

The nature of experience determines object
representations in the visual system

Yetta K. Wong¹, Jonathan R. Folstein² & Isabel Gauthier²

University of Hong Kong, Hong Kong¹

Vanderbilt University, Nashville, Tennessee, USA²

Keywords: visual representation, object recognition, fMRI, perceptual learning,
perceptual expertise

Word count: 8120

Author Note

Yetta K. Wong, Psychology Department, University of Hong Kong; Jonathan R. Folstein, Psychology Department, Vanderbilt University; Isabel Gauthier, Psychology Department, Vanderbilt University.

Correspondence concerning this article should be addressed to Yetta K. Wong, Psychology Department, University of Hong Kong, 604, Knowles Building, Pokfulam Road, Hong Kong. Email: yetta.wong@gmail.com

Abstract

Visual perceptual learning (PL) and perceptual expertise (PE) traditionally lead to different training effects and recruit different brain areas, but reasons for these differences are largely unknown. Here, we tested how the learning history influences visual object representations. Two groups were trained with tasks typically used in PL or PE studies, with the same novel objects, training duration and parafoveal stimulus presentation. We observed qualitatively different changes in the cortical representations of these objects following PL and PE training, replicating typical training effects in each field. These effects were also modulated by testing tasks, suggesting that experience interacts with attentional set and that the choice of testing tasks critically determines the pattern of training effects one can observe after a short-term visual training. Experience appears sufficient to account for prior differences in the neural locus of learning between PL and PE. The nature of the experience with an object's category can determine its representation in the visual system.

Visual learning provides a special window on how the visual system works. Theoretical accounts of visual learning (Op de Beeck & Baker, 2010; Gilbert et al., 2001; Sasaki et al., 2010; Roelfsema et al., 2010; Bukach et al., 2007) are often based on small clusters of empirical research, each located in small regions of the multi-dimensional space of factors potentially affecting learning. It is therefore difficult to extract general principles of learning. Here we seek to explore the space between two such clusters of visual training studies that stem from different traditions of research and typically produce different training effects. The first area of study is perceptual learning (PL), which investigates how practice results in improvements in judgments based on simple visual attributes such as line orientation, Gabor filters or moving dot patterns (Fiorentini & Berardi, 1980; Karni & Sagi, 1991). Behavioral improvements are often highly specific to the trained stimuli (e.g. orientation, spatial frequency or shape), task, visual field, or even trained eye (Fiorentini & Berardi, 1980; Karni & Sagi, 1991; Poggio, Fahle, & Edelman, 1992; Sigman, 2000; Fahle, Edelman, & Poggio, 1995; Ball & Sekuler, 1987; Fiorentini & Berardi, 1981; Fahle, 1997). In the brain, PL studies consistently produce training effects in early retinotopic cortex (Furmanski et al., 2004; Pourtois et al., 2008; Maertens & Pollmann, 2005; Schoups et al., 2001; Mukai et al., 2007; Schwartz et al., 2002; Yotsumoto et al., 2008; Lewis et al., 2009), where neural selectivity to simple attributes and specific retinotopic locations is consistent with highly specific behavioral improvements (Gilbert et al., 2001; Fahle & Poggio, 2002; Fahle, 2009).

A second tradition has focused on a family of phenomena called perceptual expertise (PE), inspired by studies characterizing expertise acquired outside of the laboratory with natural object categories (Gauthier et al., 2000; Xu, 2005; Tanaka et al.,

2001; 2005; James et al., 2005; Wong & Gauthier, 2010; Busey & Vanderkolk, 2002; Harley et al., 2009) and extending this work to laboratory training with familiar or novel objects. In PE studies, observers typically learn to individuate or discriminate¹ visually-similar objects within a category, such as birds (Tanaka et al., 2005), cars (Jiang et al., 2007), letters (McCandliss et al., 2003), and in some cases computer-generated novel objects (Gauthier, Williams, Tarr & Tanaka, 1998; Moore et al., 2006; Op de Beeck et al., 2006; Yue et al., 2006; Wong, Palmeri & Gauthier, 2009a; Wong, Palmeri, Rogers, Gore & Gauthier, 2009b). Contrasting with the specificity of PL effects, learning in PE studies generalizes to new objects in the trained domain (Gauthier et al., 1998; Wong et al., 2009a). PE also leads to category-specific recruitment of areas in lateral occipital region and the fusiform gyrus but generally not in the early visual cortex (Gauthier et al., 1998; 2000; Moore et al., 2006; Op de Beeck et al., 2006; Yue et al., 2006; Wong et al., 2009b; Cohen et al., 2000; van der Linden, Murre & van Turennout, 2008; van der Linden, van Turennout & Indefrey, 2010; but see Harel et al., 2010²).

Although the neural changes after PL and PE are very different, little discussion or empirical work has been devoted to understanding these differences. Comparing PL and PE studies is difficult as methods in the two fields differ on multiple dimensions. For example, PL uses simple visual attributes (Gilbert et al., 2001) while PE typically uses complex objects (Bukach et al., 2006). The training tasks in PL often involve binary

¹ Note that there are reasons to believe that individuation and discrimination trainings have different effects but for the present purposes, they both result in learning that differ from PL training.

² In that study, the extent of expertise effects may have been overestimated due to low level differences in the categories that were compared, suggested by the fact that even car novices showed significantly more activity to cars than control stimuli in early visual areas.

judgment for orientation or size, while those in PE often involve object naming (Gauthier & Tarr, 1997; Wong et al., 2009a). PE usually presents single objects at the fovea while PL often presents multiple objects in peripheral visual regions simultaneously (e.g. Sigman et al., 2000; Karni & Sagi, 1991). The training duration and testing tasks for PL and PE are different, such that not only the learning materials and the amount of experience differ, but also how learning is measured and quantified. Therefore, it is highly difficult to pinpoint specific factor(s) that account for the differences in training effects between PL and PE. Here, we sought to bridge these literatures empirically by testing a specific hypothesis: the divergent neural changes associated with PL vs. PE are driven by the nature of the experience determined by the training tasks, rather than other differences involving the complexity or number of training stimuli, or the part of the visual field used during training. By keeping everything but training experience constant, we can test whether such a manipulation is sufficient to produce PL- vs. PE-like neural patterns of learning.

The role of experience in determining visual representations remains unclear. Some authors suggest that cortical visual representations could be determined by innate factors (Mahon et al., 2009) or largely constrained by object geometry (Tanaka, 1996; Kourtzi & Dicarlo, 2006; Kayaert, Biederman & Vogels, 2005), and others have explicitly suggested that learning may only moderately alter pre-existing object representations (Op De Beeck et al., 2007; Op de Beeck, 2010). However, others suggest that learning history critically determines the visual representation of objects. For example, specific demands of the training task can lead to a shift of object representation from higher to lower visual regions (Sigman et al., 2005). Also, task-specific and item-

specific cortical networks are tuned up by training experience (Gilbert, 2007; Fahle, 2009). In one recent line of research, different training experiences with the same objects yielded qualitatively different perceptual strategies (Wong et al., 2009a) and different patterns of learning in visual cortex (Wong et al., 2009b). In particular, learning to individuate objects from a novel category (a PE task) resulted in a local increase in the right fusiform gyrus, while learning to categorize the same objects resulted in more distributed changes with increased activity in the medial portion of the ventral occipito-temporal cortex relative to more lateral areas (Wong et al., 2009; see also Song et al., 2010).

Here, we examined whether the different patterns of visual learning for PL and PE could be accounted for by training experience. We compared PL and PE with training protocols typical of each field but with matched training object sets, the same parafoveal stimulus presentation and the same training duration. The PL training task followed a visual search paradigm based on judgments of stimulus orientation used by Sigman et al., (2005), while the PE training task modeled an individuation training used in a number of PE studies (Gauthier et al., 1999; Wong et al., 2009a; Rossion, Kung & Tarr, 2004). We used two families of computer-generated “Ziggerin” objects used in prior PE work (Wong et al., 2009a; 2009b) but in two-tone silhouette versions that would allow discrimination in the visual periphery (Fig. 1).

We first examined whether modified PL and PE training protocols would replicate the typical patterns of changes in the brain found in prior PL and PE studies. If differences in training task suffice to account for large differences between PL and PE in prior work, we should observe typical PL training effects after our PL training, with

increased activity in early retinotopic cortex when subjects search for objects in the trained compared to an untrained orientation (Sigman et al., 2005). We should also observe typical PE training effects, with higher visual areas showing selectivity for the trained compared to a novel object category during shape discrimination (Gauthier et al., 1998; 1999).

Second, by comparing the patterns of learning following our modified PL and PE protocols to those obtained in prior work, we may offer some inferences as to whether the departures from classic methods (e.g., our using parafoveal presentation rather than foveal presentation for PE, or our inclusion of task-irrelevant shape differences in PL) influence patterns of visual learning.

Third, we tested how training experience determines object representations by comparing the training effects for PL and PE. If object representation is largely constrained by innate factors, object geometry or pre-existing object representations (Mahon et al., 2009; Tanaka, 1996; Kourtzi & Dicarlo, 2006; Kayaert, Biederman & Vogels, 2005; Op De Beeck et al., 2007; Op de Beeck, 2010), the training effects observed for PL and PE should be highly similar, given that the training objects were matched in the two types of training. However, if training experience matters, we should observe different training effects in PL and PE in the same task and with the same contrasts.

Finally, we examined how changes in the visual system following PL and PE are modulated by testing tasks. In the literature, the neural substrates of PL are often thought to be task-specific (Fahle, 2009; Gilbert et al., 2001; 2007; Li et al., 2004). In contrast, the results in PE are mixed. For example, there is some evidence that with increasing

expertise, category-specific activity becomes less task dependent (e.g. Gauthier et al., 2000), while another PE study found category-specific effects that depended both on the training task and on the testing context (Wong et al., 2009b; see also Harel et al., 2010). Here we included tasks each designed to tap into the learning effects of PL or PE, therefore our design allowed us to examine the influence of testing tasks on the neural patterns of changes associated with each training protocol. We compared the training effects across two different testing tasks in the same group of subjects using the same contrast. If the training effects of PL or PE are task-specific, the results should be different across the two tasks.

Methods

Subjects

Subjects were 24 undergraduate students, graduate students and staff members at Vanderbilt University. All 24 subjects completed the behavioral training and two fMRI sessions, one before and one after the training. Twelve subjects were randomly assigned to the PL group (6 females, 6 males; mean age = 25.1 years; SD = 4.87), and 12 were assigned to the PE group (7 females, 5 males; mean age = 25.1 years; SD = 4.68). All subjects reported normal or corrected-to-normal vision and gave informed consent according to the guidelines of the institutional review board of Vanderbilt University and received money for their participation.

Stimuli

Two categories of novel objects called ‘Ziggerins’ (Wong et al., 2009a) were used and transformed in silhouette format using Adobe Photoshop CS2 software (Fig. 1). Each category of 24 exemplars was defined by a unique part structure and configuration.

Training Regimens

Subjects were trained with 18 exemplars of one of two categories in one of two possible orientations (0° or in 180°), and the remaining six exemplars were reserved for the pretests and posttests (‘novel exemplars’). The trained object set, the trained orientation and which six objects were reserved were all counterbalanced across subjects within groups. There were eight one-hour training sessions and all subjects finished the training sessions within a four-week period. For all training tasks, the stimuli were presented in eight positions, 3.5° from the central fixation along a circle, and each object spanned a visual angle of approximately $1.9^\circ \times 1.9^\circ$ (Fig. 2). Throughout the eight-hour training, no object was presented at fovea and accuracy was stressed over response time. Before the training task was introduced, each subject was allowed to study the 24 training objects presented on a piece of paper with no time limit.

PL training was modeled after the visual search training in Sigman et al (2005), using silhouettes of Ziggerins instead of ‘T’ shapes (see also Lewis et al., 2009). On each trial, one of the 18 training objects was randomly selected to create an eight-object array, in which the eight objects were identical in shape but either plane-rotated 0° , 90° , 180° or 270° from the subject’s assigned training orientation (Figure 2). Subjects judged whether any object in the array was in the target orientation by key press, and targets appeared with 50% probability. On each trial, a central fixation dot was presented for 1000ms,

followed by an eight-object array for 150ms, and then the central fixation reappeared until response (Fig. 2a). Subjects were informed of their mean accuracy every 60 trials.

PE training was modified from the protocol in Greeble training studies (Gauthier & Tarr, 1997; Gauthier, et al., 1998). Eighteen two-syllable nonsense words (e.g., pimo, jepu) were randomly assigned to the 18 trained objects for each subject, and subjects learned to name objects that were presented for 150ms in one of eight positions (Fig. 2). The 18 objects were gradually introduced in four learning phases (4, 4, 4 and 6 objects respectively). Each learning phase included passive viewing (an object with its name) and verification practice (judging whether a name matched with the object) for the newly introduced objects, followed by a naming training among all the introduced objects. Corrective feedback was provided for each naming or verification trial, and subjects proceeded to the next learning phase when they achieved 90% naming accuracy among all the learned objects. The training procedures and behavioral training effects are detailed in Wong et al. (2011).

Pre- and Post-training fMRI scans

The pre and post-training fMRI scans were identical. Each fMRI scan included six experimental runs (three *visual search* and three *shape matching* runs) and 2 localizer runs. Stimuli were presented on an LCD panel and back-projected on a screen. Subjects viewed the stimuli through a mirror mounted on top of an RF coil above their head.

Experimental runs used either a *visual search* task or a *shape matching* task, each including four conditions: objects in the trained or the novel category, presented either in the trained (upright) or the inverted orientation. For the trained category, only novel

exemplars were used such that no names were associated with any objects for either group. Each run began with a 10s fixation period. Then the four conditions were presented in four 20s blocks of trials. This set of four conditions were presented four times in total (with order of the four conditions counterbalanced), separated by a 16s fixation period. The run ended with a 6s fixation period. Similar to the behavioral training, the stimuli were always presented in 8 possible locations, centered 3.1° - 3.5° from fixation along a circle (adjusted for the positions of the mirror with variable individual head sizes), and each object spanned a visual angle of 1.9° .

Visual search runs were modeled after Sigman et al. (2005) to look for neural training effects after PL training. Each block started with a target object presented for 2000ms at the center of the screen. Subjects were asked to search for this target object in six consecutive trials – it was present 50% of the time (Fig. 3). On each trial, an eight-object array was briefly presented for 150ms, followed by a central fixation for 2850ms, and objects were arranged in the same manner as the behavioral training and testing. Subjects responded by key presses, with the right index or middle finger for target present or absent trials respectively.

Shape matching runs were designed to tap into the neural changes following PE training. Each block included ten trials, in which two objects were simultaneously presented in opposite peripheral positions for 150ms, followed by a fixation dot for 1850ms (Fig. 3). Subjects judged whether the two objects were identical in shape by key presses, with the right index finger for ‘same’ and right middle finger for ‘different’ responses. Half of the trials were ‘same’ trials in each run.

Localizer runs included three types of stimuli: faces, common objects and scrambled objects. Each run began with a 10s fixation period, followed by six groups of 3 blocks (16s each), one for each object type with order counterbalanced, and ended with a 6s fixation period. Each block consisted of 16 trials, in which an object was presented for 750ms followed by a 250ms fixation. Subjects performed a one-back task, in which they pressed the right index finger key as fast as possible when they detected an object identical to the previous one. There were either two or three repeated trials in each block, with a repeat rate of 16.1%.

MRI Data Acquisition

Imaging was performed using a 3T Philips Intera Achieva scanner at the Institute of Imaging Science at Vanderbilt University. The blood oxygen level-dependent (BOLD) signals were collected using a T2*-weighted EPI acquisition (TE = 35ms, TR = 2000ms, flip angle = 79°, matrix size = 64x64, field of view = 192mm, 34 slices, slice thickness = 3mm with no gap). To increase coverage of the brain, the slices were tilted 10° from the horizontal plane so that the ventral temporal cortex and the occipital lobe were always covered, while portions of the superior parietal and superior frontal cortex may be left out due to individual differences in brain size. High-resolution T1-weighted anatomical volumes were also acquired using a 3D Turbo Field Echo (TFE) acquisition (TE = 4.6ms, TR = 8.9ms, flip angle = 8°, matrix size = 256x256, field of view = 256mm, 170 slices, slice thickness = 1mm with no gap).

fMRI Data Analysis

Data analysis was performed with Brain Voyager 1.10 (www.brainvoyager.com) and included 3D motion correction, slice scan time correction, temporal filtering (3 cycles / scan, high-pass), spatial smoothing (6mm FWHM Gaussian), and multi-study GLM (general linear model), treating subjects as a random factor. For each subject, all functional images in the two scans were co-registered to the anatomical images obtained during the pre-training scan. Data analyses included the 4th volumes onwards when the hemodynamic response should be at peak level. Whole brain analyses were focused on the posterior half of the brain (from Talairach coordinates $y = -20$ to occipital pole) and the active voxels (voxels that produced a significantly larger response for any of the conditions compared to fixation at the level of uncorrected $p = .01$) to ensure that analyses were performed on regions that were covered by functional scans in all subjects and to increase statistical power. Multiple comparisons were corrected by cluster thresholding method, where statistic images were assessed for cluster-wise significance using the Cluster Thresholding plugin in Brain Voyager, with a cluster-defining threshold of $p = 0.02$ and 2000 simulation trials. The 0.05 FWE-corrected critical cluster size was 9 or 11 voxels depending on the contrast (243 - 297 mm³). With the localizer, we identified bilateral face-selective areas (FFA) [faces – objects], bilateral ventral object-selective areas (LO) [objects – scrambled objects] and also bilateral parahippocampal gyri (PG) that responded more for objects than faces [objects – faces]. These ROIs were defined because they have been used in prior PE studies (e.g. Gauthier et al., 1999; Kourtzi & Dicarlo, 2006; Wong et al., 2009b; Wong & Gauthier, 2010; Xu, 2005). Note that these ROIs are defined at the group level with both training groups combined, with $p < .05$

corrected with false discovery rate (FDR; Genovese, 2002) except for the LFFA that required a lower threshold ($t > 2.0$), presumably because it tends to be smaller than the RFFA and more variable in location. ROI analyses were focused on the group level because unequal numbers of individual ROIs were obtained in the two groups, which made interpretation of the results difficult. In regions showing significant activation, further data analyses were performed with a region-of-interest $10 \times 10 \times 10 \text{ mm}^3$ in size, centered on the peak activity, either to extract descriptive statistics (to illustrate the pattern of a significant interaction) or for analyses on an independent part of the data set. Contiguous areas of activity were separated as multiple non-overlapping $10 \times 10 \times 10$ significant mm^3 areas if they consisted of multiple local peaks. We report spatial coordinates in Talairach space.

Results

Summary of behavioral findings

Behavioral results for this study are reported elsewhere (Wong, Folstein & Gauthier, 2011). The major findings are summarized below, to provide the context for our fMRI analyses.

PL

Similar to prior PL studies (Sigman et al., 2000; 2005), the PL group developed orientation-specific improvement for the visual search task (Table 1). A 2x2 ANOVA with Training (pretest / posttest) x Orientation (trained / inverted) on d' revealed a significant interaction between Training and Orientation, $F(1,11) = 63.9$, $p \leq .0001$ ($\eta_p^2 = .85$, $CI_{.95}$ of $\eta_p^2 = .55$ to $.91$; Fritz, Morris & Richler, 2011). Scheffé tests ($p < .05$)

revealed that an inversion effect was absent during pretest but was significant at posttest, with better performance for the trained than the inverted orientation. Similar analyses on novel exemplars of the trained category revealed a similar Training x Orientation interaction, $F(1,11) = 124.3$, $p \leq .0001$, suggesting that learning generalized to novel exemplars.

The PL improvement for the visual search task was also category-specific. A 2x2 ANOVA on Category (trained / novel) x Orientation (trained / inverted) on d' was performed on posttest results (Table 1). The Category x Orientation interaction was significant, $F(1,11) = 73.4$, $p \leq .0001$ ($\eta_p^2 = .87$, $CI_{.95}$ of $\eta_p^2 = .59$ to $.92$). Scheffé tests ($p < .05$) revealed that the performance for the trained objects was better than for novel objects, only for the trained orientation.

PE

As in prior PE studies (Gauthier et al., 1998; Wong et al., 2009), PE training improved shape discrimination performance for the trained compared to novel object category (Table 2). A one-way ANOVA on Training (pretest / posttest) on the noise threshold for the trained category revealed a main effect of Training, $F(1,11) = 6.71$, $p = .025$ (Cohen's $d = .86$, $CI_{.95}$ of $d = .09$ to 1.38), indicating that the amount of Gaussian noise required to keep subjects' accuracy at 80% increased after PE training. The improvement was category-specific since, after training, performance for the novel category was no different from pretest performance for the trained category.

Behavioral evidence of PE learning transferring to novel exemplars within the trained category was not obtained, which is different from prior studies (Gauthier et al., 1997, 1998, 2002). This was attributed to the fact that the only task (matching) that

included novel exemplars was not sensitive to training effects even with the trained exemplars. This may simply be a limitation of measurements, since other tasks showed training effects for the trained objects. Accordingly, we expected that fMRI would be more sensitive to reveal a generalization effect to novel exemplars, given that prior PE fMRI work only used novel exemplars and since training effects in PE studies can be obtained even in passive viewing or incidental tasks (e.g. Gauthier et al., 2000).

Generalization to untrained tasks

Both types of learning generalized to untrained tasks. For PL, performance improved for the untrained shape matching task (Table 2). A one-way ANOVA on Training (pretest / posttest) on the noise threshold for the trained category revealed a main effect of Training, $F(1,11) = 9.24$, $p = .011$ (Cohen's $d = .86$, $CI_{.95}$ of $d = .19$ to 1.54), indicating that more noise was required to keep accuracy at 80% during the untrained shape matching task after PL training. Similar to the PE group, the PL improvement in shape matching was specific to the trained category and was not observed for the novel category.

The PE group improved for the untrained visual search task (Table 1). For the trained exemplars, a 2x2 ANOVA with Training (pretest / posttest) x Orientation (trained / inverted) on d' revealed a significant main effect of Training, $F(1,11) = 11.5$, $p = .006$ ($\eta_p^2 = .51$, $CI_{.95}$ of $\eta_p^2 = .06$ to $.72$), with better performance at posttest than pretest. While Training x Orientation interaction was not significant for trained exemplars ($p < .2$), analyses with novel exemplars revealed a significant inversion effect (trained $d' >$ inverted d') at posttest but not at pretest, suggesting some degree of orientation specificity in the improvement of the untrained visual search task.

In sum, the PL and PE training effects replicated typical behavioral findings of PL and PE, and revealed that the behavioral improvement can generalize across untrained tasks.

Next, we report the fMRI results in four sections. First, we examined whether our PL and PE training protocols replicated the typical patterns of neural changes reported in prior studies. The PL and PE training effects were assessed using the visual search task and the shape matching task respectively. Second, we investigated how the current findings differ from prior studies, as we introduced some important departures from standard training paradigms here for the purpose of making PE and PL more comparable (e.g. involving task-irrelevant shape variability in the training objects for PL, or using parafoveal instead of foveal presentation for PE). Third, we compared the training effects between PL and PE with the same contrasts in the same task to test whether changes in object representation depended on the nature of training experience. Finally, we tested whether the PL or PE training effects were task-specific by comparing the learning patterns observed across the two testing tasks.

Typical training effects replicated

PL

An increased inversion effect, i.e., higher sensitivity to objects in the trained than an untrained orientation, has been the signature behavioral effect in PL training and was obtained for our subjects (Sigman et al., 2000; 2005; Wong et al., 2011). The same contrast was used to reveal the brain regions engaged by PL (Sigman et al., 2005; Lewis et al., 2009). We performed a whole brain analysis using the Orientation x Training

contrast [(trained – inverted orientation) x (post – pre scan)] with objects from the trained category to look for brain regions recruited after the PL training.

An increased inversion effect was observed in a widespread network of brain areas (Fig. 4A; Table 3). Importantly, early retinotopic areas showed an increased inversion effect after PL training, including different parts of the occipital pole and the calcarine fissure (Fig. 4A; Table 3). The recruitment of early visual regions by PL replicates findings from prior PL studies (Furmanski et al., 2004; Pourtois et al., 2008; Maertens & Pollmann, 2005; Schoups et al., 2001; Mukai et al., 2007; Schwartz et al., 2002; Yotsumoto et al., 2008; Lewis et al., 2009).

PE

Category selective learning is generally the focus of PE studies, with subjects performing better at shape discrimination with the trained object category compared to an untrained category (Gauthier et al., 1999; 2000; Wong et al., 2009b). This was investigated in a whole brain analysis using a Category x Training contrast [(trained – novel category) x (post – pre scan)] with objects in the trained orientation to look for brain regions showing increased category selectivity after the PE training.

Increased category selectivity was observed in bilateral ventral temporal cortex, including bilateral inferior temporal regions and the left middle temporal area, and the LIPS (Fig. 5D; Table 4). We quantified the amount of behavioral improvement using the noise threshold for 80% accuracy during shape matching, contrasting the trained and novel category after training (for details of the noise manipulations during pretests or posttests, see Wong et al. 2011), and found that the increase in category selectivity in one

of these ventral temporal areas (the left middle temporal area; -63, -37, -6; 288mm³) predicted this behavioral advantage for the trained category, $r = .66$, $p = .02$ (Fig. 6A).

Separate ROI analyses (see Methods) further revealed that PE training engaged the face- and object-selective regions. The increase in category selectivity predicted behavioral improvement (defined as above) in the LFFA ($r = .61$, $p = .034$; Fig. 6B), the LLO ($r = .61$, $p = .034$; Fig. 6C), and the RPG ($r = .59$, $p = .045$, Fig. 6D).

In sum, the recruitment of ventral temporal cortex for the trained category, with the magnitude of category selectivity predicting behavioral effects, replicates PE training effects in prior studies (Gauthier et al., 1999; Gauthier & Tarr, 2002; Op de Beeck et al., 2006; Yue et al., 2006; Xu, 2005; Wong et al., 2009b).

Comparing current findings with prior work

PL

Compared to prior studies with similar designs (Sigman et al., 2005; Lewis et al., 2009), our results were qualitatively different in several ways. For example, an increased inversion effect was found in bilateral higher visual cortex extensively, covering bilateral lingual gyrus and fusiform gyrus (Fig. 4A, Table 3), in contrast to prior studies in which the higher visual cortex was only recruited in a highly localized region (Lewis et al., 2009), or showed a negative inversion effect after training, i.e., the neural activity was higher for the inverted than upright condition (Sigman et al., 2005). The ROI analysis further revealed that our PL training engaged some brain regions selective for faces and objects. A 2x2 ANOVA with Orientation x Training revealed an increased inversion effect after training in the LFFA, $F(1,11) = 6.91$, $p = .023$ ($\eta_p^2 = .39$, $CI_{.95}$ of $\eta_p^2 = .00$ to

.64; Fig. 7A) and the LPG, $F(1,11) = 9.04$, $p = .012$ ($\eta_p^2 = .45$, $CI_{.95}$ of $\eta_p^2 = .03$ to .68; Fig. 7B).

In addition, our PL protocol led to increased inversion effects in a large part of the dorsal network, with local deactivations in bilateral superior parietal regions (Fig. 4A; Table 3). These were in contrast to the global deactivation in the dorsal network in previous studies (Sigman et al., 2005; Lewis et al., 2009). Finally, bilateral STS and the cingulate gyrus also showed increases in inversion effects after PL. Engagement of these two areas were not reported in previous studies (Sigman et al., 2005; Lewis et al., 2009).

The extensive neural network showing an increased inversion effect after PL training was not driven by pre-training differences, because no voxels with significant inversion effect were found before training (trained – inverted orientation, pre-scan only; Fig. 8).

Since the increased inversion effect in behavioral learning in PL was also specific to the trained object category (Wong et al., 2011), we performed a whole-brain analysis with the Orientation x Category x Training [(trained – inverted orientation) x (trained – novel category) x (post – pre scan)] to explore which brain regions are related to the category-specific inversion effect. Result revealed a widespread neural network covering both higher visual cortex and the dorsal network (Fig. 9), suggesting that a wide range of neural areas may support the orientation- and category-specific behavioral learning in PL.

In sum, our PL training engaged the early retinotopic cortex, a typical finding in prior PL studies (Furmanski et al., 2004; Pourtois et al., 2008; Maertens & Pollmann, 2005; Schoups et al., 2001; Mukai et al., 2007; Schwartz et al., 2002; Yotsumoto et al., 2008; Lewis et al., 2009). The qualitative differences between the current and prior

studies, including the extensive recruitment of bilateral higher visual cortex, the dorsal network, the STS and the cingulate gyrus, is probably a result of using more complex object silhouettes with shape variability instead of simple Gabor filters or shapes (see Discussion).

PE

The occipito-temporal areas showing increased category selectivity were somewhat different from those observed in some other PE studies. For example, prior work using Ziggerins (Wong et al., 2009b) or other novel objects (Gauthier et al., 1999) found increased category selectivity in the RFFA, while we observed similar effects in the LFFA instead of the RFFA. This could be related to the use of parafoveal presentation in our PE protocol, which could be further explored in future studies.

Training effects depend on training experience

Visual search engaged different neural substrates in PL vs. PE

With different training experience, the PE training led to cortical changes that were qualitatively different from that of the PL group during visual search. In the same analysis, no training effect was found in early retinotopic cortex (Fig. 4C; Table 5). Also, in higher visual cortex (e.g. the left fusiform gyrus and the middle temporal gyrus) and various parietal regions in the dorsal pathway, the inversion effect decreased after training. This was in contrast to the PL group in which a widespread neural network showed increased inversion effects after training (Fig. 4A; Table 3).

The group differences were confirmed by analyses directly comparing the two groups. The Group x Orientation x Training contrast [(PL – PE) x (trained – inverted orientation) x (post – pre scan)] was performed with objects from the trained category. An extensive neural network showed significantly larger increase in the inversion effect for PL than PE (Table 6; Fig. S1), similar to that in Fig. 4A.

ROI analyses also revealed qualitatively different visual learning between the two groups. In the LFFA, a 2x2 ANOVA with Orientation x Training revealed that the inversion effect increased after PL (Fig. 7A) but not after PE ($F_{1,11} < 1$). In both left and right PG, a 2x2x2 ANOVA with Group x Orientation x Training revealed significant three-way interactions in both the LPG, $F(1,22) = 10.4, p = .0039$ ($\eta_p^2 = .49$, $CI_{.95}$ of $\eta_p^2 = .05$ to $.70$; Fig. 7B), and the RPG, $F(1,22) = 4.35, p = .049$ ($\eta_p^2 = .28$, $CI_{.95}$ of $\eta_p^2 = .00$ to $.57$; Fig. 7C). For both areas, the inversion effect was not found before training for either group, but was found after training for PL (Scheffé tests, $p < .05$) but not for PE.

Therefore, the visual search task produced qualitatively different changes in orientation selective object representations depending on training experience.

Apart from the Orientation x Training contrast, qualitatively different training effects can also be found with the Category x Training contrast [(trained – novel category) x (post – pre scan), for objects in the trained orientation]. As shown in figure 5A and 5C, PL led to increase in category selectivity in early retinotopic cortex, higher visual cortex and the dorsal pathway, while PE resulted in decrease in category selectivity in the higher visual cortex and the parietal regions. These results demonstrate that the specific contrast used is not critical here. PL and PE show qualitatively different visual

learning effects in the visual search task, regardless of whether the baseline is an untrained orientation or a novel category of objects.

Shape matching engaged different neural substrates in PL vs. PE

During shape matching, PL resulted in a different pattern of neural changes after training compared to that in PE. The Category x Training contrast did not reveal an increase in category selectivity in the ventral temporal cortex after PL training (Fig. 5B; Table 7). Instead, increased category selectivity was observed in the LIPS, a region similar to that engaged by the PE group in the same contrast, and in an additional RSTS region. Directly comparing the two groups with the Group x Category x Training contrast did not reveal any significant difference in these regions during shape matching, perhaps because of limited statistical power for a between-subject contrast. This is consistent with behavioral results showing more similarity between the two groups during shape matching than visual search tasks (Wong et al., 2011). Although we cannot conclude that the training effects in terms of category selectivity for trained vs. novel objects differ across groups, an analysis restricted to the trained-upright objects revealed a significant interaction between groups. Specifically, the Group x Training contrast [(PL – PE) x (post – pre scan)], with trained-upright objects only, revealed significant interactions with increased activity for PE more than for PL in bilateral ventral temporal regions, and larger increases for PL than PE in the RSTS (Table 8; Fig. S1).

The ROI analyses in bilateral FFA, LO or PG did not reveal any change of category selectivity after PL training, in terms of the mean category selectivity (all $ps > .3$), or correlation between the change in category selectivity and behavioral improvement

(defined as that in the PE group; all $ps > .2$). This is in contrast to the PE group that engaged multiple regions in the ventral temporal cortex and the face- and object-selective regions for shape matching.

It is difficult to compare training effects across groups using the Orientation x Training contrast [(trained – inverted orientation) x (post – pre scan)], because this contrast only revealed limited training effects within each group (Fig. 4B & 4D; all $ps > .2$ for ROI analyses with this contrast). This is perhaps not surprising for PE, since a significantly increased inversion effect after PE training has only been obtained once (Gauthier et al., 1999), while changes in category selectivity has been considered a more reliable and robust training effect for PE (e.g., Gauthier et al., 2000; Xu, 2005; Op de Beeck et al., 2006).³

In sum, the trained objects in the trained orientation engaged different neural substrates during the shape matching task, depending on training experience. It is worth noting that the PL group's performance in shape matching improved at least as much as that of the PE group (Wong et al., 2011; Table 2). Therefore, similar behavioral improvements in shape matching ability in the two groups are associated with different neural mechanisms. For PE, the shape matching ability may be supported by multiple regions in the ventral temporal cortex and the face- and object-selective regions, while that for the PL group may be supported by the inferior parietal area and the RSTS.

Training effects depend on testing tasks

³ It is possible that the representational changes that support PE and allow for generalization to new exemplars very quickly (e.g., McGugin et al., in press) are not initially specific enough to produce inversion effects.

Training effects do not only depend on the nature of training experience, but also appeared dependent on the testing tasks. Using identical contrasts, the training effects observed in each group were different during visual search and shape matching. For PL, both the increased inversion effect (Fig. 4A) and the increased category specificity (Fig. 5A) were found in a widespread set of areas during visual search but not during shape matching (Fig. 4B, 5B). For PE, the inversion effect and category selectivity in ventral temporal areas and left parietal regions were increased during shape matching (Fig. 4D, 5D), but decreased during visual search (Fig. 4C, 5C).

This interaction between testing task and prior experience can be observed not only in distributed patterns of activity but also within small local regions. For example, in the LFFA, an increased inversion effect was found for PL during visual search ($p = .023$) but not during shape matching ($F < 1$), and this effect was not found for PE in either task (both $F_s < 1$). These results suggest that both the change in inversion effects and the category selectivity were dependent on training experience and testing tasks.

Discussion

In this study, we investigated why PL and PE studies consistently obtain contrasting patterns of neural training effects. We found that the nature of training experience alone is sufficient to explain the typical training effects of PL and PE, including the increased selectivity for objects in the trained orientation in early retinotopic cortex in PL (but not in PE) during visual search, and the increased category selectivity for trained objects in higher visual cortex in PE (but not in PL) during shape matching. These suggest that the qualitatively different patterns of learning effects

reported in the two literatures may be largely driven by differences in the training experience, rather than other factors matched in the current study such as object sets, parafoveal stimulus presentation and training duration. The divergent patterns of results in PL and PE cannot be explained by different levels of attention or differential degrees of learning in the two groups. Our findings demonstrate the critical importance of the nature of experience with an object's category as one of the factors determining its representation in the visual system (Gauthier et al., 1998; Wong et al., 2009b; Gauthier, Wong & Palmeri, 2010).

Implications for Perceptual Learning

For PL, despite the fact that our training objects were much more complex and variable than in prior work, activity increased in early retinotopic cortex during visual search for targets in the trained orientation compared to inverted targets, consistent with the engagement of early visual areas for PL (Furmanski et al., 2004; Pourtois et al., 2008; Maertens & Pollmann, 2005; Schoups et al., 2001; Mukai et al., 2007; Schwartz et al., 2002; Yotsumoto et al., 2008; Lewis et al., 2009). The results are consistent with the hypothesis that rapid recognition of multiple simultaneously presented objects leads to the recruitment of early retinotopic cortex because its high spatial resolution helps rapid access of the stimuli presented in different visual field positions (Sigman et al., 2005).

However, changes in the dorsal network and in higher visual areas in our PL were different from previous studies (Sigman et al., 2005; Lewis et al., 2009). In prior work, the opposite learning effects in different visual areas, i.e., decreases in extrastriate areas associated with increases in retinotopic cortex, were taken as evidence that large-scale

reorganization between visual areas supported learning in this task (Sigman et al., 2005). However, our results indicate that changes in early visual cortex can be obtained without a corresponding decrease of activity in higher visual regions or in the dorsal attention network (both were more active for the trained orientation in the current study). Our results suggest that PL can lead to large-scale changes in object representation, while the roles of the higher visual areas and the dorsal network may be determined by properties of the training objects. For example, it is possible that PL with shape variability in objects required influences from higher areas to establish the correct template in early visual areas for the search task. It is worth noting that the similarities and differences in neural engagement in the current and previous PL studies are all accompanied by common behavioral improvement highly specific to the training orientation and category. This demonstrates how similar behavioral effects can be supported by different patterns of changes in the visual system.

Implications for Perceptual Expertise

Our modified PE training led to increases in category selectivity in multiple regions in the ventral temporal cortex. In particular, the change in category selectivity in the left temporal area, LFFA, LLO and RPG predicted behavioral improvement, suggesting that the enhanced shape discrimination ability is related to computations in these regions. These results are consistent with prior PE studies that shape discrimination learning recruits higher visual cortex (Gauthier et al., 1999; Gauthier & Tarr, 2000; Op de Beeck et al., 2006; Yue et al., 2006; Xu, 2005; Wong et al., 2009; van der Linden, Murre

& van Turennout, 2008; van der Linden, van Turennout & Indefrey, 2010). While our modified PE training engaged different parts of the higher visual cortex compared to prior studies (e.g. Gauthier et al., 1999; Wong et al., 2009), the differences may be caused by the parafoveal object presentation instead of foveal presentation.

Even though the PL and PE training led to similar degrees of improvement in shape matching (Wong et al., 2011), the neural mechanisms supporting such behavioral learning are largely different across the two groups. While prior PE studies and the current findings consistently demonstrated that PE is associated with ventral temporal cortex (Gauthier et al., 1999; Gauthier & Tarr, 2002; Op de Beeck et al., 2006; Yue et al., 2006; Xu, 2005; Wong et al., 2009; van der Linden, Murre & van Turennout, 2008; van der Linden, van Turennout & Indefrey, 2010), increased category selectivity in the PL group was found in the LIPS and RSTS instead of the ventral temporal regions. The results were not strong enough to support a group difference in category selectivity (for trained vs. novel objects), but we nonetheless observed the engagement of different neural areas across groups for trained-upright objects during shape matching: ventral temporal regions showed a larger increase for PE than PL, while the right STS showed a larger increase for PL than PE. These difference in neural training effects cannot be explained by differences in behavioral learning across groups, since the behavioral improvement for both groups were of similar magnitude and shared similar category specificity. These differences in neural recruitment may be related to the fact that shape discrimination was explicitly trained in PE but acquired in a task-irrelevant manner in PL (in which the training was related to visual search instead of fine-level shape discrimination; Wong et al., 2011). Our results highlight the importance of *how* a certain

visual ability is acquired: similar behavioral learning can be supported by distinct neural mechanisms when the ability is learned through different types of training.

Relationship between Perceptual Learning and Perceptual Expertise

In the literature, PL and PE have been treated as different types of visual training studies. For example, PE (but not PL) is thought to involve explicit and semantic memory (e.g. the use of naming training) and thus recruit temporal regions, while PL is regarded as more implicit and perceptual and thus engages early visual areas (Fine & Jacobs, 2002; Gilbert et al., 2001). However, we showed that PL can recruit ventral temporal regions extensively, indicating that the engagement of ventral temporal areas in visual learning is not necessarily related to naming.

PL and PE are subsets of visual learning studies that are perhaps not that distinct from each other. For example, both PL and PE may engage higher visual cortex (e.g. the present study) and early visual cortex (e.g. Sigman et al., 2005; and Schoups et al., 2001 for PL; Wong & Gauthier, 2010 for PE). On the behavioral level, the degree of specificity in PL, the signature of PL, did not appear to be lower than PE when they were directly compared using the same testing tasks (Wong et al., 2011). In addition, improvements in PL have been associated to changes in the intraparietal regions and the medial frontal cortex, suggesting the involvement of decision-making and reward mechanisms in PL (Kahnt et al., 2011; Law & Gold, 2008). While prior PE studies did not investigate the role of decision-making or reward mechanisms, multiple intraparietal and medial frontal regions are engaged after PE training (Table 4-5; Wong & Gauthier, 2010), suggesting similar possibilities of involving higher mechanisms in PL and PE. In general, we did not

observe strong evidence supporting the idea that PL and PE should be considered separately (Fine & Jacobs, 2002; Gilbert et al., 2001).

Apparent discontinuities between PL and PE (and perhaps with other visual learning studies) may be salient because the space of task constraints is sparsely sampled in the literature. In the present study, training effects in PL and PE can be explained by differences in training tasks, suggesting that these two types of training may fall onto different positions along a continuum of task demands. Indeed, task constraints across training studies often differ in numerous ways. For example, our PL and PE training protocols differed along multiple dimensions, including types of discrimination (orientation / shape), amount of visual crowding on each display, and number of responses used on training tasks. To bridge the literatures between PL and PE, one should consider the differences in task demands between these two types of training.

Importance of testing task

Patterns of short-term visual learning do not only depend on the nature of training experience, but also the testing task. Within the same group of subjects, the observed training effects can be qualitatively different when the testing task was visual search or shape matching. This task dependency of the training effects was found when different contrasts were used, suggesting that this finding was not a result of specific contrasts, but rather may well represent learning effects that are not expressed in some testing conditions.

For PL, such task dependency differs from what prior studies seem to predict. Specifically, while the neural substrates of PL is often thought to be task-specific (Fahle,

2009; Gilbert et al., 2001; 2007; Li et al., 2004), this typically means that the training effects can be observed only during the trained task. Here, although we observed that the exact patterns of results were highly dependent on the specific tasks, we found training effects in both the trained task (visual search) and a task irrelevant to the PL training (shape matching). Using tasks that are very different from the training task may help to differentiate the role of various brain regions in visual learning.

It is possible that the task-specificity may decrease with additional training. For example, a prior study with Ziggerins used six object categories in the training and thus provided about 1/6 of the experience per object category compared to the present PE training (Wong et al., 2009). The task-specificity of their neural training effect was even higher than the current study in that their training effects were observed only for testing tasks that were highly similar to the training. It is possible that as we learn a category, visual learning effects progress from being specific to the training task, to generalizing to other tasks and eventually showing a relatively stable pattern where regardless of the task, the pattern of activity resembles that for the practiced task (Gauthier et al., 2000). For instance, one hallmark of expertise individuating objects is a relatively inflexible tendency to apply holistic strategies even when the task calls for part-based attention (Bukach et al., 2010).

Manipulating experience for developing visual learning theories

In the face of the many factors that can drastically influence how the visual system is affected by visual learning, and the possibility that these factors can interact as training task and testing task did in our study, how can a systems-level theory of visual

learning be possible? Given a detailed description of objects and training conditions, how can we predict which visual areas will show the greatest amount of learning? Can we do better than to state post-hoc that the neural substrates recruited are the ones that are the most informative for the task at hand (Op de Beeck & Baker, 2010)?

While we are optimistic that neuroimaging can help reveal how different aspects of experience constrain patterns of visual learning, the empirical evidence required for such models is currently lacking, in particular because of the need for manipulations of prior experience as we used here. Such designs are not common (Wong et al., 2009; Song et al., 2010), as most fMRI studies of visual learning only look at the changes that follow a single training protocol (e.g., Sigman et al., 2005; Op de Beeck et al., 2006; Jiang et al., 2007; Kourtzi et al., 2005; Schwartz et al., 2002; Mukai et al., 2007; Yotsumoto et al., 2008; Lewis et al., 2009). In such cases, training effects cannot be attributed to specific aspects of experience: even mere exposure could potentially explain these results. For this reason, we know a lot less about the role of experience than we know about factors that are frequently manipulated, such as object category or shape (e.g., Haxby et al., 2001; Eger et al., 2008; Grill-Spector, Sayres & Ress, 2006; Golcu & Gilbert, 2009).

Arguing that several factors interact and provide constraints for visual learning to influence category-selectivity is not a completely new idea (Gauthier, 2000; Malach et al., 2002; Op de Beeck et al., 2008; Op de Beeck & Baker 2010). Our daily experience with objects is associated with different goals, which may interact with other factors such as object geometry (Tanaka, 1996; Op de Beeck, 2010), eccentricity (Malach & Hasson, 2002) and non-visual experience with objects (Wong et al., 2010; James et al., 2005) to shape the topographical map of selectivity for various categories of objects (Grill-Spector

and Malach, 2004). Some results also suggest that *visual* experience may not be necessary for at least some of the coarse organization in the visual system (Mahon et al., 2009).

The contribution of the present work is to show that experience alone is sufficient to drive differences as large as those reported in the PL and PE literatures. Although this goes beyond the current data, it is logically possible that prior experience with objects could in fact constrain the shape-selective effects obtained in experiments where experience is not manipulated. For instance, in an experiment where entirely novel objects vary in shape such that some are smooth and others are spiky (e.g., Op de Beeck et al., 2006), the maps of shape-selectivity could be determined by prior experience with other spiky and smooth objects. In the light of the present demonstration that the nature of our experience with an object's category can influence its representation in the visual system, it appears premature to suggest that learning only moderately alters pre-existing object representations and does so only in a very distributed fashion (Op de Beeck & Baker, 2010; Op de Beeck et al., 2007; see also Freedman & Miller, 2008). Controlled manipulations of prior experience such as we used here can help us understand the principles governing the functional plasticity of the visual system.

Acknowledgements

This work was funded by NIH grants 1 F32 EY019445-01, 2 RO1 EYO13441-06A2, P30-EY008126 and by the Temporal Dynamics of Learning Center, NSF grant SBE-0542013.

References

- Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Res*, 27(6), 953-965.
- Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends Cogn Sci*, 10(4), 159-166.
- Bukach, C. M., Phillips, W. S., & Gauthier, I. (2010). Limits of generalization between categories and implications for theories of category specificity. *Atten Percept Psychophys*, 72(7) 1865-74.
- Busey, T., & Vanderkolk, J. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research*, 45(4), 431-448.
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., et al. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123(Pt 2), 291-307.
- Eger, E., Ashburner, J., Haynes, J. D., Dolan, R. J., & Rees, G. (2008). fMRI activity patterns in human LO carry information about object exemplars within category. *J. Cogn Neuroci*, 20, 356-370.
- Fahle, M. (1997). Specificity of learning curvature, orientation, and vernier discriminations. *Vision Res*, 37(14), 1885-1895.
- Fahle, M. (2009). Perceptual learning and sensorimotor flexibility: Cortical plasticity under attentional control? *Phil. Trans. R. Soc. B* 2009 364, 313-319.
- Fahle, M. & T. Poggio (2002). *Perceptual Learning*. Massachusetts: MIT Press.

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity.

Vision Res, 35(21), 3003-3013.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning tasks: a review. *Journal*

of vision, 2(2), 190-203.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, 287(5777), 43-44.

Fiorentini, A., & Berardi, N. (1981). Learning in grating waveform discrimination: specificity for orientation and spatial frequency. *Vision Res*, 21(7), 1149-1158.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2011). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*.

Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Curr Biol*, 14(7), 573-578.

Gauthier, I. (2000). What constrains the organization of the ventral temporal cortex?

Trends Cogn Sci (Regul Ed), 4(1), 1-2.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Research*, 37(12), 1673-1682.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition:

Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perceptual and Performance*, 28(2), 431-446.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci*, 3(2), 191-197.

- Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research*, 38(15/16), 2401-2428.
- Gauthier, I., Wong, A. C.-N., & Palmeri, T. J. (2010). Manipulating visual experience: comment on Op de Beeck and Baker. *Trends Cogn Sci*, 14(6), 235-6.
- Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15, 870-878.
- Gilbert, C. D., & Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron*, 54(5), 677-696.
- Gilbert, G. D., Sigman, M. & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, 31, 681-697.
- Golcu, D., & Gilbert, C. D. (2009). Perceptual learning of object shape. *The Journal of Neuroscience*, 29(43), 13621-13629.
- Grill-Spector, K., Sayres, R., & Ress, D. (2006). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nat Neurosci*, 9(9), 1177-85.
- Harel, A., Gilaie-Dotan, S., Malach, R., & Bentin, S. (2010). Top-down engagement modulates the neural expressions of visual expertise. *Cerebral Cortex*, 20, 2304-18.
- Harley, E. M., Pope, W. B., Villablanca, J. P., Mumford, J., Suh, R., Mazziotta, J. C., Enzmann, D., & Engel, S. (2009). Engagement of fusiform cortex and disengagement of lateral occipital cortex in the acquisition of radiological expertise. *Cerebral Cortex*, 19, 2746-54.

- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425-30.
- James, K. H., James, T. W., Jobard, G., Wong, A. C., & Gauthier, I. (2005). Letter processing in the visual system: different activation patterns for single letters and strings. *Cognitive, affective & behavioral neuroscience*, *5*(4), 452-466.
- Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, *53*(6), 891-903.
- Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of National Academy of Sciences of the United States of America*, *88*, 4966-4970.
- Kayaert, G., Biederman, I., & Vogels, R. (2005). Representation of regular and irregular shapes in macaque inferotemporal cortex. *Cerebral Cortex*, *15*, 1308-21.
- Kourtzi, Z., & Dicarlo, J. J. (2006). Learning and neural plasticity in visual object recognition. *Current Opinion in Neurobiology*, *16*, 152-158.
- Lewis, C. M., Baldassarre, A., Committeri, G., Romani, G. L., & Corbetta, M. (2009). Learning sculpts the spontaneous activity of the resting human brain. *Proc Natl Acad Sci USA*, *106*(41), 17558-63.
- Li, W., Piëch, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nat Neurosci*, *7*(6), 651-657.

- Maertens, M., & Pollmann, S. (2005). fMRI reveals a common neural substrate of illusory and real contours in V1 after perceptual learning. *Journal of cognitive neuroscience, 17*(10), 1553-1564.
- Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron, 63*, 397-405.
- McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences, 7*(7), 293-299.
- Moore, C. D., Cohen, M. X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *J Neurosci, 26*(43), 11187-11196.
- Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., & Ungerleider, L. G. (2007). Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. *J Neurosci, 27*(42), 11401-11411.
- Op de Beeck, H. P., & Baker, C. I. (2010) The neural basis of visual object learning. *Trends in Cognitive Sciences, 14*(1), 22-30.
- Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *J Neurosci, 26*(50), 13025-13036.
- Op de Beeck, H. P., Wagemans, J., & Vogels, R. (2007). Effects of perceptual learning in visual backward masking on the responses of macaque inferior temporal neurons. *Neuroscience, 145*, 775-789.

- Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, *256*(5059), 1018-1021.
- Pourtois, G., Rauss, K. S., Vuilleumier, P., & Schwartz, S. (2008). Effects of perceptual learning on primary visual cortex activity in humans. *Vision Res*, *48*(1), 55-62.
- Rossion, B., Kung, C.-C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 14521-14526.
- Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, *412*(6846), 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*(26), 17137-17142.
- Sigman, M., & Gilbert, C. D. (2000). Learning to find a shape. *Nat Neurosci*, *3*(3), 264-269.
- Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron*, *46*(5), 823-835.
- Song, Y., Hu, S., Li, X., & Liu, J. (2010). The role of top-down task context in learning to perceive objects. *The Journal of Neuroscience*, *30*(29), 9869-76.
- Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*(1), 43-47.

- Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological science : a journal of the American Psychological Society / APS*, *16*(2), 145-151.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu Rev Neurosci*, *19*, 109-139.
- Van der Linden, M., Murre, J. M. J., & van Turennout, M. (2008). Birds of a feather flock together: Experience-driven formation of visual object categories in human ventral temporal cortex. *PLOS One*, *3*(12), e3995.
- Van der Linden, M., van Turennout, M., & Indefrey, P. (2010). Formation of category representations in superior temporal sulcus. *Journal of Cognitive Neuroscience*, *22*(6), 1270-1282.
- Wong, A. C.-N., Palmeri, T., & Gauthier, I. (2009a). Conditions for face-like expertise with objects: Becoming a Ziggerin expert - but which type? *Psychological Science*, *20*(9), 1108-1117.
- Wong, A. C.-N., Palmeri, T., Rogers, B. P., Gore, J. C., & Gauthier, I. (2009b). Beyond shape: How you learn about objects affects how they are represented in visual cortex. *PLoS One*, *4*(12), e8405.
- Wong, Y. K., & Gauthier, I. (2010). A multimodal neural network recruited by expertise with musical notation. *Journal of Cognitive Neuroscience*, *22*(4), 695-713.
- Wong, Y. K., Folstein, J. R., & Gauthier, I. (2011). Task-irrelevant perceptual expertise. *Journal of Vision*, *11*(14):3, 1-15.
- Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cereb Cortex*, *15*(8), 1234-1242.

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.

Yue, X. M., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision*

Research, *46*(22), 3802-3811.

Figure captions

Figure 1. The two sets of Ziggerins in silhouette formats used for the training. The brackets illustrate the objects used as training exemplars, novel exemplars and novel category for one subject (counterbalanced across subjects).

Figure 2. The training paradigm used in the training. For PL (top), a visual search task was used in which subjects judged whether an object in the target orientation was present. For PE (bottom), a naming task was used which required subjects to name each individual object.

Figure 3. The task used in the scanner. In the visual search run (top), a target object was first presented for 2s, followed by 6 trials in which subjects were required to indicate whether objects in the same orientation as the target object was present (all objects in the display were identical in shape as the target object). In the shape matching run (bottom), subjects judged whether the two objects presented in each trial were identical in shape.

Figure 4. Areas showing significant increase (orange) or decrease (blue) using the Orientation x Training contrast during visual search or shape matching. Only the regions included in the whole brain analyses (posterior half of the brain, $y = -20$ to the occipital pole) were shown. Results were superimposed on the flattened cortical map of the left hemisphere of one of the PE subjects, only one subject was used to present results for both groups to maximize the ease of comparison in the figures. PCS – post-central sulcus; STS – superior temporal sulcus; pFs – posterior fusiform gyrus.

Figure 5. Areas showing significant increase (orange) or decrease (blue) using the Category x Training contrast during visual search or shape matching. The flattened maps are presented in the same manner as in Figure 4.

Figure 6. Scatter plots showing significant correlations for the PE group between behavioral improvement in shape matching (indicated by the difference in noise level for 80% matching accuracy for trained and novel object categories after training) and category selectivity during shape matching (defined by the Category x Training contrast) in the left middle temporal area (A), LFFA (B), LLO (C) and RPG (D).

Figure 7. Inversion effects during pre-scan (open squares) and post-scan (closed circles) for the two groups in the LFFA (A), LPG (B) and RPG (C). Error bars show the 95% CI of the Group x Orientation x Training interaction.

Figure 8. Areas showing significant increase (orange) or decrease (blue) using the Orientation contrast (trained – inverted) during visual search at pre-scan (left) or post-scan (right). PCS – post-central sulcus; STS – superior temporal sulcus; pFs – posterior fusiform gyrus.

Figure 9. Areas showing significant increase in response using the Orientation x Category x Training contrast during visual search for PL. PCS – post-central sulcus; STS – superior temporal sulcus; pFs – posterior fusiform gyrus.

Table captions

Table 1. Discriminability (d') for each condition for PL and PE during the visual search test before and after training. For PL, discriminability specifically improved for the trained orientation and for the trained object category after training, but transferred completely to novel exemplars within the trained category. For PE, visual search performance also improved after the shape individuation training.

Table 2. Performance (in terms of the estimated level of Gaussian noise required for about 80% accuracy) for PL and PE during the shape matching test before and after training (for upright objects only). For both groups, shape matching performance improved for the trained category but not for the novel category after training.

Table 3. Areas showing significant Orientation x Training effects for PL during visual search.

Table 4. Areas showing significant Category x Training effects for PE during shape matching.

Table 5. Areas showing significant Orientation x Training effects for PE during visual search.

Table 6. Areas showing significant Group x Orientation x Training effects during visual search. All areas showed a stronger Orientation x Training effect in the PL group compared with the PE group.

Table 7. Areas showing significant Category x Training effects for PL during shape matching.

Table 8. Areas showing significant Group x Training effects (for trained-upright objects) during shape matching. The response in these brain regions showed a stronger training effect for one group as indicated.

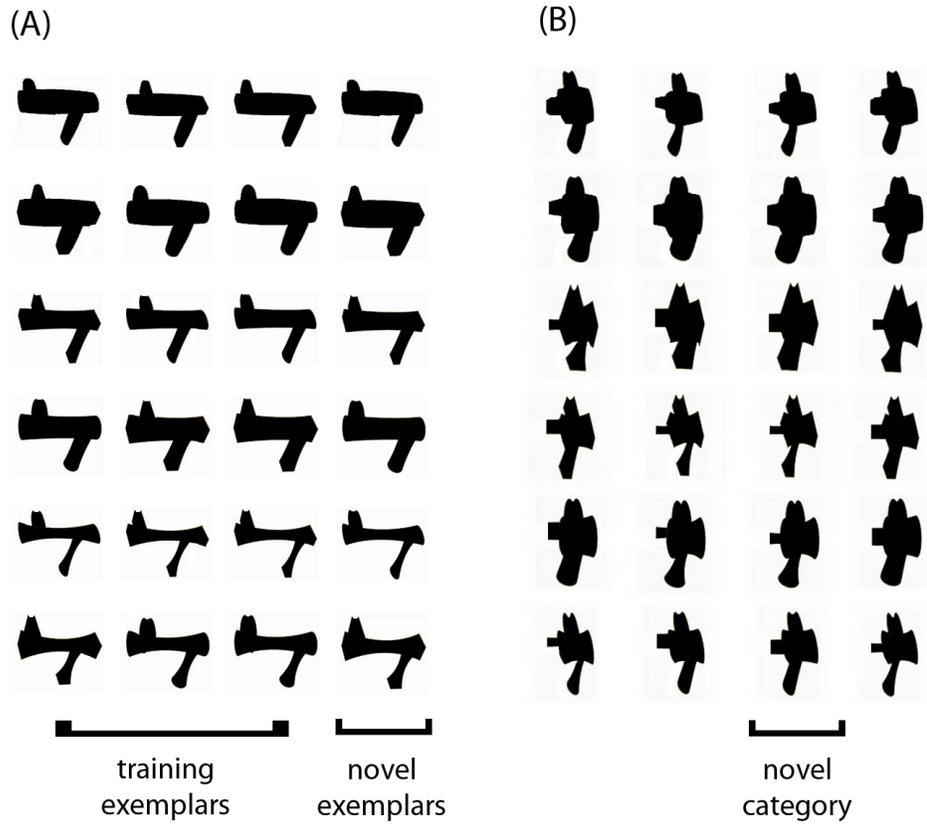


Figure 1

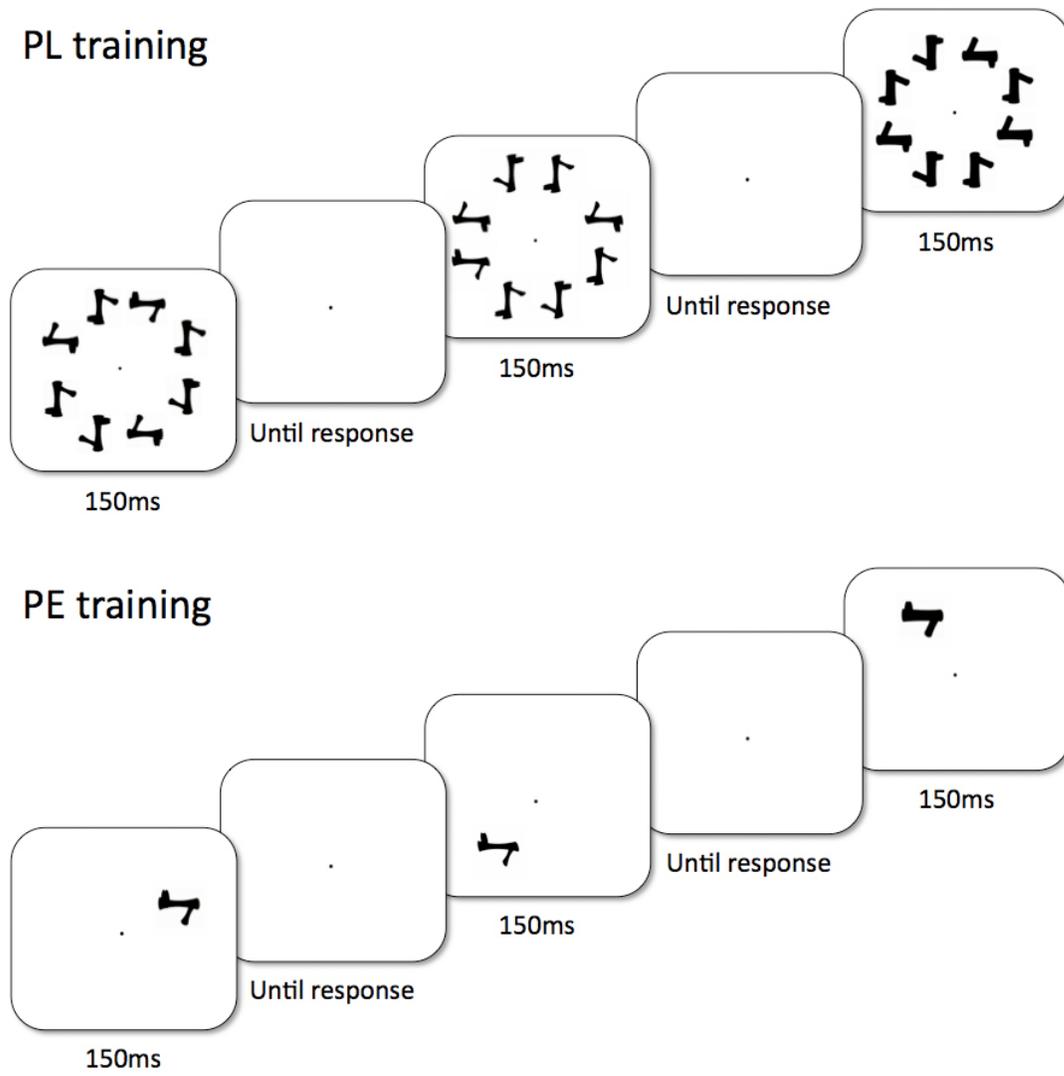
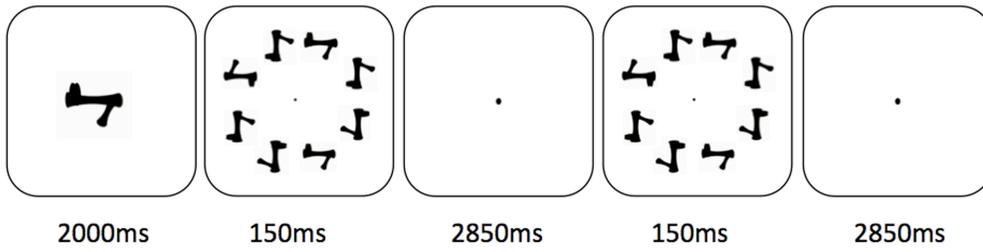


Figure 2

Visual Search Run



Matching Run

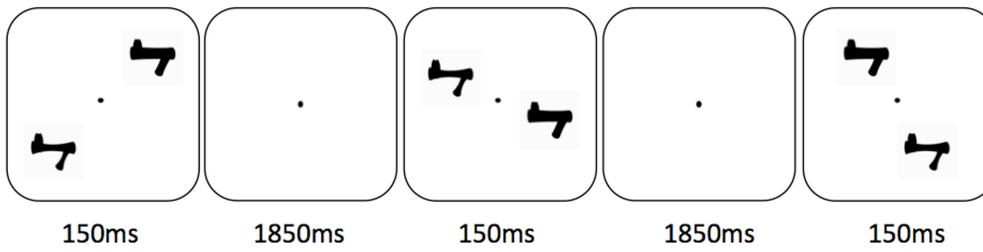


Figure 3

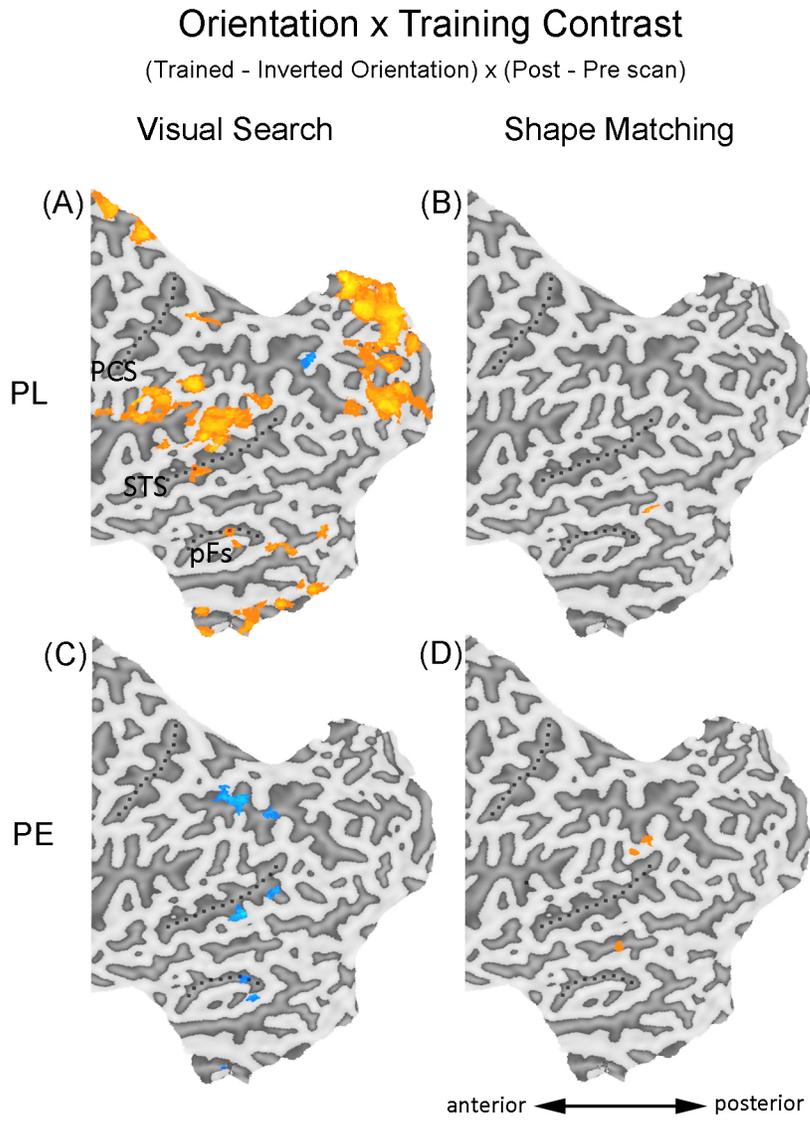


Figure 4

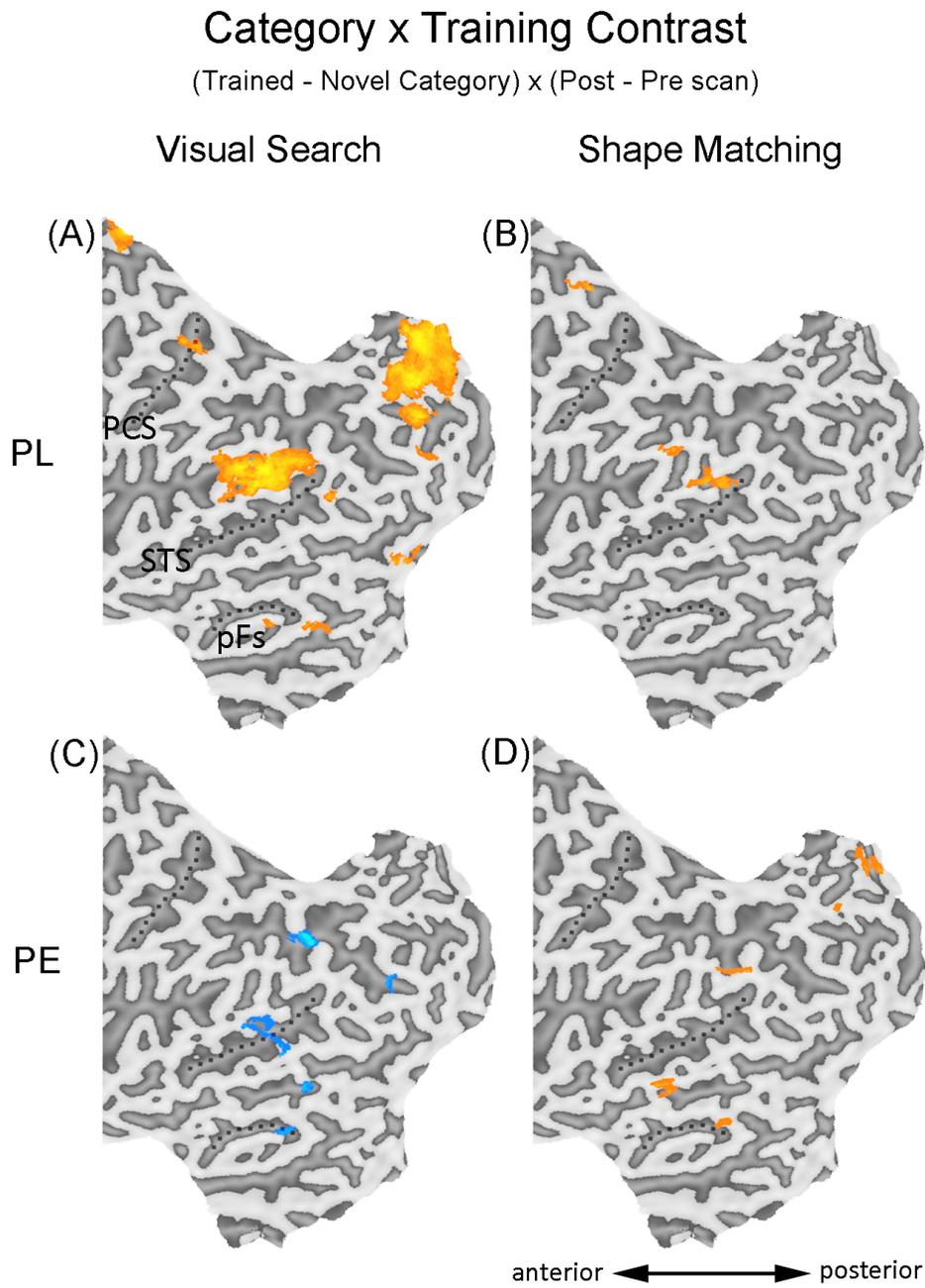


Figure 5

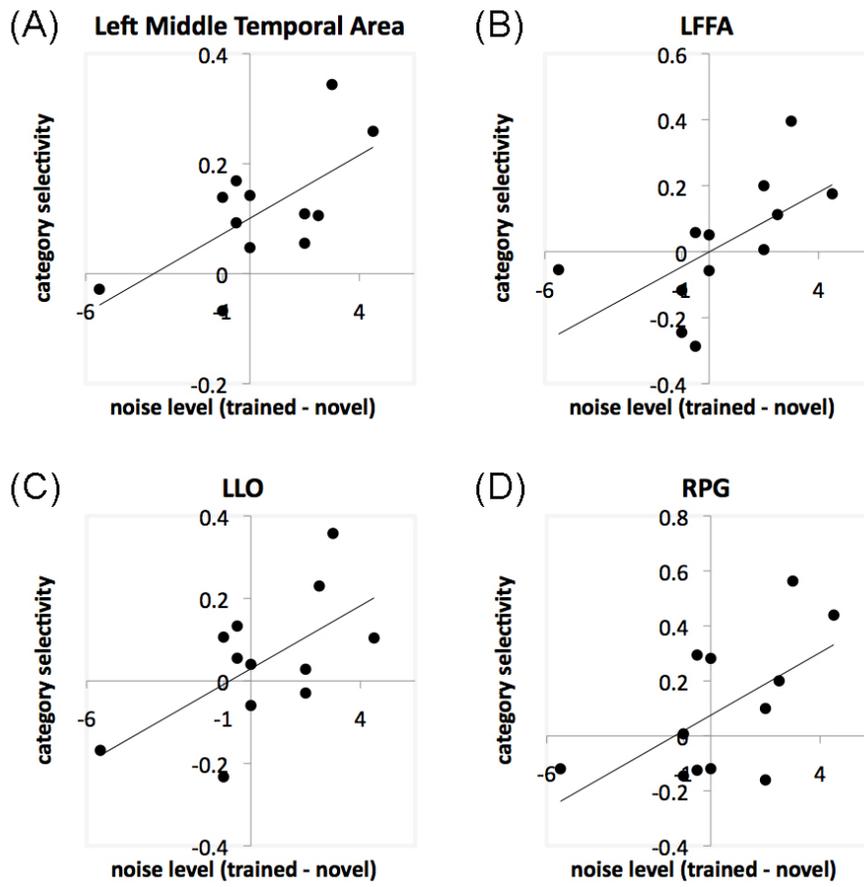


Figure 6

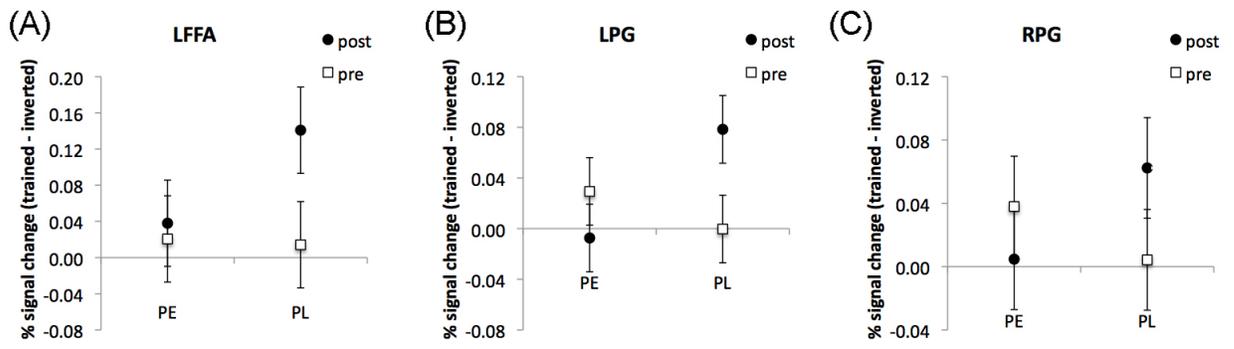


Figure 7

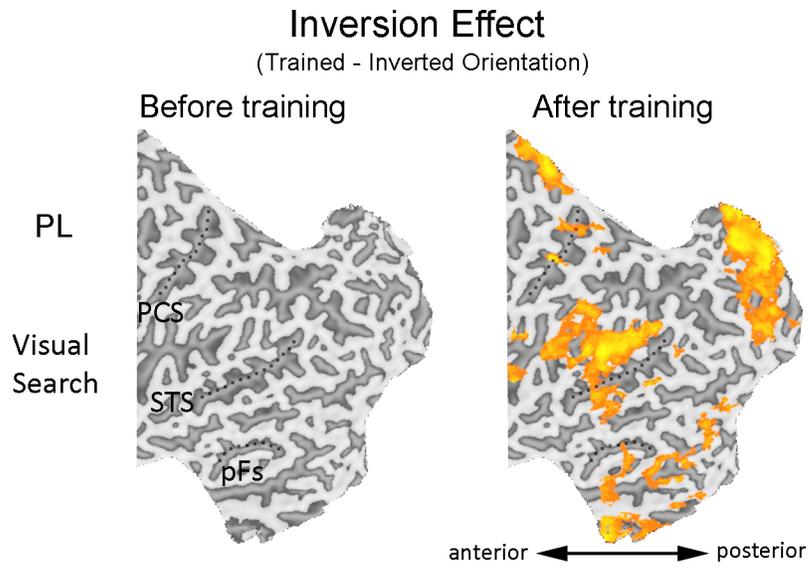


Figure 8

Orientation x Category x Training

(Trained - Inverted Orientation) x
(Trained - Novel Category) x (Post - Pre scan)

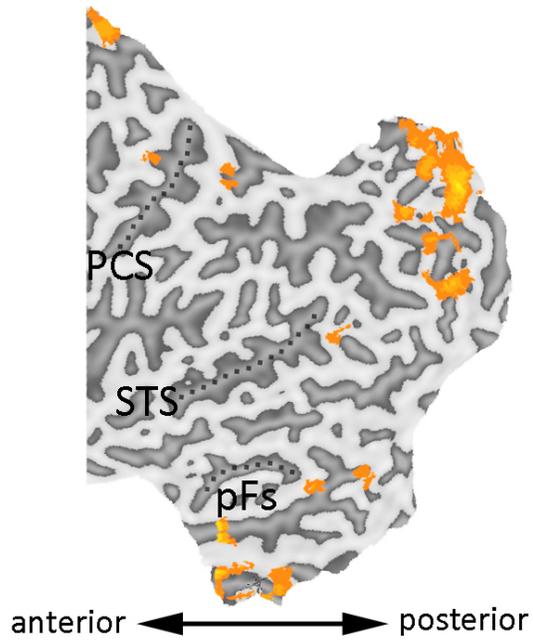


Figure 9

	Orientation	pretest	posttest		
		trained exemplar	trained exemplar	novel exemplar	novel category
PL	trained	1.02	3.38	3.56	1.01
	inverted	0.82	1.33	1.41	1.43
PE	trained	0.48	1.09	1.28	0.91
	inverted	0.58	0.91	0.76	1.03

Table 1. Discriminability (d') for each condition for PL and PE during the visual search test before and after training. For PL, discriminability specifically improved for the trained orientation and for the trained object category after training, but transferred completely to novel exemplars within the trained category. For PE, visual search performance also improved after the shape individuation training.

	pretest	posttest	
	trained category	trained category	novel category
PL	0.917	2.67	1.33
PE	0.917	2.17	1.75

Table 2. Performance (in terms of the estimated level of Gaussian noise required for about 80% accuracy) for PL and PE during the shape matching test before and after training (for upright objects only). For both groups, shape matching performance improved for the trained category but not for the novel category after training.

area	side	x	y	z	mm ³	max. t
Occipitotemporal areas						
V1/V2	R	16	-93	13	225	4.57
V1/V2	R	20	-85	2	246	4.48
lingual gyrus	L	-19	-85	-12	240	3.77
lingual gyrus	C	2	-72	-8	398	4.04
posterior fusiform gyrus	L	-23	-66	-12	648	3.77
	R	19	-68	-13	486	4.12
lingual gyrus	L	-18	-59	10	650	5.82
	R	17	-51	2	959	5.43
fusiform gyrus	L	-41	-45	-14	378	3.5
Parietal areas						
intraparietal sulcus	L	-45	-71	35	563	5.09
superior parietal lobe	L	-24	-62	44	729	-5.01
	R	13	-64	42	837	-4.72
inferior parietal lobe	L	-53	-39	32	756	5.71
superior parietal lobe	R	24	-38	56	327	5.09
inferior parietal lobe	L	-34	-35	46	216	3.76
Superior temporal gyrus						
STS	L	-56	-49	10	815	7.11
STS	R	54	-36	3	742	7.02
Sylvian fissure						
sylvian fissure	R	62	-21	11	640	5.84
Cuneus						
cuneus	C	1	-80	27	990	5.92
Cingulate gyrus						
cingulate gyrus	C	5	-49	7	809	7.35
cingulate gyrus	L	-14	-42	42	613	8.42

Table 3. Areas showing significant Orientation x Training effects for PL during visual search.

area	side	x	y	z	mm ³	max. t
Occipitotemporal areas						
inferior temporal gyrus	L	-43	-60	-21	243	4.71
inferior temporal sulcus	R	45	-53	-5	270	3.37
middle temporal gyrus	L	-63	-37	-6	288	4.02
Parietal areas						
intraparietal sulcus	L	-36	-61	48	207	3.68
Precuneus						
precuneus	R	10	-64	36	317	4.32
Cingulate gyrus						
cingulate gyrus	L	-9	-28	31	504	4.33

Table 4. Areas showing significant Category x Training effects for PE during shape matching.

area	side	x	y	z	mm ³	max. t
Occipitotemporal areas						
inferior temporal sulcus	L	-39	-70	-2	225	-3.28
middle temporal gyrus	L	-38	-58	4	666	-4.64
middle temporal gyrus	L	-49	-57	3	279	-6
fusiform gyrus	R	33	-55	-14	189	-3.32
Precuneus						
precuneus	C	6	-67	46	270	-4.18
Parietal areas						
intraparietal sulcus	L	-28	-45	42	390	-5.96

Table 5. Areas showing significant Orientation x Training effects for PE during visual search.

area	side	x	y	z	mm3	max.t
Occipitotemporal areas						
V1	R	14	-86	7	461	13.57
V1		0	-70	11	636	13
V1/V2	L	-6	-81	14	726	14.1
V1/V2	R	9	-65	7	760	17.1
lingual gyrus	L	-20	-83	-12	285	11.1
	R	18	-64	-15	558	13.08
lingual gyrus	L	-14	-53	4	626	17.7
	R	14	-47	1	812	20.1
fusiform gyrus	L	-25	-65	-16	366	12.28
	R	33	-55	-17	459	16.4
fusiform gyrus	L	-37	-53	-7	713	19.3
fusiform gyrus	L	-29	-53	-14	423	10.99
parahippocampal gyrus	R	18	-38	1	543	30.9
Parietal areas						
superior parietal lobe	L	-24	-63	45	378	12.93
	R	20	-43	54	339	13.7
inferior parietal sulcus	L	-35	-39	47	558	21.3
Superior temporal gyrus						
superior temporal sulcus	L	-54	-52	7	656	21.6
	R	53	-47	9	920	16.86
superior temporal sulcus	R	48	-11	-14	446	26.3
Sylvian fissure						
sylvian fissure	L	-56	-35	21	732	23.44
	L	-39	-31	21	627	20.61
	R	59	-44	24	432	18.93
sylvian fissure	R	62	-22	12	465	13.88
sylvian fissure	L	-38	-2	2	777	15.07
	R	58	-3	7	629	19.2
Cuneus						
cuneus	L	-13	-86	17	216	13.7
	R	15	-85	22	615	21.5
Cingulate gyrus						
cingulate gyrus	L	-19	-55	24	493	14.6
cingulate gyrus	L	-23	-46	29	732	28.4
cingulate gyrus	R	3	-42	20	821	18.1
cingulate gyrus	L	-9	-30	28	665	20
cingulate gyrus	L	-2	-22	37	715	17.5

Table 6. Areas showing significant Group x Orientation x Training effects during visual

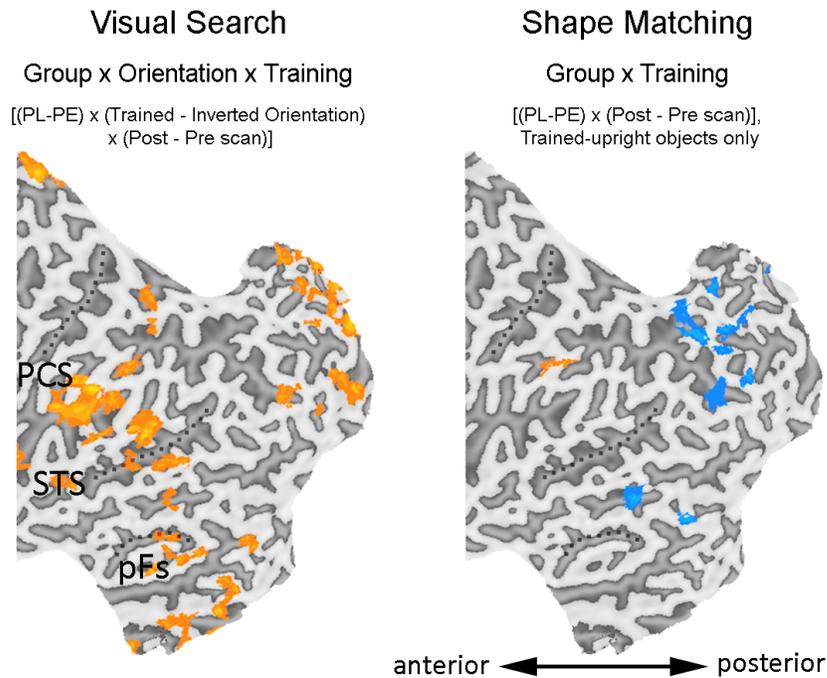
search. All areas showed a stronger Orientation x Training effect in the PL group compared with the PE group.

area	side	x	y	z	mm ³	max. t
Parietal areas						
intraparietal sulcus	R	43	-60	43	285	4.33
supramarginal gyrus	L	-43	-59	27	636	4.96
inferior parietal area	L	-53	-48	44	264	3.61
Precuneus						
precuneus	L	-9	-27	50	513	7.71
Superior temporal gyrus						
STS	R	36	-23	7	261	3.66

Table 7. Areas showing significant Category x Training effects for PL during shape matching.

Larger training effect for	area	side	x	y	z	mm3	max.t
PE	Occipitotemporal areas						
	middle occipital gyrus	L	-34	-83	14	519	-3.44
	inferior occipital sulcus	L	-27	-77	2	498	-3.9
	fusiform gyrus	L	-44	-58	-21	351	-3.72
	fusiform gyrus	L	-40	-57	-3	644	-4.63
	inferior temporal sulcus	R	52	-48	-8	216	-3.3
	Parietal areas						
	occipito-parietal sulcus	L	-13	-77	32	509	-3.9
	superior parietal lobe	L	-19	-68	44	788	-3.81
	inferior parietal sulcus	R	28	-62	48	279	-3.23
	Precuneus						
	precuneus	R	13	-71	39	348	-3.66
	precuneus		0	-58	31	374	-3.47
	precuneus	R	2	-53	47	480	-3.64
	Cingulate gyrus						
	cingulate gyrus	L	-2	-29	24	297	-3.48
	PL	Occipitotemporal areas					
inferior temporal sulcus		R	49	-12	-10	297	2.972
Parietal areas							
inferior parietal sulcus		R	46	-57	38	435	3.38
inferior parietal lobe		L	-56	-34	34	445	3.797
Sylvian fissure							
sylvian fissure		R	47	-45	27	324	3.02
sylvian fissure	R	49	-8	16	438	3.486	

Table 8. Areas showing significant Group x Training effects (for trained-upright objects) during shape matching. The response in these brain regions showed a stronger training effect for one group as indicated.



Supplementary Figure 1. Different training effects for the two groups with different training experiences. The left panel shows areas with significantly larger inversion effect for the PL group than the PE group (orange regions) during visual search. The right panel shows areas with significantly larger training effect for PL than for PE (orange) or larger training effect for PE than PL (blue) during shape matching with trained-upright objects. PCS – post-central sulcus; STS – superior temporal sulcus; pFs – posterior fusiform gyrus.