterclockwise more frequently during the preceding 21s period. At the onset of the next block, the grating appeared and began rotating again. Participants received feedback about their performance after the end of each run. This demanding fixation task was designed to prevent participants from searching the conjunction stimuli during the fMRI sessions.

In each scan session, we collected eight functional MRI runs with 12 stimulus blocks in each (two for each of the six conditions). Condition order was randomly shuffled in each run, and the duration of each run was 5 minutes. Stimulus appearance was kept comparable between the scanner and behavioral trainings by equating the retinal size and position of the stimuli.

Functional localizer

As in Experiment 1, we identified the retinotopic locations of the conjunction stimuli in visual cortex with a functional localizer. Blocks of colorful, flickering, annular

checkerboard patterns were presented at each of the four ring eccentricities. Stimulation blocks were 12s and followed by 12s blank intervals. Three blocks of data were collected for each ring location, in a random order. During the localizer scan, participants fixated on a point at the center of the screen, which dimmed periodically. Participants were asked to respond with a button-press every time the point dimmed. The localizer scan lasted 5 minutes.

Stimulus generation

Stimuli were generated using Psychtoolbox [Brainard, 1997; Pelli, 1997] running in MATLAB (Mathworks, Natick, MA).

Scanning parameters

Imaging was performed with a 3-Tesla Philips Intera Achieva scanner (Philips, Best, The Netherlands) with a 32-channel head coil. Functional imaging data were collected with a standard T2*-weighted EPI sequence (TR = 2s, TE = 35ms, image matrix = 80x80 voxels, voxel-size = 3x3x3 mm, 35 slices, no inter-slice gap). A high-resolution anatomical scan of each participant's brain was acquired during the first scanning session with an MPRAGE sequence (TR = 2.25s, TE = 2.6ms, 256x256 voxels, voxel-size = 1x1x1 mm, 220 slices, no inter-slice gap).

Behavioral data analysis

As in Experiment 1, an IES score was calculated for each behavioral training session, for each participant, by dividing median reaction time by accuracy. This IES score was log-transformed and a linear function was fitted to learning indices across training sessions for each participant. The slopes of these functions, across participants, were compared to zero using a one-sample *t*-test.

Imaging data analysis

High-resolution anatomical scans and functional imaging data were processed as in Experiment 1 (again, using the FreeSurfer and FSFAST software packages).

Functional data were analysed using a general linear model approach with a block design. For the pre-training and post-training scanning sessions, twelve predictors were included. Four predictors modelled activation when the target was presented in each of the four rings. Two predictors modelled activity during stimulation with target-only and distractor-only trials. Each of these six predictors were constructed separately for the trained and untrained conjunction stimuli, yielding twelve predictors in total. Since participant performance on the central task was high (see below), we included all trials in the analysis. All models contained motion correction parameters and a linear scanner drift parameter as regressors of no interest.

ROI definitions

Using data from the functional localizer, ROIs were defined for the four ring locations on each participant's inflated hemispheric surfaces (see Fig. 2c,d). Each ring was defined by contrasting activation during stimulation of one ring with the combined activation during stimulation of each of the other three rings. As in Experiment 1 an FDR-corrected threshold of $P < 0.001$ was used, and ROIs were spatially constrained to avoid overlap with each other, and to avoid intrusion into the IPS.

Retinotopic gain for target among distractors

Analogous to Experiment 1, the difference in target- and distractor-evoked activity was computed before and after learning for trained and untrained stimuli, by comparing retinotopically-constrained activity within trials. For each session, BOLD percent signal change was computed in each ROI for trials containing a target among distractors. The BOLD signal was referenced to implicit baseline, consisting primarily of blank inter-stimulus intervals, and was averaged across hemispheres. Next, target-evoked activity (i.e., the activity of one target and seven distractors within a ring) was averaged across ROIs, and simultaneous distractor-evoked activity (i.e., activity from rings containing eight distractors) was averaged separately across ROIs. This resulted in two values for each participant, one capturing target-evoked activity, the other capturing simultaneous distractor-evoked activity.

As in Experiment 1, we subtracted distractor-evoked activity from target-evoked activity separately for trained and untrained stimuli. A $2 \times 2 \times 2$ mixed-design ANOVA was carried out on these difference scores to test for effects of stimulus type (trained or untrained stimuli, repeated measure), time (before or after training, repeated measure), and training group (training on R/G or T/L stimuli, non-repeated measure).

We also created a composite ROI of all four rings and computed BOLD percent signal change for the conditions containing all targets and all distractors. Activity for all-distractor trials was subtracted from activity in all-target trials. These scores were evaluated using a separate $2 \times 2 \times 2$ mixed design ANOVA with the same factors as above.

RESULTS

Behavior

Participants' visual search performance improved in the training task completed outside the scanner. Slopes representing changes in IES over sessions were significantly different from zero in the R/G training group [$t(9) = -8.57$, $P < 0.001$, Cohen's $d = 5.71$] and in the T/L training group [$t(9) = -13.17$, $P < 0.001$, $d = 8.78$]. The behavioral data from these training sessions are pending publication elsewhere, alongside data from similar purely behavioral

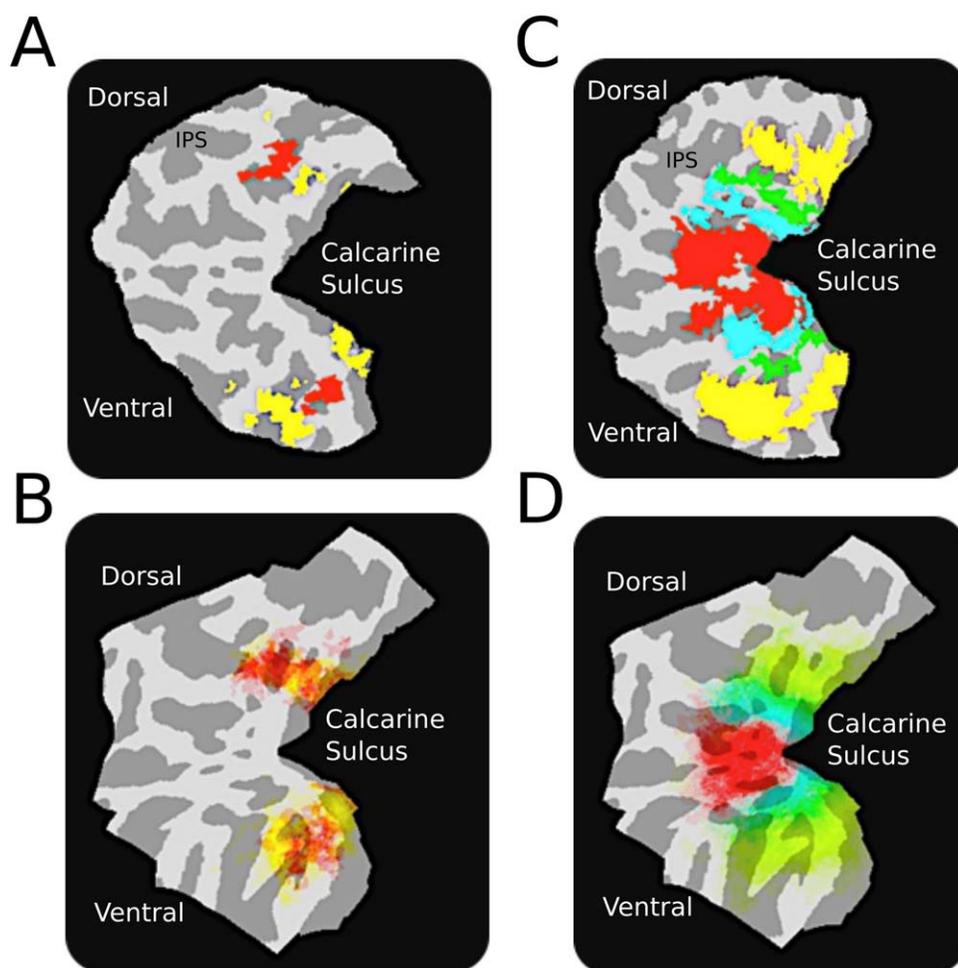


Figure 2.

Regions-of-interest (ROIs) in Experiments 1 (a,b) and 2 (c,d). (a) Experiment 1: Stimulus locations in visual cortex: cortical flat-map of visual cortex from a representative participant (left hemisphere). Gyri are in light grey; sulci are in dark grey. Yellow spots represent the four stimulus locations on the horizontal and vertical meridians, red spots represent two of the four off-meridian stimulus locations (the two in the left hemisphere). (b) Experiment 1: Overlap in stimulus locations across participants. For illustrative purposes, all individual stimulus locations were remapped onto a flattened map of left visual cortex of the FreeSurfer template brain (though all analyses were performed in participants' native space). Because the ROIs were remapped to a template brain for this illustration, sulcal and gyral locations

are approximate relative to individuals' neuroanatomy. Color conventions as in (a). Regions of greater overlap across participants are displayed in more saturated colors. (c) Experiment 2: Stimulus locations in visual cortex: cortical flat-map of visual cortex for a representative participant (left hemisphere). Different colors represent stimulus locations in the four different rings (red = innermost ring, yellow = outermost ring). Each ring contained eight stimuli. Other figure conventions as in (a). (d) Experiment 2: Overlap in stimulus locations across participants. Colors as in (c); other figure conventions as in (b). As in Experiment 1, all analyses were carried out in participants' native space and the location of anatomical landmarks illustrated on the average brain differed slightly from participant to participant.

experiments. Slopes were significantly steeper in the R/G group compared to the T/L group [independent-samples $t(18) = -3.15, P = 0.006, d = 1.48$], primarily driven by better performance in the initial session by the T/L group. In the R/G group, accuracy (hits and correct rejections) increased from 76% (SEM = 5.2%) to 95% (1.9%) and median reaction time for target-present trials decreased

from 2.6s (0.35 s) to 0.85s (0.04 s). In the T/L group, accuracy increased from 81% (2.8%) to 91% (1.9%) and reaction time decreased from 2.0s (0.11 s) to 0.94s (0.04 s).

Performance levels in the orthogonal fixation task during scanning were high but not at ceiling (pre-training: 81% (1.8%), post-training: 84% (2.1%)). As expected, performance did not differ significantly between pre- and

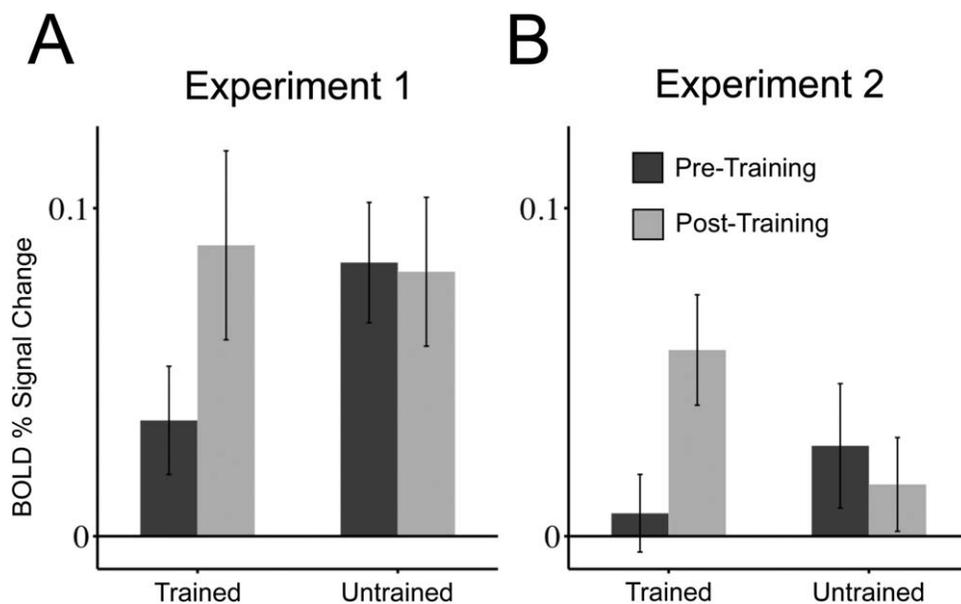


Figure 3.

Target-evoked activity increases relative to distractor-evoked activity for trained stimuli in Experiment 1 (a) and 2 (b). Pre-training activity levels are depicted in dark grey, post-training levels in light grey. (a) Contrast between target- and distractor-evoked activity (in BOLD % signal change) for the trained stimuli (complex motion trajectories) and the untrained control stimuli (unidirectional motion trajectories) in Experiment 1. With training, target-evoked activity increased relative to distractor-evoked activity for the trained stimuli. (b) Contrast between target- and distractor-evoked activity for trained and untrained stimuli in Experiment 2. As in Experiment 1, with training, target-evoked activity increased relative to distractor-evoked activity for

learned stimuli. However, in Experiment 2, activity was measured during passive-viewing conditions while participants were actively engaged in an attentionally-demanding fixation task. Finding an effect under these conditions suggests that the change in neural activity reflects bottom-up changes in perceptual processing of the feature conjunction stimuli and, therefore, that conjunction learning is a form of perceptual learning. Although the magnitudes of the BOLD % signal changes in Experiment 2 appear smaller than in Experiment 1, a direct comparison of the magnitudes is not warranted due to major differences in the design of the two experiments (e.g., event-related vs. block-design and different stimuli).

post-training scans [$t(19) = 1.55$, $P = 0.14$]. There was no significant difference in accuracy on the fixation task between trained- and untrained-stimulus trials either before training [$t(19) = -1.24$, $P = 0.23$] or after training [$t(19) = 1.11$, $P = 0.28$].

Functional Imaging

Figure 3b shows the retinotopically-specific response gain for the target among distractors (target signal - distractor signal; see Fig. 3a). Retinotopic target gain was analyzed with a $2 \times 2 \times 2$ mixed design ANOVA as described above. No main effects of stimulus type (trained vs. untrained stimuli: $F(1,18) = 0.66$, $P = 0.43$, $\eta^2 = 0.04$) or time [pre/post-training: $F(1,18) = 2.98$, $P = 0.10$, $\eta^2 = 0.14$] were observed. However, the ANOVA yielded a significant interaction between stimulus type and time [$F(1,18) = 4.54$, $P = 0.047$, $\eta^2 = 0.20$]: retinotopic target response gain increased for the trained but not untrained stimuli (i.e., the T/L stimuli for the R/G trainees and vice

versa). The ANOVA did not yield significant interactions between group (i.e., R/G trainees vs. T/L trainees) and stimulus type [$F(1,18) = 0.63$, $P = 0.44$, $\eta^2 = 0.03$], group and time [$F(1,18) = 0.41$, $P = 0.53$, $\eta^2 = 0.02$], or group, stimulus type, and time [$F(1,18) = 2.03$, $P = 0.17$, $\eta^2 = 0.10$], suggesting that the significant observed effects generalized across training groups.

A second $2 \times 2 \times 2$ mixed design ANOVA with the same factors as above was carried out on the MRI-signal difference between conditions with all targets and all distractors (i.e., stimulus displays with no singleton). The only significant effect was a main effect of time [$F(1,18) = 4.69$, $P = 0.04$, $\eta^2 = 0.21$]. There was no main effect of stimulus type [$F(1,18) = 0.08$, $P = 0.78$, $\eta^2 = 0.004$] and no interaction between stimulus type and time [$F(1,18) = 0.001$, $P = 0.97$, $\eta^2 < 0.01$]. Again, no interactions were observed between training group and time [$F(1,18) = 2.79$, $P = 0.11$, $\eta^2 = 0.13$], training group and stimulus type [$F(1,18) = 2.82$, $P = 0.11$, $\eta^2 = 0.14$], or training group, time, and stimulus type [$F(1,18) = 0.35$, $P = 0.56$, $\eta^2 = 0.02$].

EXPERIMENT 2 DISCUSSION

Experiment 2 replicates the finding that learning feature conjunctions in a visual search task leads to a retinotopically-specific increase in target-evoked activity relative to distractor-evoked activity in visual cortex (an increase in the salience of the search target relative to distractors). It extends the finding to include a new class of conjunction stimuli: positional conjunctions (in the T/L detection task). More importantly, Experiment 2 shows that training-induced increases in target-evoked activity reflect bottom-up changes in perceptual processing of the conjunction stimuli, not top-down, task-driven activity. Learned changes in processing of the conjunction stimuli persist even when endogenous attention is directed away from the conjunction stimuli and participants are engaged in an orthogonal task at performance levels that are consistent across subjects and suggest strong engagement of participants' attention in the distractor task. The persistence of learning-related differences in target- and distractor-evoked activity under these conditions suggests that these effects are driven by true changes in the bottom-up processing of the stimuli, not by task demands of visual search.

Although the magnitude of the changes in the BOLD signal appear smaller in Experiment 2 than Experiment 1 (see Fig. 3), a direct comparison of magnitudes across experiments is not warranted because of major differences in the design of the two experiments (e.g., event-related vs. block-design, different stimuli). Thus, it remains an intriguing question for future research whether any such differences in magnitude might be due to a factor of theoretical interest (e.g., greater attention allocated to the stimuli in Experiment 1) rather than a byproduct of the experimental design factors that differ in the present case.

Experiment 2 also included all-target and all-distractor conditions (i.e., stimulus displays containing no singleton target or distractor). The only significant effect observed in these data was a generalized increase in target-evoked activity (relative to distractor-evoked activity) from the first scan to the second, across both learned and unlearned stimulus types. This nonspecific increase in target activity is not in itself of great theoretical consequence. However, the failure to find a significant interaction between stimulus type and time for all-target versus all-distractor conditions could be of theoretical importance. The fact that there was no sign of this interaction under conditions where 32 conjunction stimuli learned as targets were presented simultaneously, despite the measurable increases in activity driven by just one target in the retinotopic analysis, suggests that learned search targets only elicit increased activity relative to distractors when they are perceived in the context of such distractors. In other words, learning-induced increases in target-evoked activity appear to be *context-dependent*.

This result suggests that instead of thinking of this form of learning as learning target salience *per se*, it might be

more accurate to think of it as learning of target salience in a specific context of other visual stimuli. Put another way, one might conceptualize the learning as learning of the *Gestalt* of the entire search array, rather than learning of targets or distractors in isolation.

GENERAL CONCLUSION

In two experiments, we show that learning various types of complex feature conjunctions in a visual search task is associated with a retinotopically localized change in the neural response to the stimuli. Specifically, we find that the neural signal-to-noise ratio of targets relative to distractors improves with training, such that learned targets elicit more activity than distractors after training, even when the physical intensity of the target and distractor stimuli remains identical and activity evoked by the two types of stimuli is measured simultaneously. This difference in activity evoked by learned targets versus distractors is measurable even when the stimuli are presented under conditions that effectively prevent participants from attending to the stimuli or performing visual search. However, the activity difference is only evident when a target is presented among distractors in a configuration like those encountered during training; arrays of all targets do not elicit a significantly different amount of activity after training than arrays of all distractors.

These results lend clear support to the hypothesis that conjunction learning in visual search results in altered bottom-up processing of conjunction stimuli. In other words, practicing a visual search task with complex conjunction-defined stimuli leads to bottom-up changes in the neural processing of the conjunctions encountered during training that persist even in the absence of active volitional search for those learned conjunctions. Indeed, such bottom-up detection of learned feature conjunctions even occurs when volitional attention is directed toward a demanding, unrelated task at a different location in the visual field. Thus, our results suggest that learning such conjunction stimuli in a visual search task leads to a measurable change in the way those stimuli drive the perceptual system, one that transcends performance of the search task and persists even when participants are occupied with a distractor task. This may indicate that this type of conjunction learning is akin to certain types of single-feature perceptual learning that lead to changes in neural tuning in the visual system [Furmanski et al., 2004; Schoups et al., 2001; Schwartz et al., 2002; Shibata et al., 2012; Yotsumoto et al., 2008].

At the same time, the lack of a significant difference between the all-target and all-distractor conditions post-training in Experiment 2 suggests that this learning is context-dependent. Rather than targets becoming more salient *per se*, it appears that learning makes them more salient *in the context of the distractors*. Put another way, the results are consistent with perceptual learning of the

Gestalt formed by the target and distractors together. Context-dependent learning is well-described in other learning domains (e.g., Pavlovian conditioning) [Bouton, 1993]. Recently, there have been demonstrations that related types of context-dependent learning can occur in tasks involving complex visual stimuli [Crossley, Ashby, & Maddox, 2014]. Moreover, there are indications of contextual cues influencing learning in other perceptual domains [Christensen et al., 1998; Flanagan, et al., 2008]. Yet to our knowledge, ours is the first study to investigate the context-dependency of this type of visual feature conjunction learning using neuroimaging techniques. It will be an important topic for further research to investigate similarities and differences in the nature and neural mechanisms of visual feature conjunction learning versus other forms of learning (e.g., perceptual learning, associative learning).

The results of Experiment 2 also bear on a long-running assumption that attention is critical for the perception of visual feature conjunctions. Influential early models developed to explain the difficulty of visual search for feature conjunctions [Treisman and Gelade, 1980] suggested that attention was required to bind visual features together in order to perceive conjunction-defined objects. While this may or may not be true for novel feature conjunctions, our results suggest an important caveat: particular feature conjunctions can be *learned* as a Gestalt unit, after which it appears little or no attention is required for their perception. This may be an important mechanism for perceptual optimization in ecological vision. Many meaningful stimuli in the visual environment contain information encoded in complex combinations of visual features; if each of those conjunctions required attention to perceive, the visual world would be an overwhelming and bewildering place requiring massive cognitive resources to navigate. The challenge of ecological vision would be made much more tractable by learned ‘chunking’ of frequently-encountered, meaningful conjunctions of visual features into Gestalts that can be perceived efficiently with minimal attention. Our results suggest that such a learning process is indeed possible.

In summary, our results suggest that measurable bottom-up changes in the perceptual processing of visual feature conjunctions can occur with training and persist even when trained observers are occupied with a distractor task. This suggests that such feature conjunction learning is a form of perceptual learning, not a mere byproduct of changes in cognitive or attentional processing during performance of a task such as visual search. Moreover, this form of learning appears to be influenced by the visual context of the stimuli, suggesting that rather than targets and distractors being learned as independent stimuli, it might be that search arrays are learned as Gestalt units. This type of perceptual learning could provide an important mechanism for perceptual optimization in ecological vision.

REFERENCES

- Bahrami B, Lavie N, Rees G (2007): Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Curr Biol* 17:509–513.
- Ball K, Sekuler R (1987): Direction-specific improvement in motion discrimination. *Vis Res* 27:953–965.
- Bouton ME (1993): Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psych Bull* 114: 80–99.
- 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* 60:1243–1258.
- Christensen TA, Waldrop BR, Hildebrand JG (1998): Multitasking in the olfactory system: Context-dependent responses to odors reveal dual GABA-regulated coding mechanisms in single olfactory projection neurons. *J Neurosci* 18:5999–6008.
- Crist R, Kapadia M, Westheimer G, Gilbert CD (1997): Perceptual learning of spatial localization: Specificity for orientation, position, and context. *J Neurophysiol* 78:2889–2894.
- Crossley MJ, Ashby FG, Maddox WT (2014): Context-dependent savings in procedural category learning. *Brain Cognit* 92:1–10.
- Dale AM, Fischl B, Sereno M (1999): Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage* 194:179–194.
- Duncan R, Boynton G (2003): Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron* 38:659–671.
- Fischl B, Sereno MI, Dale AM (1999): Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage* 195–207. 207.
- Flanagan JR, Bittner JP, Johansson RS (2008): Experience can change distinct size-weight priors engaged in lifting objects and judging their weights. *Curr Biol* 18:1742–1747.
- Frank SM, Reavis EA, Tse PU, Greenlee MW (2014): Neural mechanisms of feature conjunction learning: Enduring changes in occipital cortex after a week of training. *Hum Brain Mapp* 35: 1201–1211.
- Frank SM, Reavis EA, Greenlee MW, Tse PU (2016): Pretraining cortical thickness predicts subsequent perceptual learning rate in a visual search task. *Cereb Cortex* 26:1211–1220.
- Furmanski CS, Schluppeck D, Engel SA (2004): Learning strengthens the response of primary visual cortex to simple patterns. *Curr Biol* 14:573–578.
- Gibson E (1963): Perceptual learning. *Annu Rev Psychol* 14:29–56.
- Goldstone RL (1998): Perceptual learning. *Annu Rev Psychol* 49: 585–612.
- Heathcote A, Mewhort DJK (1993): Representation and selection of relative position. *J Exp Psychol Hum Percept Perform* 19: 488–516.
- 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.
- Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999): Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22:751–761.
- Pelli D (1997): The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis* 10:437–442.

- Reavis EA, Frank SM, Tse PU (2015): Caudate nucleus reactivity predicts perceptual learning rate for visual feature conjunctions. *NeuroImage* 110:171–181.
- Sasaki Y, Náñez JE, Watanabe T (2010): Advances in visual perceptual learning and plasticity. *Nat Rev Neurosci* 11:53–60.
- Schoups A, Vogels R, Qian N, Orban G (2001): Practising orientation identification improves orientation coding in V1 neurons. *Nature* 412:549–553.
- Schwartz S, Maquet P, Frith C (2002): Neural correlates of perceptual learning: A functional MRI study of visual texture discrimination. *Proc Natl Acad Sci USA* 99:17137–17142.
- Shibata K, Chang L-H, Kim D, Náñez JE, Sr, Kamitani Y, Watanabe T, Sasaki Y (2012): Decoding reveals plasticity in V3A as a result of motion perceptual learning. *PLoS ONE* 7:e44003.
- Shiu LP, Pashler H (1992): Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Percept Psychophys* 52:582–588.
- Sigman M, Gilbert CD (2000): Learning to find a shape. *Nat Neurosci* 3:264–269.
- Sireteanu R, Rettenbach R (1995): Perceptual learning in visual search: Fast, enduring, but non-specific. *Vision Res* 35:2037–2043.
- Tootell RBH, Hadjikhani N, Hall EK, Marrett S, Vanduffel W, Vaughan JT, Dale AM (1998): The retinotopy of visual spatial attention. *Neuron* 21:1409–1422.
- Townsend J, Ashby F (1978): Methods of modeling capacity in simple processing systems. In: Castellan N, Restle F, editors. *Cognitive Theory*, Vol.3. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 199–239.
- Treisman A, Gelade G (1980): A feature-integration theory of attention. *Cogn Psychol* 12:97–136.
- 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.