Learning optimal eye movements to unusual faces

Matthew F. Peterson and **Miguel P. Eckstein**

University of California, Santa Barbara, Department of Psychological & Brain Sciences

Abstract

Eye movements, which guide the fovea's high resolution and computational power to relevant areas of the visual scene, are integral to efficient, successful completion of many visual tasks. How humans modify their eye movements through experience with their perceptual environments, and its functional role in learning new tasks, has not been fully investigated. Here, we used a face identification task where only the mouth discriminated exemplars to assess if, how, and when eye movement modulation may mediate learning. By interleaving trials of unconstrained eye movements with trials of forced fixation, we attempted to separate the contributions of eye movements and covert mechanisms to performance improvements. Without instruction, a majority of observers substantially increased accuracy and learned to direct their initial eye movements towards the optimal fixation point. The proximity of an observer's default face identification eye movement behavior to the new optimal fixation point and the observer's peripheral processing ability were predictive of performance gains and eye movement learning. After practice in a subsequent condition in which observers were directed to fixate different locations along the face, including the relevant mouth region, all observers learned to make eye movements to the optimal fixation point. In this fully learned state, augmented fixation strategy accounted for 43% of total efficiency improvements while covert mechanisms accounted for the remaining 57%. The findings suggest a critical role for eye movement planning to perceptual learning, and elucidate factors that can predict when and how well an observer can learn a new task with unusual exemplars.

Keywords

perceptual learning; eye movements; face recognition; ideal observer

1. Introduction

Perceptual learning, whereby training leads to significant and sustained improvement in perceptual tasks, has been studied at the behavioral and neural level for many years (Fine & Jacobs, 2002; Gilbert, Sigman, & Crist, 2001; Goldstone, 1998). The brain's ability to improve perceptual performance across a wide range of modalities (visual: Ahissar &

Hochstein, 1993; Fiorentini & Berardi, 1980; Matthews, Liu, Geesaman, & Qian, 1999; auditory: Atienza, Cantero, & Dominguez-Marin, 2002; Polley, Steinberg, & Merzenich, 2006; olfactory: Moreno et al., 2009; Wilson & Stevenson, 2003; somatosensory: Pleger et al., 2003; Sathian & Zangaladze, 1998) and tasks (motion discrimination: Ball & Sekuler, 1982, 1987; texture segregation: Karni & Sagi, 1991, 1993; auditory frequency discrimination: Hawkey, Amitay, & Moore, 2004; wine discrimination: Bende & Nordin, 1997) suggests that learning is mediated by a complex and, at some level, generalized set of neural mechanisms and corresponding behaviors. Focusing on visual learning, past research has implicated modulations at the neural and algorithmic levels, such as internal noise reduction (Dosher & Lu, 1998; Lu & Dosher, 1998), signal amplification (Gold, Bennett, & Sekuler, 1999; Lu & Dosher, 1999), feature/receptive field tuning (Li, Levi, & Klein, 2004; Saarinen & Levi, 1995), and attentional reallocation (Ahissar & Hochstein, 1993; Gilbert et al., 2001; Ito, Westheimer, & Gilbert, 1998; Peterson, Abbey, & Eckstein, 2009; Trenti, Barraza, & Eckstein, 2010). Common to many of these mechanisms is the fundamental concept of improved efficiency at selecting, processing, and integrating task-relevant information or features (Beard & Ahumada, 1999; Dosher & Lu, 1998; Eckstein, Abbey, Pham, & Shimozaki, 2004; Gold et al., 1999; Hurlbert, 2000; Peterson & Eckstein, 2012).

Although these studies have increased our understanding of the mechanisms mediating perceptual learning, most investigations have not considered the role active vision, and specifically eye movements, plays in perceptual learning (but see Chukoskie, Snider, Mozer, Krauzlis, & Sejnowski, 2013; Droll, Abbey, & Eckstein, 2009; Holm, Engel, & Schrater, 2012 for exceptions). This would seem to be an important factor to explore, as the inhomogeneity in visual of processing across the visual field suggests that during active vision, familiarization with a perceptual environment might lead to changes in saccade strategies and contribute to performance improvements. While the physical world surrounds us across all angles, the visual system is limited to a slightly greater than 180 degree field of view at any given time, with only a tiny portion of this area surrounding fixation (corresponding to the fovea) given access to high-resolution, high-sensitivity processing. This architecture creates a need for the brain to intelligently guide the eyes through head, body, and eye movements such that task-relevant information, in the form of light, impinges areas of the retina that correspond to high-powered processing by visual cortex. The critical role that eye movement behavior plays in perception can be seen in such common but important tasks as visual search (Eckstein, Beutter, Pham, Shimozaki, & Stone, 2007; Hayhoe & Ballard, 2005; Najemnik & Geisler, 2005, 2009; Rao, Zelinsky, Hayhoe, & Ballard, 2002; Tavassoli, Linde, Bovik, & Cormack, 2009; Zelinsky, Rao, Hayhoe, & Ballard, 1997), reading (Rayner, 1998), and face recognition (Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012). Indeed, humans display a remarkable ability to enact eye movement strategies that are consistent with optimal fixation model predictions (Najemnik & Geisler, 2005, 2008; Peterson & Eckstein, 2012). Given the vital nature of this interaction, surprisingly little work has assessed the functional role of eye movement strategy modulation to perceptual learning beyond that conferred by modification to covert mechanisms, and how the brain learns these strategies.

Here, we assess how practice changes observers' eye movement strategies and evaluate their functional role in performance improvements. We chose a task, face identification, for

which humans have already learned optimized eye movement strategies to typical, naturally occurring faces (Peterson & Eckstein, 2012). We constructed face images where all discriminatory information was confined to a small region encompassing the mouth, creating a situation where the optimal eye movement strategy for this synthetic face set diverged greatly from the optimal strategy for normal faces. Without any special instructions, observers were asked to identify these faces over the course of 1600 trials. We measured changes in fixation patterns and isolated the contribution to accuracy improvements due to eye movement modulations by interleaving trials where eye movements were allowed with trials where fixation was confined to a specific region. We found that observers fell into three distinct groups defined by their eye movement modulation: Non Movers, Partial Movers, and Complete Movers. Adapting fixation strategy was found to significantly increase performance beyond that possible with only modulations to covert mechanisms. The magnitude of overall improvement, and the ability of observers to modify their eye movements without instruction, was seen to be influenced mainly by two factors: 1) The distance of the observer's initial, normal fixation region from the new optimal location, and 2) The observer's peripheral processing ability. We conclude that eye movements can be an essential element in maximizing learning of new perceptual tasks, and that the ability to learn these new strategies can be predicted by the observer's ability to notice and process task-relevant information across the visual field.

2. General methods

2.1. Participants

Fourteen undergraduate students (eight female, six male, age range 20 to 23) from the University of California, Santa Barbara participated in the study for course credit. All observers had normal or corrected-to-normal vision and no history of neurological disorders. Each observer completed all four tasks.

2.2. Display

All stimuli were presented using a linearly calibrated 17-inch CRT monitor set to 8-bit grayscale with mean luminance of 25 cd/m^2, resolution of 600 by 800 pixels, and refresh rate of 100 Hz. Observers sat 63 cm from the monitor, with each pixel subtending .037 degrees visual angle.

2.3. Eye tracking

The left eye of each participant was tracked using an SR Research Eyelink 1000 Tower Mount eye tracker sampling at 250 Hz. A nine-point calibration and validation were run before each 100 trial session with a mean error of no more than 0.5° visual angle. Using Eyelink's suggested criteria, saccades were classified as events where eye velocity was greater than 22°/sec and eye acceleration exceeded 4000°/sec^2. Periods of forced fixation were enforced by aborting the current trial if the eye position registered more than 1° from the center of the fixation cross.

2.4. Procedure

The entire study consisted of four distinct sections, each of which is described in detail in sections 3 through 6. In general, grayscale face images were randomly selected from a small set of possible images and briefly shown to observers with additive Gaussian white noise. Observers were then presented with high contrast, noise-free versions of the possible face images and used the mouse to click on the face they thought they had seen.

2.5. Ideal observer, efficiency, and the learning factor

Performance is dictated by an interaction between the visual information available for a task and the visual system's ability to extract and process this information. The amount of task-relevant information can be assessed using ideal observer theory, a technique that specifies an algorithm that makes Bayesian optimal decisions given the statistical properties of the possible signals (here, face images) and the added stochastic noise (Green, 1966). The ideal observer provides a gold standard for maximum task performance. Human behavior and its associated performance are thus conceptualized as the result of some noisy process (the visual system) that incorporates only a portion of the available information into its decisions. This proportion is quantified with the absolute efficiency metric, η, which is defined as the ratio of the ideal observer's signal contrast energy, $E_{IO|human}$, to that of the human's, E_{human}, for a given performance threshold (Barlow, 1980; Burgess, Wagner, Jennings, & Barlow, 1981; Eckstein et al., 2004). Here, the signal is the original face image, which is common to both the ideal observer and the humans and whose contrast energy is designated by E_0. The signal was then modified using a contrast multiplier (a scalar value between 0 and 1 that attenuates signal strength and thus decreases stimulus information), denoted as C_{human} for the psychophysical trials (kept constant across trials and observers) and $C_{IO|human}$ for the ideal observer, where the ideal observer's multiplier value was chosen so as to match the human's perceptual accuracy. Thus, the total contrast energy is the original signal's contrast energy multiplied by the square of the contrast multiplier. Using these properties, the absolute efficiency is computed as:

$$\eta = \frac{E_{IO|human}}{E_{human}} = \left(\frac{C_{IO|human}^2 E_0}{C_{human}^2 E_0} \right) = \left(\frac{C_{IO|human}}{C_{human}} \right)^2 \quad (2.5.1)$$

Efficiency is a monotonic transform of human performance: improvement in an observer's proportion correct, the most classic behavioral trademark of learning, directly translates to increased efficiency. Efficiency formalizes this learning in terms of the increase in the amount of task-relevant information the observer is able to incorporate into the perceptual decision.

In this study, we measured learning in two conditions: when eye movements were allowed (*free*) and when fixation was constrained to a specific location (*fixed*), with these trial types run in an interleaved fashion. For any given condition, c, and time frame of interest, t, the total amount of learning, $\eta_{c,t}$, is given by the difference in efficiency at t, $\eta_{c,t}$, compared to the efficiency at the beginning of the study, $\eta_{c,0}$:

$$\Delta\eta_{c,t}=\eta_{c,t}-\eta_{c,0} \quad (2.5.2)$$

A difference in efficiency when eye movement behavior is allowed to change is potentially a consequence of modifications to both overt (eye movement) and covert mechanisms. Here, covert mechanism is used as a general term that encompasses any learning-associated changes that are not eye movements. This may include things like changes to covert attention, feature selection, internal noise reduction, and stimulus enhancement, among many others. The key here is that these covert mechanisms are assumed common to both the *free* and *fixed* conditions. Thus, we take the change in efficiency for the *free* condition as a measure of the total amount of learning dependent on changes to both covert mechanisms *and* eye movements (Fig. 1):

$$\Delta\eta_{total}=\Delta\eta_{free}=\eta_{free,end}-\eta_{free,0} \quad (2.5.3)$$

We also note that the change in efficiency in the *fixed* trials, η_{fixed}, is solely dependent on modifications to covert mechanisms (Fig. 1):

$$\Delta\eta_{covert\_mechanisms}=\Delta\eta_{fixed}=\eta_{fixed,end}-\eta_{fixed,0} \quad (2.5.4)$$

Assuming both trial types begin at the same efficiency (which is the case here), we can then isolate the contribution of eye movements to learning by noting that the total change in efficiency is the sum of the change due to covert mechanisms, which is common to both *free* and *fixed* trials, and the change attributable to eye movements alone, $\eta_{eye\_movements}$ (Fig. 1):

$$\Delta\eta_{eye\_movements}=\eta_{free,end}-\eta_{fixed,end} \quad (2.5.5)$$

$$\Delta\eta_{total}=\Delta\eta_{covert\_mechanisms}+\Delta\eta_{eye\_movements} \quad (2.5.6)$$

Dividing the additional efficiency conferred by eye movement modulation by the total efficiency increase gives a metric termed the learning factor, *LF*, that quantifies the percent of total learning attributable to eye movements:

$$LF=\frac{\Delta\eta_{eye\_movements}}{\Delta\eta_{total}} \times 100\% \quad (2.5.7)$$

Finally, we define the relative efficiency between two conditions, $\eta_{rel,cond\ 1-cond\ 2}$, as the ratio of each condition's absolute efficiency (Eckstein, Beutter, & Stone, 2001). This gives a measure of the difference in the amount of information an observer is using between two conditions.

$$\eta_{rel,cond1-cond2} = \frac{\eta_{cond1}}{\eta_{cond2}} \quad (2.5.8)$$

3. Task 1: Preferred points of fixation for face recognition

Humans have a reliable, consistent, and individualized initial eye movement strategy when identifying faces. Most people select an initial fixation location toward the vertical meridian of the face and displaced downward from the eyes, about a third of the way toward the nose tip (Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012, 2013). We assessed each observer's personal preferred point of fixation using a simple, fast face identification task with normal faces (Peterson & Eckstein, 2012).

3.1. Stimuli

Images were 600 by 600 pixels, frontal view photographs of ten different male faces with neutral expressions. Photos were taken in-house under constant, diffuse lighting conditions. Hair, clothing, and background were excluded using a black cropping mask, which created a visible face area of 18.1° with a 7.2° separation between the center of the eyes and center of the mouth in the vertical direction. The visible face area was lowered in contrast (average RMS contrast = 13.8%), energy normalized, and embedded in a zero-mean, white Gaussian noise field with a standard deviation of 1.96 cd/m^2 (corresponding to a noise RMS contrast of 0.078 and a noise spectral density of 8.4e-6 deg^2) that was independently sampled on each stimulus presentation (Fig. 2A).

3.2. Procedure

Each of 400 trials began with a fixation cross located 13.3° from the center of the monitor at either the extreme left or extreme right edge of the screen (location randomly selected). The observer initiated the trial by fixating the cross and pressing the spacebar. After a random, uniformly distributed delay between 500 and 1500 ms, the cross was removed and a noisy stimulus image was presented at the center of the monitor. Before the stimulus image appeared, fixation at the peripheral cross was enforced such that if gaze deviated by more than 1° from the center of the cross the trial was aborted and restarted. Once the face appeared, observers were free to move their eyes, with the 350 ms display time allowing for a single in-face fixation. The face image was then replaced for 500 ms by a high contrast white noise mask (standard deviation of 5.88 cd/m^2 with a noise RMS contrast of 23.5% and spectral density of 7.6e-5 deg^2) followed by a response screen. Observers used the mouse to click on the face they believe they had just seen. A white box then framed the correct answer (Fig. 2A).

3.3. Results

Consistent with previous reports using short display times (Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012), observers first fixated an area toward the midline of the face and displaced downward to just below the eyes (mean distance below center of the eyes was 1.15°; Fig. 2B,C). Crucially, there was significant variation between observers with a standard deviation across observers of 1.12° (95% confidence interval = [0.81°,1.80°]), also consistent with previous reports (Peterson & Eckstein, 2013; Fig. 2B,C). These differences

in fixation behavior were used to assess the effect of starting location on the ability to notice and learn the critical informative regions of a novel, composite face stimulus described in Section 4.

4. Task 2: Learning unusual faces

The main task in this study required observers to identify faces that were artificially manipulated to constrain all discriminating information to the mouth region. Task 2 interleaved trials where eye movements were allowed with trials were fixation was restricted to the observer's preferred face identification fixation location, as assessed in Task 1. This allowed us to measure both perceptual decision learning (performance improvements) and the corresponding changes in eye movement behavior.

4.1. Stimuli

We created composite stimuli using five frontal view, 600 by 600 pixel neutral expression male faces (different identities from Task 1). One of the faces was selected as the "base face". For each of the four "donor faces", we used Photoshop to extract the mouth region and then blended the mouth into the base face's corresponding region, creating four composite images where all discriminating information was confined to the mouth (Fig. 3A). Images were lowered in contrast (average RMS contrast = 14.2%) and embedded in zero-mean, white Gaussian noise with a standard deviation of 0.98 cd/m^2 (noise RMS contrast of 3.9% and spectral density of 2.1e-6 deg^2) that was independently sampled on each stimulus presentation (Fig. 3D). To confirm that the discriminatory information was confined to the mouth region we implemented a region of interest (ROI) ideal observer analysis. The ROI localizes and quantifies information content by making identification decisions on all the small (1° by 1°) regions that can be extracted from the noisy face images (Fig. 3B; see Peterson & Eckstein, 2012 for a full derivation). An ideal observer with a simulated foveated visual system (foveated ideal observer; FIO) was used to predict performance at all possible fixation points, with the results also confirming that the theoretically optimal point of fixation is on the mouth (Fig. 3C; see Peterson & Eckstein, 2012 for a full derivation).

4.2. Procedure

We attempted to separate the contributions of overt (eye movements) and covert mechanisms to perceptual learning performance enhancements by interleaving two distinct trial types over the course of sixteen, 100-trial sessions. On odd-numbered trials, observers were free to explore the stimulus as they chose. On even-numbered trials, eye movements were precluded by aborting the trial if the observer's gaze was detected straying from a pre-assigned location. Observers were given no instructions as to the special nature of the stimuli, only that the task would be difficult but that the faces were, indeed, discriminable.

4.2.1. Odd trials: Free eye movements—Free eye movement trials (*free*) began with the observer fixating a cross at the left or right edge of the screen (indicated by the red circle in Fig. 3D) as in Task 1 and pressing the spacebar when ready. Following a random 500 – 1500 ms delay during which eye movements were not allowed, the cross was removed and

one of the four faces was displayed in the center of the screen for 600 ms, during which eye movements were allowed. The image was then replaced by a gray screen with a fixation cross in the center of the monitor. Once observers moved their eyes to the cross, they had to maintain fixation for a random 500 – 1500 ms delay. The cross was then removed and high contrast, noise-free versions of the four face images appeared in the four corners of the screen. Observers were allowed to move their eyes freely for 500 ms before the images were replaced by boxes with identifying names (Al, Bill, Carl, and Dave). Observers used the mouse to select the box with the name that corresponded to the face image they believed they had seen. The correct answer was then outlined in white before proceeding to the next trial (Fig. 3D). The re-fixation and brief response-image presentation time were employed to prevent observers from using a trivial side-by-side visual comparison method to learn the stimulus properties.

4.2.2. Even trials: Forced fixation—Forced fixation trials (*fixed*) began with a cross located at the observer's unique preferred point of fixation for face identification (indicated by the blue circle in Fig. 3D). This location was determined by computing the mean position of the end of the observer's first into-face saccade from Task 1 (e.g., black dot in Fig. 2B). After pressing the spacebar, a 500 – 1500 ms delay ensued followed by the presentation of a noisy face image for 250 ms (Fig. 3D). Critically, if the observer's eye position strayed more than 1° from the center of the cross the trial would restart with a newly sampled noisy face image. After stimulus presentation, the procedure (refixate, response image, response names, feedback) was identical to Section 4.2.1.

4.3.1 Results: Eye movements—Observers fell into three distinct groups based on their eye movement behavior during the *free* trials in the final session. Using the final (16th) session as our test data, we evaluated each observer on two metrics. First, we tested whether the observer's fixations had migrated more than 0.5 degrees visual angle from their preferred fixation location. Three of the fourteen participants did not reach this migration criterion, placing them in a group termed the Non Movers (NM; ps > .1 for each observer; representative observer's fixation behavior shown in the left panel of Fig. 4A, group data represented by the black line in Fig. 4B). Next, we tested whether the observer's fixations had migrated to within the informative region of the stimulus, an area extending one degree above the mouth center. Five observers finished fixating within this region, comprising the group termed the Complete Movers (CM; all ps > .1; see right panel of Fig. 4A and light gray line in Fig. 4B). The remaining six observers, termed the Partial Movers (PM), significantly moved their fixations downward by at least 0.5 degrees from their preferred point (all ps < .01), but still fixated significantly above the mouth region (all ps < .01; see middle panel of Fig. 4A and dark gray line in Fig. 4B; see Supplementary Figure 1 for the last session's group data).

4.3.2 Results: Perceptual performance—As a group, performance, measured in terms of proportion correct (*PC*), increased significantly in both the Free and Fixed conditions, with an average difference between the final two and first two sessions ($PC_{last\text{-}first} = PC_{last} - PC_{first}$) of 0.218 in the *free* condition ($t = 6.42$, $p < .001$, one-tailed, black line in Fig. 5, left panel) and 0.180 in the *fixed* condition ($t = 5.03$, $p < .001$, one-tailed; gray line in Fig. 5,

left panel). Importantly, performance in the first two sessions was not significantly different between the two conditions ($PC_{free-fixed} = PC_{free} - PC_{fixed} = 7.1e\text{-}4$, $t = 0.04$, $p = .97$, two-tailed), but trended toward significance in the last two sessions ($PC_{free-fixed} = 0.039$, $t = 1.83$, $p = .09$, two-tailed). However, the performance changes were starkly different across the groups defined in Section 4.3.1. NMs did not significantly improve in either condition (*free:* $PC_{last-first} = 0.077$, $t = 1.56$, $p = .13$, one-tailed; *fixed:* $PC_{last-first} = 0.043$, $t = 1.14$, $p = .15$, one-tailed; Fig. 5). PMs significantly improved in both conditions (*free:* $PC_{last-first} = 0.253$, $t = 22.71$, $p < .001$, one-tailed; *fixed:* $PC_{last-first} = 0.247$, $t = 5.76$, $p = .001$, one-tailed; Fig. 5) but saw no differentiation between the conditions during the last two sessions ($PC_{free-fixed} = -0.010$, $t = 0.28$, $p = .79$, two-tailed; Fig. 5). Finally, CMs significantly improved in both conditions (*free:* $PC_{last-first} = 0.26$, $t = 3.43$, $p = .01$, one-tailed; *fixed:* $PC_{last-first} = 0.18$, $t = 2.63$, $p = .03$, one-tailed; Fig. 5), with significantly greater performance in the *free* condition ($PC_{free-fixed} = 0.08$, $t = 4.16$, $p = .01$, two-tailed; Fig. 5).

5. Task 3: Guided exploration

Only five of the fourteen observers completely modulated their eye movement behavior, while three observers failed to learn the task at all. What drives these differences in eye movement and general task-learning behavior? We hypothesized that the interaction of two main factors leads to the observed differences among individuals. First, individuals display distinct eye movement patterns during normal face identification, with some looking further up the face (and thus further from the informative mouth region) than others (Peterson & Eckstein, 2013). Second, there may be substantial individual variability in the ability to process the mouth region's visual information content as a function of peripheral distance. Both of these factors could lead to situations where some observers were more likely than others to notice differences in the mouth region, which would lead to variability in learning among participants, as well as possible differential benefits to eye movement modulation dependent on peripheral processing ability. To assess these factors, observers identified the same modified faces while forced to fixate at five different locations along the vertical meridian of the face.

5.1. Stimuli

Images were the same composite pictures as described in Section 4.1.

5.2. Procedure

Each of 1500 trials (15 sessions of 100 trials each) began with a fixation cross located at one of five positions. Each location was selected 20 times per session in pseudorandom order. Three of the positions, which were common across all observers, corresponded to the eyes, nose tip, and mouth center (aligned along the face's vertical midline with a spacing of 3.3° visual angle; white dots in Fig. 6A). The fourth position was determined by the observer's preferred fixation point from Task 1 (the same location as in the *fixed* condition from Task 2; red dot in Fig. 6A). The fifth position corresponded to the observer's mean fixation location from the final two sessions of the *free* condition in Task 2 (green dot in Fig. 6A). The procedure was identical to the *fixed* condition of Task 2 (Section 4.2.2).

5.3. Results

NMs showed no signs of learning over the sixteen learning sessions and thus entered the guided exploration section with effectively no knowledge of the relevant stimulus properties. This lack of learning was almost immediately rectified through guided exploration of the stimulus, with a significantly above chance average PC when fixating the mouth of 0.70 ($t = 3.40$, $p = .038$, one-tailed).

A second notable effect is the clear advantage, for all observers, of fixating closer to the mouth region. Consistent with the foveated properties of the visual system, where visual information in the periphery is conferred degraded processing, and as predicted by a foveated ideal observer (Fig. 3C; Peterson & Eckstein, 2012), group performance decreased monotonically with fixation distance from the mouth ($PC_{mouth} = .704\pm.037$, $PC_{nose} = .605\pm.049$, $PC_{eyes} = .529\pm.044$; $t_{mouth > nose} = 4.69$, $p < .01$; $t_{nose > eyes} = 3.87$, $p < .01$; Bonferonni corrected; Fig. 6B).

6. Task 4: Learning completion

Each observer performed well above chance when fixating the mouth region in Task 3. With the statistical structure of the stimuli now known to everybody, we assessed the effects of guided exploration by asking observers to identify the composite faces in the same manner as in Task 2 (interspersed *free* and *fixed* trials). Would the learned knowledge of the special structure of the composite faces translate to eye movement and performance changes in the NM group? How might the PMs further alter their eye movements with more complete knowledge of their peripheral processing ability?

6.1. Stimuli and procedure

Images and experimental design were the same as described in Section 4.1. Observers completed five more sessions of Task 2 (100 trials per session).

6.2. Results

The statistical knowledge of the composite stimuli produced an immediate effect on the NM eye movements, with average fixation distance from the mouth decreasing from 6.58° just before exploration to 1.38° just after ($t = 8.52$, $p < .01$, one-tailed; Fig. 7A). The change in fixation behavior was accompanied by a robust increase in perceptual performance in both the *free* ($PC_{post-pre,free} = .313$, $t = 4.18$, $p = .03$, one-tailed) and *fixed* ($PC_{post-pre,fixed} = .205$, $t = 2.78$, $p = .05$, one-tailed) conditions (Fig. 7B). Meanwhile, PMs, who had stabilized their fixations in the session before exploration to a region above the mouth, further adjusted their eye movements after experience with fixating further down the face, with the average distance above mouth decreasing from 3.42° pre-exploration to 2.06° after ($t = 2.66$, $p = .02$, one-tailed; Fig 7A). This additional movement was associated with an additional boost to perceptual performance in the *free* condition ($PC_{post-pre,free} = .098$, $t = 2.05$, $p = .048$, one-tailed), and a marginal boost in the *fixed* condition ($PC_{post-pre,fixed} = .053$, $t = 1.98$, $p = .052$, one-tailed; Fig. 7B). CMs showed a trend toward moving even closer to the mouth center, with distance from the center decreasing from 0.35° pre-exploration to 0.01° after ($t = 1.91$, $p = .064$, one-tailed; Fig 7A). This adjustment was also associated with an increase

in performance in the *free* condition ($PC_{post-pre,free}$ = .146, t = 3.96, p < .01, one-tailed) but only a marginal trend in the *fixed* ($PC_{post-pre,fixed}$ = .069, t = 1.70, p = .082, one-tailed; Fig. 7B). Across all observers, performance in both the *free* and *fixed* conditions increased significantly ($PC_{post-pre,free}$ = .161, t = 4.60, p < .01, one-tailed; $PC_{post-pre,fixed}$ = .091, t = 3.32, p < .01, one-tailed).

7. Discussion

7.1. The importance of eye movement modulation for perceptual learning

Perceptual learning has been shown to change which stimulus features and spatial locations the visual system selects for task-specific processing (Ahissar & Hochstein, 1993; Dosher & Lu, 1998; Droll et al., 2009; Eckstein et al., 2004; Peterson et al., 2009). In this work, we showed that humans are able to adjust their eye movements to optimize the sampling of a novel visual stimulus, with these modulations contributing significantly to task performance maximization beyond the contributions of covert mechanisms alone. But what is the magnitude of this learning in terms of the extra amount of information the observer is able to incorporate into the perceptual decision?

To quantify and compare these improvements in a meaningful way, we would like to convert the raw performance data, given in terms of proportion correct, to a measure of information utilization efficiency. For this, we use the absolute efficiency metric, η, and the learning factor, LF, as described in Section 2.5. Before guided exploration, only CMs realized benefits from adapting eye movement plans. After exploration, the NMs also displayed a great advantage when allowed to choose their own eye movements, rivaling that of the CMs ($LF_{post, NMs}$ = 60%±10%, $LF_{post, CMs}$ = 63%±18%; Fig. 8A). Across all observers, eye movement modulations accounted for 43%±11% of the total improvements in efficiency post-exploration (Fig. 8A).

A corroborating result can be seen in the guided exploration data by ascertaining the amount of information an observer uses when forced to fixate their preferred point relative to when forced to fixate the mouth (i.e., the relative efficiency metric defined in Section 2.5). This can be thought of as an upper bound on the additional benefit of optimizing eye movements (assuming the optimal fixation is, indeed, at the mouth). Overall, observers were 43%±6% as efficient when fixating their preferred location compared to fixating the mouth (Fig. 8B, left). This suggests that, all else being equal, an optimal shift in eye movement strategy could theoretically support a greater than 100% gain in efficiency.

The findings are consistent with studies showing the benefits of learning eye movement strategies on perceptual performance in visual search tasks (Chukoskie et al., 2013; Droll et al., 2009; Koehler, Akbas, Peterson, & Eckstein, 2012). Yet, the importance or role of eye movements outside the scope of visual search has not been demonstrated. The results from this study are in contrast to a recent report where perceptual learning in an object recognition task was thought to be mediated solely through improved stimulus feature extraction (conceptualized as an expanded field of view) rather than through feature selection via eye movements (Holm, Engel, & Schrater, 2012). Certain key differences between that study and the current one provide useful insights into when eye movement modulation can and

does improve perceptual learning. In the object recognition task (Holm et al., 2012), observers located noisy contour-defined objects embedded in a field of distracting contour pieces. The objects were large, occupying approximately 25% of the display, with informative regions spanning the entire object contour. This study did not incorporate a forced fixation paradigm, making it impossible to quantify how different eye movement strategies might directly affect performance.

In the current study, observers identified exemplars from a single stimulus class where all task-relevant information was confined to a single region. This can be thought of as an extreme version of normal face recognition. Natural faces are comprised of a specific spatial configuration of discriminating features with differential information between regions. For instance, the eyes and eyebrows have been shown to contain a wealth of discriminatory information across the population, with the mouth and nose-tip regions containing significant but reduced information content (Peterson & Eckstein, 2012). This specific, highly consistent distribution of information, when combined with the foveated properties of the human visual system, leads to an optimal location for a first fixation about a third of the way down from the eyes to the nose tip (Peterson & Eckstein, 2012). Crucially, this location is always the best place to initially fixate. With the composite faces used in this study, the optimal location was artificially displaced downward to the mouth. Importantly, this optimal location was the same on every stimulus presentation. Furthermore, our forced fixation and guided exploration conditions revealed a significant and sizeable advantage for fixating close to the mouth. It should be noted that while the use of face stimuli with discriminatory information concentrated within a single feature might not seem ecologically valid, there are real world situations that may be close analogs. For instance, when trying to discriminate two twins, or even similar looking siblings that are close in age, one is often trying to learn particular discriminatory features within the face. At first the twins are hard to identify but once one learns the discriminatory feature the task becomes easier. While these situations may be atypical, they seem to arise in the natural environment, and with time people seem able to compensate for these unusual situations.

7.2. The factors that affect when eye movement learning occurs and its magnitude

We have shown that eye movement modulation can have a significant impact on perceptual learning. However, of the fourteen observers in this study, three failed to learn on their own while six only partially modified their fixation patterns. In this section, we discuss three factors that might influence whether eye movement learning will occur at all, and the magnitude of the modulation.

First, this study utilized a modified version of classic face identification tasks. Face recognition recruits a large network of specialized neural mechanisms (Haxby, Hoffman, & Gobbini, 2000; Kanwisher, McDermott, & Chun, 1997). Additionally, face identification is a highly overtrained skill with large social consequences. Indeed, the mechanisms and behaviors that support successful identification, including face-specific eye movement behavior, begin developing from birth (Farroni, Csibra, Simion, & Johnson, 2002; Morton & Johnson, 1991; Nelson, 2001). These factors lead to individuals developing distinct and highly consistent eye movement strategies for face identification, with the initial fixation

executed rapidly and automatically (Peterson & Eckstein, 2013). This automatic, stereotyped looking behavior might be difficult to modulate. Evidence for the well-practiced nature of normal face looking behavior can be seen by comparing the spread of the fixation distributions for each observer during Task 1 (normal faces) versus the fully learned state with the composite faces. When identifying normal faces, the average fixation standard deviation within an observer was 0.86±0.08 degrees visual angle. Over the final five sessions of Task 2, when the PMs and CMs were in a learned state, the Movers' fixation standard deviation had grown to 1.22±0.08 degrees, while the NMs retained their tight grouping with a spread of 0.65±0.07 degrees. After guided exploration, with all observers in a learned state, the average spread rose to 1.22±0.11 degrees (Fig. 8C). This suggests either a difference in saccadic targeting precision or uncertainty in the exact location to target. In either case, learning in this task required the observer to depart from familiar routines, which could make learning more difficult than with a less over-trained stimulus class and task (Chukoskie et al., 2013; Droll et al., 2009; Koehler et al., 2012).

Along these same lines, learning new visual tasks requires a willingness and ability to actively explore the stimulus and adjust or create new information acquisition and processing strategies. Furthermore, specific regions are selectively sampled from across the face, with information from the eye region exerting the greatest influence on perceptual decisions (Caldara et al., 2005; Schyns, Bonnar, & Gosselin, 2002; Sekuler, Gaspar, Gold, & Bennett, 2004). These automatic, overlearned mechanisms are incompatible with successful completion of the current study's task. With all information confined to the mouth area, a strategy that directs gaze to the upper part of the face and forms decisions using information from the eye region would be incapable of discriminating the four faces.

Exploration of the stimulus space, whereby different regions of the face are selectively sampled and used to form perceptual decisions, could theoretically be accomplished through both covert and overt attention mechanisms. The current paradigm does not allow us to make strong claims as to the relative utility of covert and overt exploration. However, the data from Task 3 (guided exploration) offers insights into possible mechanisms that might influence an observer's ability to learn. Clearly, whether by covert or overt means, the observer must be able to see the areas of the stimulus that are relevant to the task. Two differences between the Movers and Non Movers are apparent when looking at the guided exploration data. First, there was a trend toward NMs' preferred points of fixation for regular face identification being further from the mouth than the two Mover groups ($M_{distance\ from\ mouth\ center,\ NMs} = 6.37°$, $M_{distance\ from\ mouth\ center,\ Movers} = 5.28°$, $t = 1.58$, $p = .07$, one-tailed; Fig. 6B). This suggests that pre-existing oculomotor strategies can play a role, either facilitory or inhibitory, in the likelihood of learning a new task depending on the distance of the observers' typical preferred point of fixation from the feature that contains the discriminatory information. Second, performance as a function of fixation distance from the mouth decreased more precipitously for NMs than Movers ($PC_{fixate\ nose\ -\ fixate\ mouth,\ NMs} = 0.177$, $PC_{fixate\ nose\ -\ fixate\ mouth,\ Movers} = 0.078$, $t = 2.21$, $p = .02$, one-tailed; Fig. 6B). Taken together, these differences in preferred fixation location and peripheral processing ability resulted in large and significant differences between NMs and Movers in the mouth

region's visibility at the preferred fixation locations ($PC_{fixate\ preferred,\ NMs} = 0.399$, $PC_{fixate\ preferred,\ Movers} = 0.604$, $t = 2.02$, $p = .03$, one-tailed; Fig. 6B).

Finally, directed exploration can be beneficial even for people who appear to have fully learned the task. NMs entered the guided exploration sessions in a state of complete unawareness of the relevant stimulus properties as evidenced by their chance-level performance. Yet, they emerged from exploration with a well-formed perceptual strategy consistent with a trend toward optimality as seen with the large improvements in perceptual performance and drastic modulation of eye movements (Fig. 7AB). PMs, on the other hand, began guided exploration with at least some knowledge of the stimulus as evidenced by their significantly greater than chance performance. However, PMs did not fully modulate their eye movements and thus may have not fully optimized their perceptual strategy. This potential additional benefit was realized during exploration as can be seen by the reduced efficiency when forced to fixate their final chosen fixation relative to fixating the mouth ($\eta_{rel,final-mouth} = .737\pm.067$, $t = 3.92$, $p < .01$; Fig. 8B, right). This suggests that even when it may seem that observers have learned the task, guided exploration of the stimulus space may supplement voluntary learning mechanisms.

7.3. Factors to be explored in future investigations: stimulus duration and spatial scale

This paper focuses on eye movement modulation during perceptual learning for rapid identification of faces at a spatial scale corresponding to "normal" conversational interpersonal distances. The decisions regarding stimulus presentation time and the visual size of the face images were deliberately made. Here, we review the possible effects of these experimental parameters and offer possibilities for future investigations.

Regarding presentation time, recent studies have found that the first and second into-face fixations are the critical eye movements that support information acquisition for successful face identification (Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012), and thus we reasoned that the learning of these initial fixations was imperative to identification-specific learning effects. Furthermore, maintaining gaze for a prolonged period of time on the face of a person who is not being interacted with directly is deemed socially unacceptable in many cultures, making face exploration through quick glances relevant to real world scenarios. It is less clear how subsequent fixations affect recognition, or if they are germane to identification at all. By incorporating a speeded paradigm, we were able to create a situation where learning must take place during normal face-identification behavior. That being said, real world face identification is not necessarily restricted to quick glances. It is likely that novel faces with unusual information distributions could be learned more quickly over extended viewing times with many allowed fixations. Indeed, it has been shown that eye movement patterns to novel faces are distinct from those to familiar faces, and that eye movements during face learning seem to serve a functional role in later recognition (Barton, Radcliffe, Cherkasova, Edelman, & Intriligator, 2006; Henderson, Williams, & Falk, 2005). How might longer presentation times affect learning? As stated above, the first fixations seem to be generated by automatic behavior that has been developed by the brain to maximize fast face identification accuracy (Peterson & Eckstein, 2012). A case can be made that the failure of NMs to learn without explicit guidance can be linked to difficulty in

releasing from this overtrained behavior. Many previous studies show that fixations tend to diverge and spread out across the face with longer display times, usually resulting in a rough "feature targeting" pattern with saccades between the eyes and the lower nose/mouth (Barton et al., 2006; Henderson et al., 2005; Neumann, Spezio, Piven, & Adolphs, 2006; Pelphrey et al., 2002; Yarbus, 1967). If the NMs continued to predominantly fixate the upper face region, this would presumably preclude learning. If, however, NMs expanded their fixational coverage area with increasing stimulus display time, it is possible that the statistics of these novel faces could be learned without guided exploration. It would be interesting to see if this is in fact what would happen, and how learning with longer display times would translate to eye movement strategies for rapid identification.

A second manipulation that was not explored in the current work is the effect of the visual size of the stimulus. We scaled the face images so as to roughly approximate the retinal size of a face at a "normal conversational distance", a technique we have used extensively in the past and is common in the literature (Barton et al., 2006; Blais et al., 2008; Peterson & Eckstein, 2012, 2013). This decision was also made because it seems that novel faces that are deemed important enough to remember are often learned through visual experience at these closer distances. Of course, in real world situations humans look at other faces across a wide range of spatial scales with some invariance in recognition performance. An interesting question then is how learning might be affected by image size. The results found in this paper argue that a crucial factor mediating successful learning is the brain's ability to reliably access task-relevant visual information. The NMs tended to show very steep decreases in visibility with eccentricity that, combined with their natural tendency to look high up on the face, led to the task-relevant mouth region falling in areas of the visual field that correspond to poor processing power and resource allotment. Larger faces would create a situation where NMs would direct their initial fixations even further, in terms of retinal eccentricity, from the mouth. The prediction would be no change with this group. Smaller faces, such as those seen when viewing others from an appreciable distance, would move the informative mouth region closer to fixation. If the mouth was close enough so that the NMs could process it with similar fidelity as the CMs and PMs with the original-sized faces, the possibility exists that NMs would learn without the assistance of guided exploration. However, the NMs' neural implementation of their face recognition algorithm might still elect to sample information from inapposite regions of the face even when the mouth is clearly visible. We believe further investigations into the effect of scale on learning, and how learning eye movements at one scale might translate to new fixation strategies at another scale, would be a fruitful endeavor toward a complete understanding of eye movement learning during perceptual learning.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Ahissar M, Hochstein S. Attentional control of early perceptual learning. Proceedings of the National Academy of Sciences. 1993; 90(12):5718–5722.

Atienza M, Cantero JL, Dominguez-Marin E. The Time Course of Neural Changes Underlying Auditory Perceptual Learning. Learning & Memory. 2002; 9(3):138–150.10.1101/lm.46502 [PubMed: 12075002]

Ball K, Sekuler R. A specific and enduring improvement in visual motion discrimination. Science. 1982; 218(4573):697–698.10.1126/science.7134968 [PubMed: 7134968]

Ball K, Sekuler R. Direction-specific improvement in motion discrimination. Vision Research. 1987; 27(6):953–965. 16/0042-6989(87)90011-3. [PubMed: 3660656]

Barlow HB. The absolute efficiency of perceptual decisions. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 1980; 290(1038):71–82.

Barton JJS, Radcliffe N, Cherkasova MV, Edelman J, Intriligator JM. Information processing during face recognition: the effects of familiarity, inversion, and morphing on scanning fixations. Perception. 2006; 35(8):1089–1105. [PubMed: 17076068]

Beard BL, Ahumada AJ Jr. Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. Journal of the Optical Society of America. A, Optics, Image Science, and Vision. 1999; 16(3):755–763.

Bende M, Nordin S. Perceptual Learning in Olfaction: Professional Wine Tasters versus Controls. Physiology & Behavior. 1997; 62(5):1065–1070.10.1016/S0031-9384(97)00251-5 [PubMed: 9333201]

Blais C, Jack RE, Scheepers C, Fiset D, Caldara R. Culture Shapes How We Look at Faces. PLoS ONE. 2008; 3(8):e3022.10.1371/journal.pone.0003022 [PubMed: 18714387]

Burgess AE, Wagner RF, Jennings RJ, Barlow HB. Efficiency of human visual signal discrimination. Science (New York, NY). 1981; 214(4516):93–94.

Caldara R, Schyns P, Mayer E, Smith ML, Gosselin F, Rossion B. Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. Journal of Cognitive Neuroscience. 2005; 17(10):1652–1666.10.1162/089892905774597254 [PubMed: 16269103]

Chukoskie L, Snider J, Mozer MC, Krauzlis RJ, Sejnowski TJ. Learning where to look for a hidden target. Proceedings of the National Academy of Sciences. 2013; 110(Supplement_2):10438–10445.10.1073/pnas.1301216110

Dosher BA, Lu ZL. Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. Proceedings of the National Academy of Sciences of the United States of America. 1998; 95(23):13988–13993. [PubMed: 9811913]

Droll JA, Abbey CK, Eckstein MP. Learning cue validity through performance feedback. Journal of Vision. 2009; 9(2):18, 1–23.10.1167/9.2.18 [PubMed: 19271928]

Eckstein MP, Abbey CK, Pham BT, Shimozaki SS. Perceptual learning through optimization of attentional weighting: human versus optimal Bayesian learner. Journal of Vision. 2004; 4(12):1006–1019. 10:1167/4.12.3. [PubMed: 15669908]

Eckstein MP, Beutter BR, Pham BT, Shimozaki SS, Stone LS. Similar neural representations of the target for saccades and perception during search. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 2007; 27(6):1266–1270.10.1523/JNEUROSCI. 3975-06.2007 [PubMed: 17287501]

Eckstein MP, Beutter BR, Stone LS. Quantifying the performance limits of human saccadic targeting during visual search. Perception. 2001; 30(11):1389–1401. [PubMed: 11768491]

Farroni T, Csibra G, Simion F, Johnson MH. Eye contact detection in humans from birth. Proceedings of the National Academy of Sciences. 2002; 99(14):9602–9605.10.1073/pnas.152159999

Fine I, Jacobs RA. Comparing perceptual learning tasks: a review. Journal of Vision. 2002; 2(2):190–203. 10:1167/2.2.5. [PubMed: 12678592]

Fiorentini A, Berardi N. Perceptual learning specific for orientation and spatial frequency. 1980; 287(5777):43–44.10.1038/287043a010.1038/287043a0

Gilbert CD, Sigman M, Crist RE. The neural basis of perceptual learning. Neuron. 2001; 31(5):681–697. [PubMed: 11567610]

Gold J, Bennett PJ, Sekuler AB. Signal but not noise changes with perceptual learning. Nature. 1999; 402(6758):176–178.10.1038/46027 [PubMed: 10647007]

Goldstone RL. Perceptual learning. Annual Review of Psychology. 1998; 49:585–612.10.1146/annurev.psych.49.1.585

Green, DM. Signal detection theory and Psychophysics. New York: Wiley; 1966.

Hawkey DJC, Amitay S, Moore DR. Early and rapid perceptual learning. Nature Neuroscience. 2004; 7(10):1055–1056.10.1038/nn1315

Haxby JV, Hoffman EA, Gobbini MI. The distributed human neural system for face perception. Trends in Cognitive Sciences. 2000; 4:223–233.10.1016/S1364-6613(00)01482-0 [PubMed: 10827445]

Hayhoe M, Ballard D. Eye movements in natural behavior. Trends in Cognitive Sciences. 2005; 9(4): 188–194. 16/j.tics.2005.02.009. [PubMed: 15808501]

Henderson JM, Williams CC, Falk RJ. Eye movements are functional during face learning. Memory & Cognition. 2005; 33:98–106.10.3758/BF03195300 [PubMed: 15915796]

Holm L, Engel S, Schrater P. Object Learning Improves Feature Extraction but Does Not Improve Feature Selection. PLoS ONE. 2012; 7(12):e51325.10.1371/journal.pone.0051325 [PubMed: 23251499]

Hsiao J, Cottrell G. Two Fixations Suffice in Face Recognition. Psychological Science. 2008; 19(10): 998–1006.10.1111/j.1467-9280.2008.02191.x [PubMed: 19000210]

Hurlbert A. Visual perception: Learning to see through noise. Current Biology: CB. 2000; 10(6):R231–233. [PubMed: 10744966]

Ito M, Westheimer G, Gilbert CD. Attention and Perceptual Learning Modulate Contextual Influences on Visual Perception. Neuron. 1998; 20(6):1191–1197.10.1016/S0896-6273(00)80499-7 [PubMed: 9655506]

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience. 1997; 17(11):4302–4311. [PubMed: 9151747]

Karni A, Sagi D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proceedings of the National Academy of Sciences. 1991; 88(11):4966–4970.

Karni A, Sagi D. The time course of learning a visual skill. Nature. 1993; 365(6443):250–252.10.1038/365250a0 [PubMed: 8371779]

Koehler K, Akbas E, Peterson M, Eckstein MP. Human versus Bayesian Optimal Learning of Eye Movement Strategies During Visual Search. Journal of Vision. 2012; 12(9):1142–1142.10.1167/12.9.1142

Li RW, Levi DM, Klein SA. Perceptual learning improves efficiency by re-tuning the decision "template" for position discrimination. Nature Neuroscience. 2004; 7(2):178–183.10.1038/nn1183

Lu ZL, Dosher BA. Characterizing human perceptual inefficiencies with equivalent internal noise. Journal of the Optical Society of America. A, Optics, Image Science, and Vision. 1999; 16(3): 764–778.

Lu ZL, Dosher BA. External noise distinguishes attention mechanisms. Vision Research. 1998; 38(9): 1183–1198.10.1016/S0042-6989(97)00273-3 [PubMed: 9666987]

Matthews N, Liu Z, Geesaman BJ, Qian N. Perceptual learning on orientation and direction discrimination. Vision Research. 1999; 39(22):3692–3701. [PubMed: 10746139]

Moreno MM, Linster C, Escanilla O, Sacquet J, Didier A, Mandairon N. Olfactory perceptual learning requires adult neurogenesis. Proceedings of the National Academy of Sciences. 2009; 106(42): 17980–17985.10.1073/pnas.0907063106

Morton J, Johnson MH. CONSPEC and CONLERN: A two-process theory of infant face recognition. Psychological Review. 1991; 98(2):164–181.10.1037/0033-295X.98.2.164 [PubMed: 2047512]

Najemnik J, Geisler WS. Optimal eye movement strategies in visual search. Nature. 2005; 434(7031): 387–391.10.1038/nature03390 [PubMed: 15772663]

Najemnik J, Geisler WS. Eye movement statistics in humans are consistent with an optimal search strategy. Journal of Vision. 2008; 8(3)10.1167/8.3.4

Najemnik J, Geisler WS. Simple summation rule for optimal fixation selection in visual search. Vision Research. 2009; 49(10):1286–1294.10.1016/j.visres.2008.12.005 [PubMed: 19138697]

Nelson CA. The development and neural bases of face recognition. Infant and Child Development. 2001; 10(1–2):3–18.10.1002/icd.239

Neumann D, Spezio ML, Piven J, Adolphs R. Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. Social Cognitive and Affective Neuroscience. 2006; 1(3):194–202.10.1093/scan/nsl030 [PubMed: 18985106]

Pelphrey KA, Sasson NJ, Reznick JS, Paul G, Goldman BD, Piven J. Visual scanning of faces in autism. Journal of Autism and Developmental Disorders. 2002; 32(4):249–261. [PubMed: 12199131]

Peterson MF, Abbey CK, Eckstein MP. The surprisingly high human efficiency at learning to recognize faces. Vision Research. 2009; 49(3):301–314.10.1016/j.visres.2008.10.014 [PubMed: 19000918]

Peterson MF, Eckstein MP. Looking just below the eyes is optimal across face recognition tasks. Proceedings of the National Academy of Sciences. 2012; 109(48):E3314–E3323.10.1073/pnas. 1214269109

Peterson MF, Eckstein MP. Individual Differences in Eye Movements During Face Identification Reflect Observer-Specific Optimal Points of Fixation. Psychological Science. 201310.1177/0956797612471684

Pleger B, Foerster AF, Ragert P, Dinse HR, Schwenkreis P, Malin JP, Tegenthoff M. Functional Imaging of Perceptual Learning in Human Primary and Secondary Somatosensory Cortex. Neuron. 2003; 40(3):643–653.10.1016/S0896-6273(03)00677-9 [PubMed: 14642286]

Polley DB, Steinberg EE, Merzenich MM. Perceptual Learning Directs Auditory Cortical Map Reorganization through Top-Down Influences. The Journal of Neuroscience. 2006; 26(18):4970–4982.10.1523/JNEUROSCI.3771-05.2006 [PubMed: 16672673]

Rao RPN, Zelinsky GJ, Hayhoe MM, Ballard DH. Eye movements in iconic visual search. Vision Research. 2002; 42(11):1447–1463. [PubMed: 12044751]

Rayner K. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin. 1998; 124(3):372–422.10.1037/0033-2909.124.3.372 [PubMed: 9849112]

Saarinen J, Levi DM. Perceptual learning in vernier acuity: What is learned? Vision Research. 1995; 35(4):519–527.10.1016/0042-6989(94)00141-8 [PubMed: 7900292]

Sathian K, Zangaladze A. Perceptual learning in tactile hyperacuity: complete intermanual transfer but limited retention. Experimental Brain Research. 1998; 118(1):131–134.10.1007/s002210050263 [PubMed: 9547071]

Schyns PG, Bonnar L, Gosselin F. Show me the features! Understanding recognition from the use of visual information. Psychological Science. 2002; 13(5):402–409. [PubMed: 12219805]

Sekuler AB, Gaspar CM, Gold J, Bennett PJ. Inversion Leads to Quantitative, Not Qualitative, Changes in Face Processing. Current Biology. 2004; 14(5):391–396. 16/j.cub.2004.02.028. [PubMed: 15028214]

Tavassoli A, van der Linde I, Bovik AC, Cormack LK. Eye movements selective for spatial frequency and orientation during active visual search. Vision Research. 2009; 49(2):173–181.10.1016/j.visres.2008.10.005 [PubMed: 18992270]

Trenti EJ, Barraza JF, Eckstein MP. Learning motion: human vs. optimal Bayesian learner. Vision Research. 2010; 50(4):460–472.10.1016/j.visres.2009.10.018 [PubMed: 19883676]

Wilson DA, Stevenson RJ. Olfactory perceptual learning: the critical role of memory in odor discrimination. Neuroscience & Biobehavioral Reviews. 2003; 27(4):307–328.10.1016/S0149-7634(03)00050-2 [PubMed: 12946684]

Yarbus, A. Eye Movements and Vision. New York: Plenum Press; 1967.

Zelinsky GJ, Rao RPN, Hayhoe MM, Ballard DH. Eye Movements Reveal the Spatiotemporal Dynamics of Visual Search. Psychological Science. 1997; 8(6):448–453.10.2307/40063232

Highlights

- Assessed the role of eye movement modification for perceptual learning.

- Eye movements can account for a large portion of learning (here, 43%).

- Learning can be predicted by default looking behavior and peripheral ability.

- Guided exploration of a novel stimulus greatly facilitates learning.
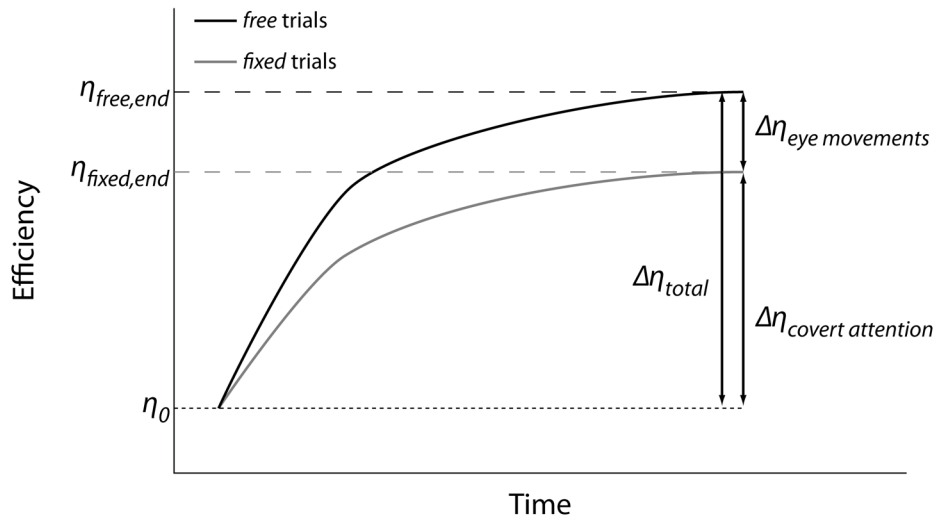
Fig. 1.
Learning is defined as an increase in task efficiency with pratice. Changes in efficiency for *fixed* trials (gray solid line) are mediated by covert mechanisms only. Efficiency changes in *free* trials (black solid line) are a consequence of modifications to both covert and overt (eye movement) mechanisms. Thus, the additional improvement in efficiency afforded by eye movements can be thought of as the difference in efficiency changes between the *free* and *fixed* conditions.
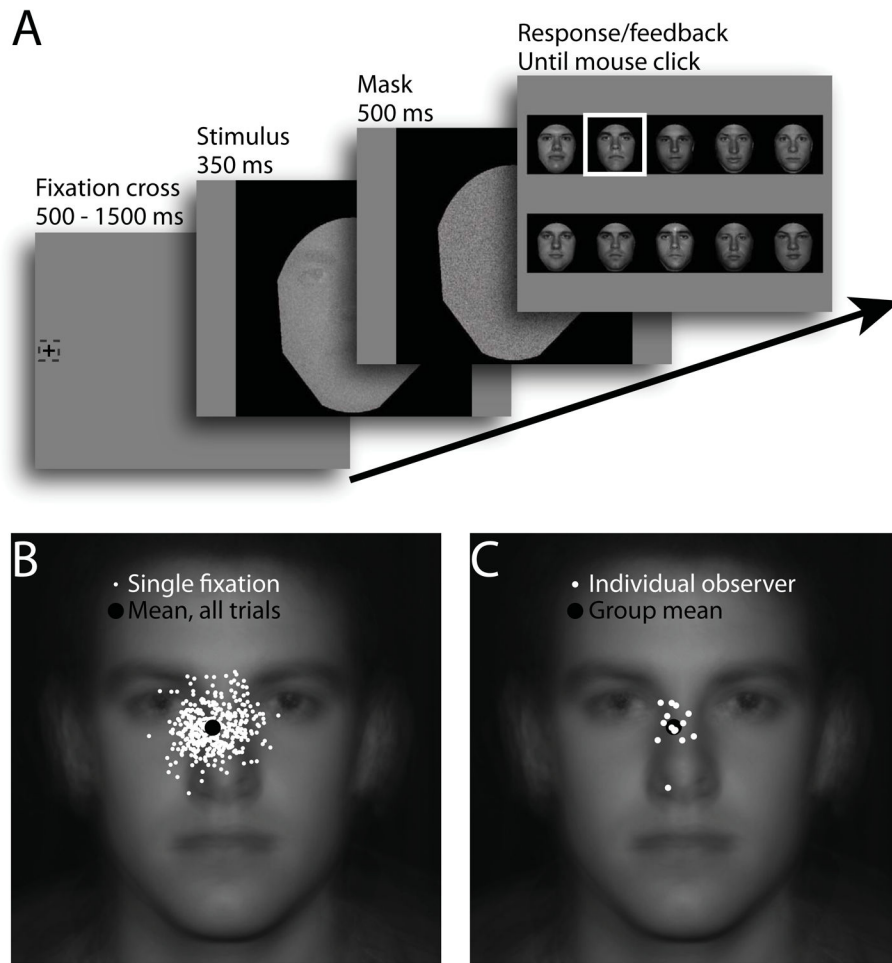
Fig. 2.
Eye movement behavior during normal face identification. (A) Observers freely moved their eyes from an initial peripheral fixation (13.3 degrees from the center of the image) into a briefly displayed, centrally presented noisy face image. (B) Each of 400 trials resulted in a single eye movement (saccade endpoints indicated by white dots), with a representative observer's data shown here (black dot representing the mean across the 400 trials). (C) The mean landing points of the into-face eye movements are shown for each of the fourteen observers (white dots) along with the average landing point for the group (black dot).
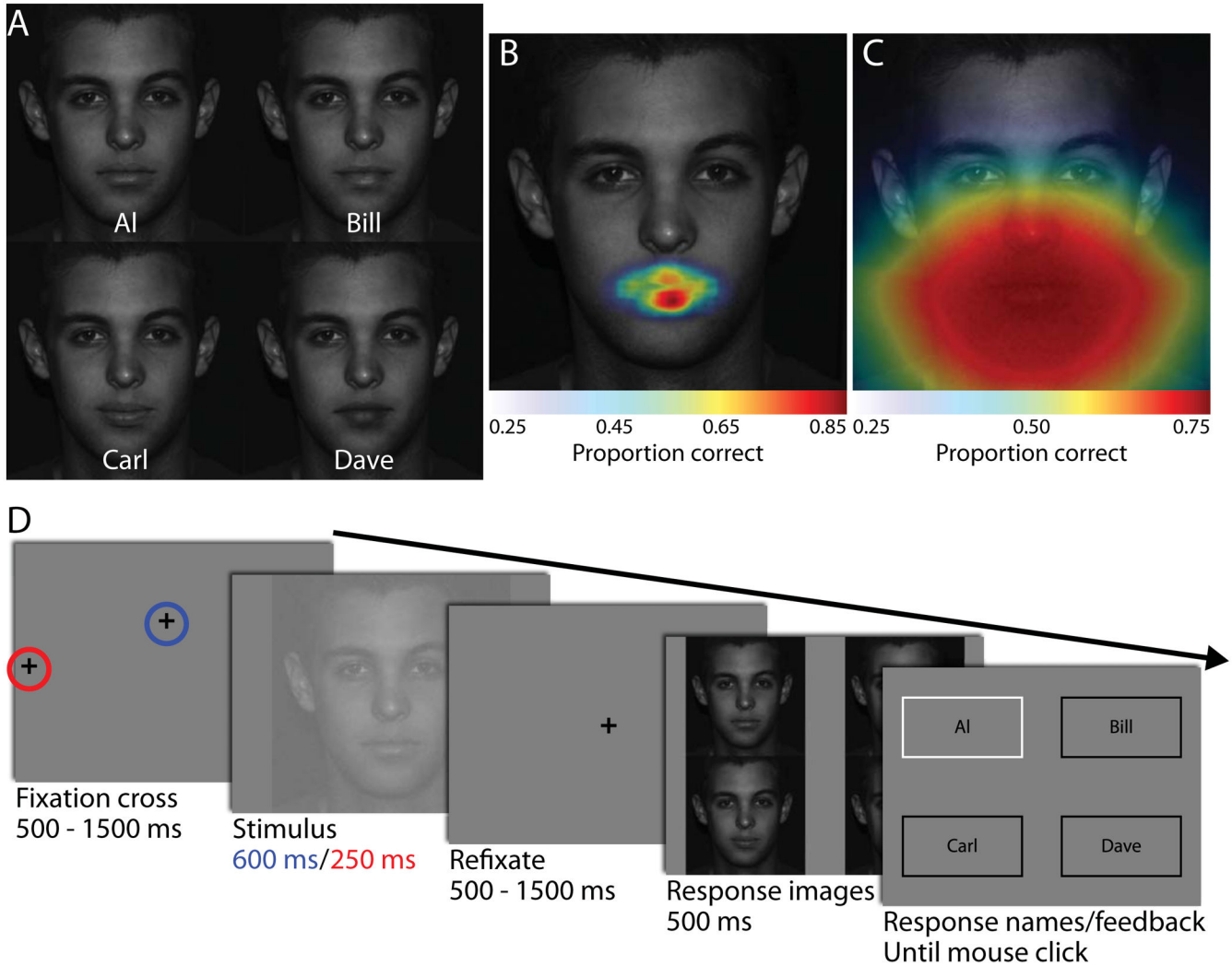
Fig. 3.
Identifying unusual faces. (A) Novel face images were created by blending four different mouths into a single base face. Faces were exactly the same except for the region outlined in red on the left-most face. (B) Simulation results from a region of interest ideal observer, which quantifies the amount of task-relevant information available in each small region of the face. Lack of color indicates no information, while red colors indicate regions of high information content. (C) The foveated ideal observer is a model that simulates the human visual system's inhomogenous contrast sensitivity across the visual field. Predicted performance for any given point of fixation is shown, again with blue areas indicating fixation locations predicted to lead to low performance and red for high-performance fixations. (D) Odd trials began with an initial fixation to the extreme left or right of the monitor (*free*, circled in red) while even trials began with a fixation at the observer's personal preferred point (*fixed*, circled in blue). After a brief stimulus presentation, observers refixated the center of the screen before a briefly displayed response image screen.
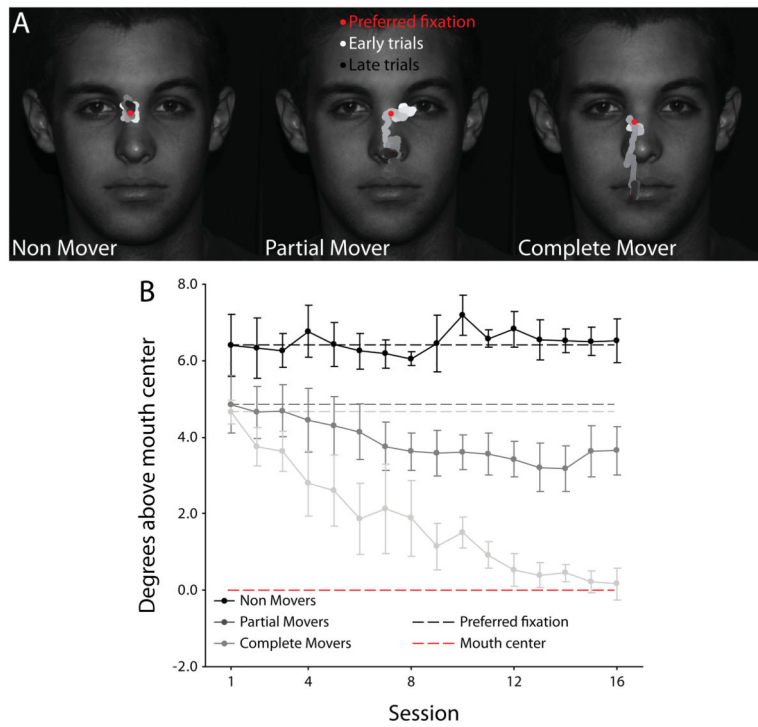
Fig. 4.
Three different observer types based on eye movement behavior. (A) Each observer fell into one of three categories based on the extent of their eye movement modulation. Representative observers are shown for each group, with the red dot indicating the observer's preferred point of fixation for normal faces and the white (early sessions) to black (late sessions) trace showing the observer's average fixation location with a 50-trial moving window. (B) Group data averaged across observers showing mean fixation location in the vertical dimension as a function of session. Grey dotted lines indicate each group's average preferred fixation for normal faces. Red dotted line indicates the center of the informative mouth region. Error bars represent one standard error of the mean (*SEM*) across observers.
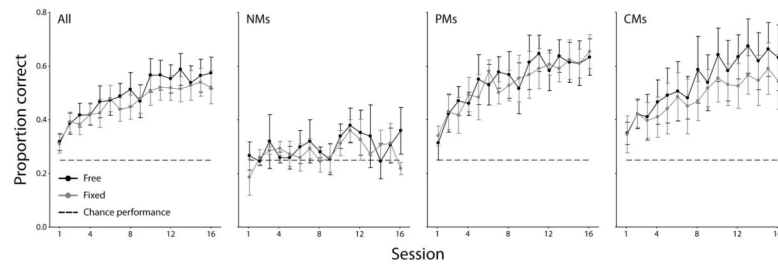
Fig. 5.
Perceptual learning as performance improvement. Perceptual performance, in terms of proportion correct, is shown as a function of learning session. The two mover groups improved substantially above chance, but only the Complete Movers reaped benefits when allowed to move their eyes. Error bars represent one *SEM* across observers.
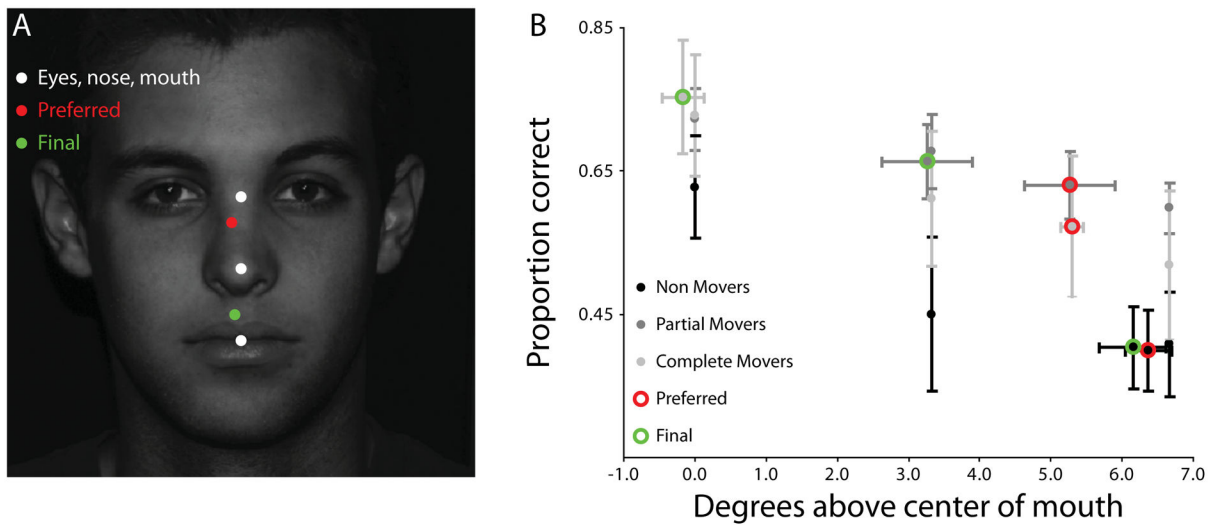
Fig. 6.
Guided exploration with the novel stimuli. (A) Observers were forced to fixate various locations during stimulus presentation. Three locations (eyes, nose, mouth; white dots) were common to all observers. The observer's preferred fixation (red dot) and mean fixation location in *free* trials at the end of Task 2 (green dot) determined the final two locations. (B) Average performances in terms of proportion correct are shown as a function of fixation location for each group, showing a general advantage for fixating closer to the mouth. Error bars represent one *SEM* across observers.
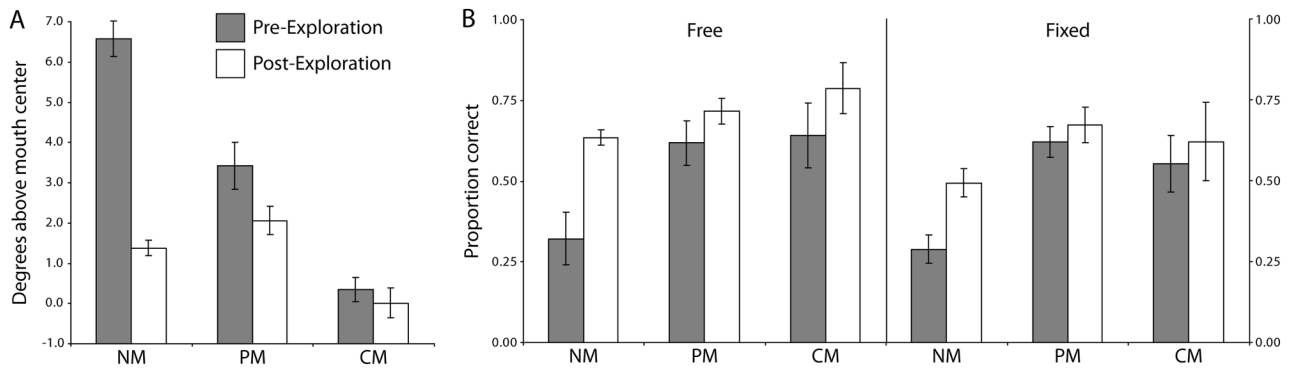
Fig. 7.
Effects of guided exploration on task behavior. Data is shown for the three groups for the five sessions immediately before (grey bars) and immediately after (white bars) guided exploration. (A) Average distance of fixation above the mouth center in the *free* condition, showing successful learning for the NMs after guided exposure to the mouth region. (B) Average performance in terms of proportion correct. Error bars represent one *SEM* across observers.
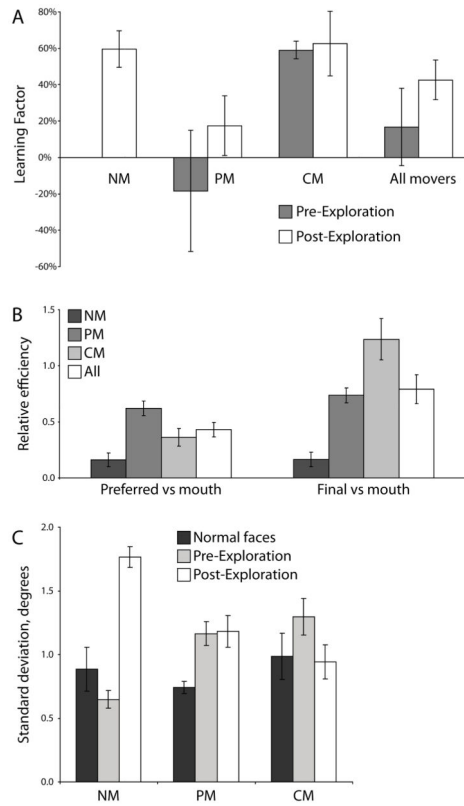
Fig. 8.
Predictors of learning. (A) Percentage of total learning due to eye movement modulation above and beyond covert mechanisms is quantified using the Learning Factor. Before guided exploration, only CMs saw a definitive advantage when allowed to choose their gaze behavior. All groups benefited once the stimulus was fully explored. (B) On the left, each group showed a different ability to extract information from their preferred point of fixation relative to optimal as measured by the relative efficiency metric. On the right, pre-exploration eye movement strategy efficacies are reflected in each group's efficiency at the their final fixation location versus optimal. (C) The average spread of each observer's fixations when identifying normal faces versus novel composite faces. Error bars represent one *SEM* across observers.