Initial eye movements during face identification are optimal and similar across cultures

Charles C.-F. Or

Matthew F. Peterson

Miguel P. Eckstein

Culture influences not only human high-level cognitive processes but also low-level perceptual operations. Some perceptual operations, such as initial eye movements to faces, are critical for extraction of information supporting evolutionarily important tasks such as face identification. The extent of cultural effects on these crucial perceptual processes is unknown. Here, we report that the first gaze location for face identification was similar across East Asian and Western Caucasian cultural groups: Both fixated a featureless point between the eyes and the nose, with smaller between-group than within-group differences and with a small horizontal difference across cultures (8% of the interocular distance). We also show that individuals of both cultural groups initially fixated at a slightly higher point on Asian faces than on Caucasian faces. The initial fixations were found to be both fundamental in acquiring the majority of information for face identification and optimal, as accuracy deteriorated when observers held their gaze away from their preferred fixations. An ideal observer that integrated facial information with the human visual system's varying spatial resolution across the visual field showed a similar information distribution across faces of both races and predicted initial human fixations. The model consistently replicated the small vertical difference between human fixations to Asian and Caucasian faces but did not predict the small horizontal leftward bias of Caucasian observers. Together, the results suggest that initial eye movements during face identification may be driven by brain mechanisms aimed at maximizing accuracy, and less influenced by culture. The findings increase our understanding of the interplay between the brain's aims to optimally accomplish basic perceptual functions and to respond to sociocultural influences.

Introduction

Recent research in cultural psychology suggests that culture can influence not only human high-level cognitive processes, such as decision making and social interaction, but also low-level sensory and perceptual operations (for a review, see Nisbett & Miyamoto, 2005). Studies have suggested that when making perceptual judgments (e.g., when orienting a line inside a tilted square frame, see Ji, Peng, & Nisbett, 2000), Western Caucasians (white people from Western culture; Caucasians, in short) tend to focus on salient features, analyze their attributes, and assign them to categories, but ignore the context, such as the


doi: 10.1167/15.13.12

Received October 13, 2014; published September 18, 2015

ISSN 1534-7362 © 2015 ARVO

Downloaded From: https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934452/ on 11/03/2018
characteristics of the neighboring features and the background. This is often described as the “analytic” style. In contrast, East Asians (Asians, in short) appear to attend more to contextual information surrounding the features, and consider more the relationships and similarities among objects. This is referred to as the “holistic” style (Nisbett, Peng, Choi, & Norenzayan, 2001).

Similarly, it has been suggested that eye movements, which allow humans to direct their high-resolution fovea to areas of interest in the environment, also appear to vary across cultures. For example, Chua, Boland, and Nisbett (2005) monitored the eye movements of Chinese and Caucasian participants in a scene-perception task where they viewed scenes with objects on complex backgrounds, such as a tiger in the woods. They found that Caucasians looked at the objects sooner and fixated them more frequently than the Chinese observers. On the other hand, Chinese participants made more saccades to the background than Caucasians. In relation to eye movements to faces, Caucasians display a triangular eye-movement pattern with fixations over the eyes, nose, and mouth (Yarbus, 1965). In contrast, a series of studies (Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Kelly et al., 2011; Kelly, Miellet, & Caldara, 2010; Rodger, Kelly, Blais, & Caldara, 2010) has reported that Asians tend to fixate the face centrally on the nose region over multiple eye movements, suggesting that they adopt a more holistic strategy by focusing more on the context and associations among face features such that a central fixation would be an optimal strategy. Other studies, however, have suggested that the eye-movement patterns differ instead when participants viewed own-race versus other-race faces, regardless of their own cultural background (Fu, Hu, Wang, Quinn, & Lee, 2012; Goldinger, He, & Papesh, 2009; Wheeler et al., 2011; Wu, Laeng, & Magnussen, 2012; Xiao, Xiao, Quinn, Anzures, & Lee, 2013).

There are, however, studies that have not found any cultural effect on eye movements. Rayner, Li, Williams, Cave, and Well (2007) recorded generally consistent eye-movement patterns across Caucasian and Chinese participants for a variety of tasks, including reading, face processing, scene perception, and visual search. Notably, the Chinese participants did not spend more time than the Caucasian participants in viewing the background of a complex scene, though the scenes they used contained more salient objects than those used by Chua et al. (2005). The cultural consistency was replicated in a later study (Rayner, Castelhano, & Yang, 2009), when Caucasian and Chinese participants viewed normal and unusual scenes. These results suggest that cultural effects on eye movements may depend upon specific task requirements, and highlight the question about the extent of cultural influences on eye movements.

This article reexamines the issue of cultural influence on eye-movement patterns to faces, with a focus on face identification and initial fixations, as Hsiao and Cottrell (2008) have reported that the first, and to a lesser extent the second, eye movement is paramount in acquiring visual information critical for identification in a standard face learning-and-recognition task (i.e., participants first studied a series of faces and then performed a face memory test containing studied and novel faces). Nevertheless, the majority of previous studies have analyzed an ensemble of eye movements over an extended period of time during face scanning, the majority of which were executed after the information critical for face identification had been acquired. Here, we isolate early fixations to determine whether culture influences eye movements that support the perceptual processing of identity information.

A separate literature has shown that brain processes guide eye movements in order to maximize the acquisition of visual information and successfully accomplish fundamental perceptual tasks such as visual search (Findlay, 1997; Najemnik & Geisler, 2005, 2008; S. Zhang & Eckstein, 2010) and reading (Legge, Klitz, & Tjan, 1997; for a review, see Rayner, 2009). This is also the case for face recognition. In particular, Peterson and Eckstein (2012) have suggested that Caucasian observers’ initial eye movements to Caucasian faces maximize perceptual accuracy (see also Hsiao & Liu, 2012, who reported that the optimal horizontal fixation position was left to the center of the face). For the majority of Caucasian observers, this optimal point of fixation is a featureless point just below the eyes (but see Peterson & Eckstein, 2013, for individual differences). Forcing participants to fixate elsewhere along the vertical midline decreased identification performance. This empirical finding could be explained by the distribution of discriminatory information across faces, the varying resolution of the human visual system, and a Bayesian rational decision mechanism. The authors derived an ideal observer that performed the face-recognition tasks by integrating information optimally across the face while considering the limitations introduced by the reduced sensitivity away from the fovea (thus named the foveated ideal observer [FIO]). Performance was maximized when the FIO made its first fixation just below the eyes, the same face region as empirically found with human observers, suggesting that humans are optimal in planning their eye-movement strategy during face identification.

In the current study, we asked whether culture influences the initial eye movements during face identification by adopting the 1-in-10 face-identification experimental paradigm from Peterson and Eck-
Here, Asian and Caucasian participants were instructed to identify Asian and Caucasian faces. In addition, we evaluated whether the initial eye-movement strategies across cultural groups could be predicted by any potential differences in the culture-specific distributions of visual information across faces using an FIO model for analyzing Asian and Caucasian faces. For example, we might find that there were significant differences in eye movements across cultural groups that reflected distinct optimal points of fixation specific to faces of each culture. On the other hand, differences across cultural groups might not be captured by variations in optimal points of fixation for faces of each culture. Finally, there might not be any differences between initial eye movements to faces across cultural groups, which might or might not be explained by the FIO model.

**Overview of experiments**

Five experiments were conducted in which participants performed a variety of face-identification tasks with Asian and Caucasian face stimuli in separate conditions. In Experiment 1, participants performed 1-in-10 face-identification tasks (Figure 1; Peterson & Eckstein, 2012), in which they were instructed to freely move their eyes towards the grayscale, noisy face (presented for 350 ms) and identify the face among 10 choices of the same race. Their initial eye-movement patterns to Asian and Caucasian faces were compared. In Experiment 2, participants performed 1-in-10 face-identification tasks again, but with longer stimulus exposure (1500 ms) in order to assess fixation patterns in later saccades. In Experiment 3, participants performed face-familiarity tasks, in which they determined whether famous faces in color were familiar in order to assess whether initial eye-movement patterns in the first two experiments would be generalized to other faces. In Experiment 4, additional 1-in-10 face-identification tasks (modified from Experiments 1 and 2 to restrict the number of saccades on faces) were performed in order to evaluate the functional contributions of initial eye movements to the overall acquisition of visual information supporting face identification (compare Hsiao & Cottrell, 2008). We also examined whether cultural diversity would be observed in later saccades during prolonged (5000 ms) face presentations. In Experiment 5, the 1-in-10 face-identification tasks were modified such that participants were forced to fixate at various vertical positions of the face throughout the trials (Figure 1). The functional importance of the Asian and Caucasian participants’ preferred initial fixation positions was evaluated by comparing the identification accuracies for different fixation positions. In addition, we applied the FIO model (Peterson & Eckstein, 2012) to Asian and Caucasian faces in order to assess whether the cross-cultural gaze behavior can be explained by brain processes guiding eye movements to optimize perceptual performance based on the interaction between the distribution of visual information within faces of each culture and the foveated nature of the human visual system.
General methods

Participants

The experiments were completed by 16 East Asian (eight female and eight male) and 16 Western Caucasian (eight female and eight male) participants. Each participant was asked to fill out a questionnaire about cultural background and experience. The Asian participants (14 from China, one from Hong Kong, one from South Korea) arrived in the United States from 1 to 12 months prior to the experiment, and none had extended travel outside of East Asian countries prior to arrival in the United States. Their estimated number of non-Asian friends and relatives averaged 4.47 (SD: 6.51). The Caucasian participants (15 from the United States, one from France) had limited non-Caucasian experience: None had made extended stays outside of the West, and they estimated to have on average 17.8 (SD: 15.4) non-Caucasian friends and relatives. Thus, neither Asian nor Caucasian participants had extensive experience seeing faces of the other race. All participants were unaware of the purpose of study, had normal or corrected-to-normal visual acuity, and received an honorarium for the study. Their ages ranged between 18 and 27 years. All participants provided informed consent following the guidelines of the institutional review board of the University of California, Santa Barbara.

Stimulus display

The stimuli were generated by a Dell Vostro 400 computer installed with MATLAB version 7.9, passed to an NVIDIA GeForce 7900 GS Graphics card, and were displayed on a ViewSonic Graphics Series G90fb monitor at a refresh rate of 100 Hz, with a spatial resolution of 800 × 600 pixels and a grayscale resolution of 8 bits/pixel. The participants viewed the screen binocularly at a distance of 71 cm in a dark observation chamber, such that each square pixel had a width of 0.037°. Prior to testing, the monitor was linearly gamma corrected using Spyder3Pro 4.0.2. The mean luminance after gamma correction was 27.7 cd/m².

Eye tracking

During the course of the experiment, the left eye of each participant was monitored using an EyeLink CL Desktop Mount eye tracker sampling at 250 Hz. A nine-point calibration and validation procedure was conducted before each experimental session, with a mean error of no more than 0.5° of visual angle.

Saccades were classified as events in which eye velocity was greater than 35°/s and eye acceleration exceeded 9500°/s².

Stimuli

The face stimuli to be described were used in all experiments and modeling except Experiment 3. Separate databases were used for Asian and Caucasian faces. For Asian faces, 10 images were taken from the face database of the Visual Cognition Laboratory at the University of Hong Kong. These photos were taken under symmetric lighting conditions. Glasses, jewelry, and makeup were removed before the volunteers were photographed. For Caucasian faces, 10 images of undergraduate students at the University of California, Santa Barbara, were taken in-house with constant diffuse lighting, distance, and camera settings. All Asian and Caucasian face images were male, posed at the frontal view, and showed neutral expression. These images were cropped to remove background, hair, face outline, and ears. The face was then centered on the image and scaled to measurements detailed in Table 1. The images were subsequently converted to have 8-bit grayscale resolution, and normalized such that the luminance profile of each image averaged to the mean luminance of the screen and that the contrast was reduced to 17% of the original images. Zero-mean white Gaussian noise, with a standard deviation of 2.44 cd/m², was added to each image.

Experiment 1: Free eye movements (350-ms presentation)

Methods

Procedure

Each participant performed both Asian and Caucasian face identification (Figure 1; 500 trials per race in five sessions of 100 trials each; the order of the two face-race conditions was counterbalanced across participants). Following the calibration and validation procedure outlined previously under Eye tracking, participants fixated a cross displayed on a uniform, mean-luminance gray background and pressed the space bar to proceed with a trial. To prevent any initial fixation bias (Arizpe, Kravitz, Yovel, & Baker, 2012; Hsiao & Cottrell, 2008), the fixation cross appeared randomly at one of two positions near the side of the monitor (13.3° from the center of screen) such that it did not overlap with the impending stimulus region. After the space bar was pressed, the fixation cross remained visible for a random period of 500–1500 ms,
during which eye movements were not allowed. (Any gaze greater than 1.1° from the center of the fixation cross aborted and restarted the trial; the 1.1° tolerance was chosen such that it covered the area of the fixation cross.) As soon as the fixation cross vanished, a noisy, contrast-reduced face stimulus (randomly chosen with equal probability from the current race’s 10 face images) was presented centrally on the screen for 350 ms, during which eye movements were unrestricted. The stimulus was replaced by a centrally presented, high-contrast white Gaussian noise mask (average luminance = mean luminance of screen, standard deviation = 7.33 cd/m²) for 500 ms, followed by a response screen showing noise-free, 100%-contrast versions of the 10 possible faces. Participants used a mouse to click on the face that they thought matched the previously presented stimulus and were provided the correct answer before the next trial commenced.

### Table 1. Measurements of 20 Asian and Caucasian face stimuli used in the experiments.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Visual angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean distance between the bottom of the hairline and the bottom of the chin</td>
<td>20.7°</td>
</tr>
<tr>
<td>Mean distance between the centers of the two eyes (interocular distance)</td>
<td>6.93°</td>
</tr>
<tr>
<td>Mean vertical distance between the eyes and the nose (eye–nose distance)</td>
<td>4.80°</td>
</tr>
<tr>
<td>Standard deviation of the vertical position of the eyes</td>
<td>0.26°</td>
</tr>
</tbody>
</table>

Results

As the 10 faces in each task were repeatedly (and randomly) presented throughout the experiment, the first fixation position could be shifted over time due to face learning. To rule out such a possibility, we computed the 25-trial running average of the first fixation position over both the vertical and horizontal dimensions (Supplementary Figure S1) and used a mixed-model repeated measures ANOVA to analyze three main effects (trial group, observer culture, face race) and their interactions (Supplementary Table S1). Notably, the main effect of trial group and its associated interactions were not significant for either the vertical or horizontal dimensions. Thus, the first fixation position did not vary systematically throughout trials despite increasing familiarity of the stimuli.

We subsequently averaged the first fixation positions over all trials, separately for each observer, face race, and dimension (Figure 2A), and used a mixed-model repeated measures ANOVA to analyze two main effects (observer culture, face race) and their interaction (Table 2). Notably, the mean first fixation positions did not differ significantly across observer cultures. The only significant effect was the vertical difference across face races. While the vertical positions to all faces were just below the eyes, fixations on Asian faces were slightly higher than fixations on Caucasian faces (0.78° vs. 1.36° below the eyes—i.e., 16.3% vs. 28.3% of the eye–nose distance; average difference = 0.58°, or 12% of eye–nose distance; Table 1). Despite small vertical differences between fixating Asian and Caucasian faces, the average first fixations of each observer were highly correlated across the two face races for both dimensions (Figure 2B).

Although no significant differences were found across observer cultures, individual differences on the first fixations were observed within a cultural group (see also Mehoudar, Arizpe, Baker, & Yovel, 2014; Peterson & Eckstein, 2013). Nevertheless, such variations were homogenous across observer cultures and face races (Levene’s test of homogeneity of variance: $F(3, 60) = 1.52$, $p = 0.22$ for the vertical dimension, $F(3, 60) = 0.395$, $p = 0.76$ for the horizontal dimension).

However, it is possible that potential differences in the first fixation patterns were not revealed by
The differences in the mean values (as in Figure 2) but in other properties of the fixation distributions (see fixation heat maps in Figure 3A), such as the standard deviation. In addition, the ANOVA test assumes parametric properties for the distributions of sample means. For these reasons, we chose to use the entire fixation distribution and test for potential differences in the first- and second-order moments (i.e., means and standard deviations of the distributions) across conditions using a nonparametric resampling technique known as the block bootstrapping method (see earlier under Statistical analyses of fixations). Asian and Caucasian observers did not differ significantly in their mean first fixations for either the horizontal or vertical dimension (Table 3). However, there were significant differences in mean fixations between Asian and Caucasian face stimuli for the vertical (but not the horizontal) dimension, with fixations significantly higher on Asian faces than on Caucasian faces (agreeing with the parametric ANOVA results). Beyond the distribution means, the standard deviations (Figure 3B) did not differ significantly except between face stimuli for Asian observers in the vertical dimension (Table 3).

Table 2. Mixed-model repeated measures ANOVA results on the first fixation positions (Experiment 1), analyzed separately for the vertical and horizontal dimensions. Notes: F: face race; O: observer culture. The significant p value (i.e., p < 0.05) is marked with an asterisk (*).
Experiment 2: Free eye movements (1500-ms presentation)

Introduction

While Experiment 1 shows similar first eye-movement strategies across observer cultures, the results could be specific to a short stimulus presentation time. In addition, investigation of later saccades was not possible, as only one fixation could be recorded within the 350-ms stimulus presentation. Thus, we used a longer presentation time in Experiment 2 (1500 ms) in order to examine the generalizability of the results on initial fixations to longer presentation times and also assess any cultural differences on fixation patterns in later saccades.

Methods

The experimental procedure was the same as in Experiment 1, except that the face stimulus was presented for 1500 ms before being masked (Figure 1). All observers from Experiment 1 participated in this experiment.

Results

The first fixation

Fixation patterns were examined using the methods described in Experiment 1. Analysis on the 25-trial running averages showed that the first fixation positions did not change significantly through time (Supplementary Table S2). Individual averages of the first fixation positions were analyzed using a mixed-model repeated measures ANOVA (Table 4). There was no significant cultural effect in the vertical dimension. The only significant vertical effect was the face race: Both observer cultures initially fixated higher on Asian faces (Asian vs. Caucasian faces: 1.09° vs. 1.46° below the eyes—i.e., 22.7% and 30.4% of eye–nose distance; average difference = 0.37°, or 8% of eye–nose distance). For the horizontal dimension, only the effect of observer culture just reached significance, where Caucasian observers fixated more leftward than Asian observers (0.893° vs. 0.375° to the left of the vertical midline—i.e., 12.9% vs. 5.4% of interocular distance; average difference = 0.518°, or 7% of interocular distance). Similar to Experiment 1, these parametric ANOVA results were largely consistent with those from nonparametric bootstrapping (Supplementary Table S3).

Comparisons between Experiments 1 and 2

Finally, we compared the first fixation positions (averaged by observer) across the two experiments. Figure 4 shows that the first fixation positions were highly correlated between the two presentation times (vertical dimension: $r = 0.863$, $p < 0.0001$; horizontal
Despite a slightly, but significantly, higher average first fixation position with shorter presentation times (average vertical difference $= 0.20^\circ$; ANOVA results in Supplementary Table S4), the first fixation positions were generally similar across stimulus presentation times. The second fixation positions, shown in Figure 5A, were more scattered than the first fixations, though they still largely concentrated on the eyes and nose region. Due to the multimodal nature of distributions of the second fixations, a method different from those used for the first fixations was implemented to compare the distributions of the second fixations across conditions.

**Table 3.** The 95% confidence intervals (CI) of the differences of bootstrapped means and standard deviations of the distributions of the first fixation positions from Experiment 1 (free eye movements, 350-ms face presentations). Notes: Signiﬁcant differences are marked with asterisks (*).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$F$</th>
<th>$df$</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vertical dimension</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main effect (within-subjects)</td>
<td>6.77</td>
<td>1</td>
<td>0.014*</td>
</tr>
<tr>
<td>Main effect (between-subjects)</td>
<td>1.58</td>
<td>1</td>
<td>0.22</td>
</tr>
<tr>
<td>Two-way interaction</td>
<td>0.57</td>
<td>1</td>
<td>0.46</td>
</tr>
<tr>
<td>Residual</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Horizontal dimension</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main effect (within-subjects)</td>
<td>0.20</td>
<td>1</td>
<td>0.66</td>
</tr>
<tr>
<td>Main effect (between-subjects)</td>
<td>4.45</td>
<td>1</td>
<td>0.043*</td>
</tr>
<tr>
<td>Two-way interaction</td>
<td>0.20</td>
<td>1</td>
<td>0.66</td>
</tr>
<tr>
<td>Residual</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Mixed-model repeated measures ANOVA results on the first fixation positions (Experiment 2), analyzed separately for the vertical and horizontal dimensions. Notes: F: face race; O: observer culture. Significant p values (i.e., $p < 0.05$) are marked with asterisks (*).
We counted fixation frequencies in distinct face areas covering the eyes, nose, and mouth, respectively, for each observer (definition of face areas in Figure 5B; Goldinger et al., 2009). Note that fixation frequencies for the left and right eyes were combined, as no significant interocular differences were found (Supplementary Table S5). We then computed the proportion of fixations in each face area by dividing the fixation frequency by the total number of second fixations for that observer. The proportions for fixating the eyes, nose, and mouth (Figure 5B) were subsequently subjected to three separate mixed-model repeated measures ANOVAs to compare the fixation distributions across observer cultures and face races (Table 5). Similar to the first fixations, the distributions of the second fixation positions were not significantly different between Asian and Caucasian observers. There were, however, significant differences across face races. Observers of both cultures directed their second fixations more frequently to the eyes than to the nose for Asian faces, and vice versa for Caucasian faces.

Similar to the second fixation, cultural effects were also absent in the third and subsequent saccades (there were 3.65 ± 0.57 fixations on average), and observers preferred the eyes when viewing Asian faces but preferred the nose when viewing Caucasian faces (see supplementary materials).

**Experiment 3: Famous-face familiarity**

**Introduction**

Although Experiment 1 showed that the first fixation positions did not differ across observer cultures, the results could be due to specific laboratory conditions, such as the noisy, monochromatic nature and the small sample size (10 Asian and 10 Caucasian) of our face stimuli. Thus in this experiment, we set out to examine the first fixation positions using a larger sample of more natural color face images in a familiarity task.

**Methods**

**Participants**

A subset of observers from Experiment 1 (six Asian, eight Caucasian; 14 in total) participated in this experiment.
Stimuli

Two hundred high-quality face images of celebrities were obtained from a Google image search, of which 100 were Asians and the other 100 were Caucasians. All faces were posed at approximately frontal views. These images were scaled, rotated, and translated in the same way as the database images used in Experiment 1, but here the color and background were retained and no noise was added. Thus hair, ears, and the face outline were displayed in addition to the eyes, nose, and mouth. The images were displayed at 100% contrast.

Procedure

Each observer performed two conditions in a random order, one on identifying 100 famous Asian faces and the other on identifying 100 famous Caucasian faces. The experimental procedure was the same as in Experiment 1, except that the famous faces were presented in a random sequence without repetition and observers were instructed to indicate whether the face images were familiar to them immediately after each face presentation. Each image was displayed for 350 ms.

Results

Figure 6A shows the first fixation positions to the famous faces. From the ANOVA (Table 6), no significant cultural effect was found in either dimension. The only significant effect was for face race in the horizontal dimension, with a slightly larger leftward bias to Caucasian faces than to Asian faces (0.41° vs. 0.13° to the left of the midline of the face—i.e., 5.9% vs. 1.9% of interocular distance). The results from block bootstrapping showed similar statistical results (Supplementary Table S8) as the ANOVA.
Comparisons between Experiments 1 and 3

Next, the first fixation positions found here were compared with those from Experiment 1. We found highly significant correlations (vertical dimension: \( r = 0.756, p < 0.0001 \); horizontal dimension: \( r = 0.533, p < 0.004 \)) between the first fixations of the two experiments (Figure 6B). Also, the first fixation positions had negligible differences across the two experiments, as the mixed-model repeated measures ANOVAs (Supplementary Table S9) show that the main effect of experiment and its associated interaction effects were all nonsignificant with one exception: the significant two-way interaction between experiment and face race in the vertical dimension. This interaction effect was likely due to the absence of a vertical face-race effect in Experiment 3, as opposed to slightly higher fixations to Asian faces in Experiment 1. However, the corresponding post hoc tests marginally failed to reach significance (Supplementary Table S9). Such discrepancy across experiments could be due to the use of a different set of face stimuli and a smaller sample size in Experiment 3. In general, these results indicate that the first fixation positions stay just below the eyes regardless of the differences of the face stimuli used in the two experiments. The consistency of the first fixation positions across observer cultures is not the result of the monochromatic, noisy nature of the test stimuli, nor restricted to the 10 Asian and 10 Caucasian face stimuli used in Experiment 1.

### Experiment 4: Restricted number of fixations

**Introduction**

Hsiao & Cottrell (2008) have shown that face-identification accuracy in a standard face learning-and-recognition task asymptotes after the second eye movement. We set out to investigate whether this result could be generalized to the 1-in-10 face-identification paradigm that we used (Figure 1). The contributions of the first, second, third, and subsequent eye movements to perceptual accuracy of face identification were evaluated. In addition, we included here a new condition where the stimulus presentation was extended to 5000 ms in order to examine any potential effects of observer culture on later saccades, as we failed to find such effects within 1500 ms of face presentations (Experiment 2). A new group of observers was recruited to ensure that our previous findings were not the result of a limited number of participants.

### Methods

The experimental procedure was largely similar to that in Experiment 1 (with the same sets of face stimuli), except that the face stimulus was masked immediately after the participant’s eyes moved away from the last permissible fixation (a total of either one, two, or three fixations were allowed; see Figure 7A) or was allowed to be presented for 5000 ms irrespective of the number of saccades (Figure 1). For each face race, the task consisted of 1,600 trials: 400 for each of the four stimulus presentation conditions (one fixation, two fixations, three fixations, 5000 ms). The conditions were randomized such that participants could not anticipate the duration of the stimulus presentation on any trial. The task was completed by 16 participants (eight Asian, eight Caucasian), six of whom (three Asian, three Caucasian) participated in the previous experiments and the remaining 10 of whom (five Asian, five Caucasian) were new, selected based on the same criteria as the original group of participants. To investigate the culture effect on fixation distributions over prolonged stimulus presentation, an additional group of 12 participants (six Asian, six Caucasian), who had no prior exposure to previous experiments, performed only the 5000-ms condition for 400 trials. Thus the 5000-ms condition was performed by 28 participants in total (14 Asian, 14 Caucasian), matching the numbers of participants in previous studies (Blais et al., 2008; Rodger, Kelly et al., 2010).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>F</th>
<th>df</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical dimension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main effect (within-subjects)</td>
<td>0.09</td>
<td>1</td>
<td>0.78</td>
</tr>
<tr>
<td>Main effect (between-subjects)</td>
<td>0.50</td>
<td>1</td>
<td>0.49</td>
</tr>
<tr>
<td>Two-way interaction</td>
<td>0.44</td>
<td>1</td>
<td>0.52</td>
</tr>
<tr>
<td>Residual</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal dimension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main effect (within-subjects)</td>
<td>11.21</td>
<td>1</td>
<td>0.006*</td>
</tr>
<tr>
<td>Main effect (between-subjects)</td>
<td>1.48</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>Two-way interaction</td>
<td>2.13</td>
<td>1</td>
<td>0.17</td>
</tr>
<tr>
<td>Residual</td>
<td>12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Mixed-model repeated measures ANOVA results on the first fixation positions (Experiment 3), analyzed separately for the vertical and horizontal dimensions. Notes: F: face race; O: observer culture. Significant p values (i.e., \( p < 0.05 \)) are marked with asterisks (*).
Results

Perceptual accuracy

To analyze how the number of fixations contributed to identification performance, we calculated the ratio of performance (in terms of proportion of correct responses) for each fixation condition (one, two, or three fixations allowed) relative to the performance in the 5000-ms condition for each of the 16 observers who performed all four fixation conditions (Figure 7B). Averaged across all conditions, viewing faces with one fixation (average stimulus duration: 578 ms) achieved 83.4% ± 1.4% ($p < 0.0001$) of the performance of 5000-ms face presentations. Viewing faces with two and three fixations (average stimulus durations: 986 and 1444 ms, respectively) resulted in, respectively, 90.8% ± 1.0% ($p < 0.0001$) and 93.8% ± 0.8% ($p < 0.0001$) of the 5000-ms performance. The durations that the eyes actually fixated on the face averaged 333, 741, 1201, and 4779 ms, respectively, for the four conditions. These results showed that the first fixation contributed markedly to face identification, and the second, third, and subsequent fixations played a considerably smaller role.

The first fixation

Similar to Experiment 1, we performed mixed-model repeated measures ANOVAs on individual running averages of the first fixation positions, separately for the 16 observers who performed all four stimulus presentation conditions (Supplementary Table S10) and the 12 observers who participated in the 5000-ms condition only (Supplementary Table S11). Again, we confirmed no significant changes in the first fixation positions across trials.

We subsequently averaged the first fixation positions over trials for each observer, face race, and stimulus presentation condition. A mixed-model repeated measures ANOVA (three main effects: stimulus presentation condition, face race, observer culture) was performed on these individual averages for each dimension, considering only the 16 observers who performed all four stimulus presentation conditions (Table 7). The cultural effects were not significant in either dimension. Replicating Experiments 1 and 2, the effect of face race was significant in the vertical dimension (fixations on Asian faces higher than on Caucasian faces) but not in the horizontal dimension. Importantly, the first fixation positions were consistent.
regardless of the duration of stimuli, as the main effect of stimulus presentation condition and its associated interaction effects were all nonsignificant. Thus we averaged the first fixation positions over the four stimulus presentation conditions for subsequent analyses (Figure 8). All these statistical results were supported by those from block bootstrapping (Supplementary Table S12).

Comparisons between Experiments 1 and 4

The fixation data of 22 observers in this experiment (after excluding the six observers who performed the previous experiments) were compared with the first fixation data from 32 observers in Experiment 1 by use of a mixed-model repeated measures ANOVA in each dimension (Table 8). Importantly, the results show that the first fixation positions from the two observer groups agreed well, with no significant differences arising from the main effect of participant group and its associated interaction effects.

Combining data from Experiments 1 and 4

Table 8 also shows the effects of face race and observer culture with this enlarged sample size of 54 observers by pooling the fixation data from the two groups, thus increasing markedly the power of the statistical tests. For both groups of participants, we confirmed that observer culture did not have any significant effect on the first fixation positions in the vertical dimension; neither did the iMap analysis (Caldara & Miellet, 2011) show any cultural differences. Interestingly, a small horizontal difference (0.59°, or 8.5% of interocular distance more leftward for Caucasian observers than for Asian observers) was found, despite no significant differences when analyzing the two participant groups separately (but note the significant differences found in Experiment 2). Regardless of observer group, the first fixations were slightly lower (mean difference: 0.46°, or 9.6% of eye–nose distance) when directed to Caucasian faces than to Asian faces (but with no horizontal difference). Figure 8 additionally shows that the overall distributions of the first fixations from this experiment (including all four stimulus presentation conditions) do not qualitatively differ from those from Experiment 1. As in Experiment 2 (stimulus duration: 1500 ms), cultural effects were absent in the second, third, and subsequent saccades (5000-ms condition: 10.89 ± 1.85 fixations in total; see supplementary materials).

Experiment 5: Forced-fixation tasks

Introduction

To investigate whether the first fixation has functional importance to identification accuracy, observers from Experiment 1 participated in an additional face-identification task in which they maintained fixation at one of five predetermined positions on the face during stimulus presentation (Figure 1, Forced-fixation condition). We assessed accuracy as a function of fixation position.

Methods

On completion of Experiments 1, 2, and 3, all 32 original participants performed 1,500 trials (10 sessions of 150 trials each) of a forced-fixation task (Figure 1). Each session began with the calibration and validation procedure outlined earlier under Eye tracking. To begin

<table>
<thead>
<tr>
<th>Parameter</th>
<th>F</th>
<th>df</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vertical dimension</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Main effects (within-subjects)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>1.50</td>
<td>1</td>
<td>0.24</td>
</tr>
<tr>
<td>F</td>
<td>9.86</td>
<td>1</td>
<td>0.007*</td>
</tr>
<tr>
<td><strong>Main effect (between-subjects)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>0.31</td>
<td>1</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Two-way interactions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T × O</td>
<td>0.76</td>
<td>1</td>
<td>0.40</td>
</tr>
<tr>
<td>T × F</td>
<td>0.14</td>
<td>1</td>
<td>0.72</td>
</tr>
<tr>
<td>O × F</td>
<td>1.13</td>
<td>1</td>
<td>0.31</td>
</tr>
<tr>
<td><strong>Three-way interaction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T × O × F</td>
<td>1.64</td>
<td>1</td>
<td>0.22</td>
</tr>
<tr>
<td>Residual</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Horizontal dimension</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Main effects (within-subjects)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>0.41</td>
<td>1</td>
<td>0.53</td>
</tr>
<tr>
<td>F</td>
<td>2.11</td>
<td>1</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>Main effect (between-subjects)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>0.91</td>
<td>1</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>Two-way interactions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T × O</td>
<td>0.28</td>
<td>1</td>
<td>0.61</td>
</tr>
<tr>
<td>T × F</td>
<td>1.21</td>
<td>1</td>
<td>0.29</td>
</tr>
<tr>
<td>O × F</td>
<td>0.01</td>
<td>1</td>
<td>0.92</td>
</tr>
<tr>
<td><strong>Three-way interaction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T × O × F</td>
<td>1.34</td>
<td>1</td>
<td>0.27</td>
</tr>
<tr>
<td>Residual</td>
<td>14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Mixed-model repeated measures ANOVA results on potential variations of the first fixation positions across stimulus presentation conditions (Experiment 4), analyzed separately for the vertical and horizontal dimensions. The analysis considered only the 16 observers who performed all four stimulus presentation conditions (T). Notes: F: face race; O: observer culture. Significant p values (i.e., p < 0.05) are marked with asterisks (*).
each trial, participants fixated a cross located randomly at one of five possible positions on the uniform gray screen. Four of the possible fixation positions (black broken squares in Figure 1) were common for all participants and were located along the vertical midline of the screen. Adjacent locations were 3.7° apart and corresponded to the forehead, the midway point between the eyes (i.e., the nose bridge), the nose tip, and the center of the mouth. The fifth position (illustrated by a red broken circle in Figure 1), which was different for each observer, was the mean location of the observer’s first fixation onto the face from Experiment 1 (we called it their preferred fixation). Following a random delay after a press of the space bar (as in Experiment 1), the fixation cross was replaced by a noisy, contrast-reduced face stimulus for 200 ms. During stimulus presentation, eye movements were prohibited such that participants were required to keep fixation on the position of the previously presented (and subsequently vanished) fixation cross (with the 1.1° tolerance as previously). Eye tracking during the forced-fixation study allowed for verification that observers were indeed fixating at the indicated points of fixation and were not moving their eyes during the presentation of the stimulus. If an eye movement was detected any time during stimulus presentation, the trial was aborted and restarted with a new stimulus. The rest of the trial was identical to that of a task with free eye movements (see Procedure in Experiment 1).

Results

Figure 9 shows face-identification accuracy as a function of fixation position for Asian observers (blue dots) and Caucasian observers (red triangles), plotted separately for the two face races. For each observer and face race, the accuracy for fixating at each of the four positions (forehead, eyes, nose, mouth) was subtracted from the accuracy for fixating at that observer’s preferred point of fixation. The resulting differences were averaged across observers and face races, then plotted as a function of the fixation position’s distance from the preferred fixation position, in the form of separate histograms for own-race faces and other-race faces (Figure 10). We found that face-identification accuracy decreased progressively as the distance between the fixation position and the preferred fixation...
position increased, regardless of whether observers viewed faces of their own race or the other race. These results indicate that first fixation at the preferred point of fixation is functionally important and optimal, leading to maximal identification performance irrespective of observer culture and face race.

Foveated ideal observer (FIO)

Introduction

What the empirical results do not explain is why humans from both cultural groups achieve maximal performance when looking below the eyes (with a small variation across faces of different races). We hypothesized that the functionally important initial fixation reflects the brain’s ability to learn eye movements that maximize perceptual performance by considering (a) the spatial distribution of information in faces that is critical for identification and (b) the fact that visual information is limited by a decline in resolution and sensitivity from the fovea to the periphery. The interaction of these two factors results in a decrease of available information further from one’s fixation point. Constrained by the varying spatial resolution of the human visual system, an FIO (Peterson & Eckstein, 2012) is a Bayesian rational model that generates predictions of face-identification performance for any given set of fixation locations (Figure 11). Here we evaluated whether the FIO predicted face-identification performance for both Asian and Caucasian observers when they fixated at various positions on the faces in Experiment 5.

Model overview

An FIO (Peterson & Eckstein, 2012; Figure 11) is a modified version of the ideal observer (Geisler, 2011; Kersten, Mamassian, & Yuille, 2004; Sekuler, Gaspar, Gold, & Bennett, 2004) that makes statistically optimal identification decisions considering the visual system’s spatially varying contrast sensitivity across the visual field. The fovea is a small region of the retina that corresponds to the point of fixation, generally the center of the visual field. Dense photoreceptor sampling and a preferential allotment of cortical resources enable visual stimulation of this region to be processed with high resolution and sensitivity. Outside the fovea, the visual information available to the brain is degraded, especially at higher spatial frequencies. This relationship between visual-information quality and distance from fixation in the visual field (eccentricity) can be quantified through the contrast sensitivity function (CSF). The CSF specifies the amount of contrast required for a signal to be reliably detected as a function of its spatial frequency. This inherent quality of the visual system imposes a constraint on the type and amount of visual information the brain can use when forming perceptual decisions, and this modulation is a reliable consequence of where on the face one looks. This modulation is what the FIO attempts to model.

Mechanistically, for any possible point of fixation the noisy stimulus image can be divided into regions defined by their distance and direction from the center of gaze, with each region assigned a CSF. The contrast of each spatial-frequency component is attenuated in accordance with the CSF through filtering in the Fourier domain. This filtered noisy image is compared to similarly filtered noise-free versions of each of the possible faces through cross-correlation. This results in 10 response variables that constitute the input to a Bayes optimal decision mechanism (i.e., an ideal...
The ideal observer possesses perfect information about the original face images (pixel-by-pixel grayscale values), the statistical properties of the additive noise (zero mean, spatially independent, normally distributed with a known variance), and the modulatory effects of the CSF. The ideal observer can thus calculate the statistical properties of the 10 response variables for each possible scenario (i.e., for each face stimulus). For each possible scenario, the ideal observer calculates how likely it would be to observe the computed responses if the stimulus was truly that face, and selects the most likely option as its identification decision. With perfect information of the response variables’ statistical properties, as here, this ensures optimal performance.

FIO accuracy is estimated for each possible fixation, with the level of accuracy directly corresponding to the amount of task-relevant information available. Thus, the FIO gives an objective measure of how much information is retained given the necessary functional effects of foveation. A rational system specifically optimized for this task would be expected to exhibit an initial fixation to the faces consistent with maximizing task performance. This is the evidence that the FIO seeks to evaluate. A more detailed description of the FIO follows.

**Model construction**

We simulated the spatial variation in contrast sensitivity using a canonical function (Peli, Yang, & Goldstein, 1991):

$$\text{CSF}(f, r) = c_0 f^{a_0} \exp(-b_0 f - d_0^{p_0} f),$$

where $f$ is the spatial frequency (in cycles per degree of visual angle) and $r$ is the distance (in degrees of visual angle) from the center of the fovea. The constants $a_0$, $b_0$, and $c_0$ were chosen to set the maximum nominal contrast at 1 and the peak spatial frequency at the...
fovea at 4 c/°, and $d_0$ is an eccentricity factor that determines the steepness of the decline in sensitivity with eccentricity. This eccentricity factor is dependent on the direction $\theta$ from the fovea (in polar angle) and $n_0$ is the steep roll-off factor, which further modulates the eccentricity effects.

For any given fixation point $k$, the image was divided into small spatial bins extending radially outward (Figure 11). Each bin was assigned a CSF according to its distance $r$ from fixation and its polar angle $\theta$. Going bin by bin, the original, unfoveated image was filtered with the current CSF. Then the current bin’s spatial region was extracted and placed into the corresponding area of the final, composite foveated image (Figure 11). In the work presented here, we used 480 spatial bins (30 eccentricity ranges and 16 angles), as finer resolutions did not significantly affect the results.

The FIO used the non-prewhitening matched filter with an eye filter (NPWE) observer model to form decision variables (Burgess, 1994; Eckstein & Abbey, 2001; Y. Zhang, Pham, & Eckstein, 2004). The unfoveated noisy stimulus was an additive combination of a known signal (the face image, denoted $s_0$) and a zero-mean, Gaussian white-noise process (denoted $n_{ex}$). The process of simulated foveation spatially correlated the noise. Instead of correcting for these correlations, the NPWE foveated the possible signals in the same way as the stimulus, and used these matched templates to form decision variables.

We denote the family of CSFs for a given fixation point $k$ with the linear operator $E_k$. The foveated noisy stimulus, then, is $\tilde{g} = E_k(s_0 + n_{ex}) + n_{in}$, where $n_{in}$ is an additive white Gaussian internal noise process. This noisy template was cross-correlated to each possible signal (i.e., the 10 faces for each race condition), which resulted in a vector of matched-template responses $r_k = \{r_{l,k}, \ldots, r_{10,k}\}$, where $r_{l,k} = (E_k s_0)^T(E_k(s_0 + n_{ex}) + n_{in})$. The FIO then computed posterior probabilities for each possible state of the world. Here, the possibilities correspond to hypotheses $H$ about the identity of the underlying face image $s_0$. That is, for a given race condition there are 10 possible states of the world, $H_1$ through $H_{10}$. The FIO’s decision rule is

$$\text{decision} = \arg \max_f \left( P(H_f | r_k) \right)$$

$$= \arg \max_f \left( \frac{P(H_f) P(r_k | H_f)}{P(r_k)} \right). \quad (2)$$

In this case, each hypothesis is equally likely to occur, allowing us to discard the prior probability $P(H_f)$. Similarly, the probability $P(r_k)$ of observing the data is common to all hypotheses and can be removed. This leaves the likelihood $P(r_k | H_f) = l_{i\beta}$ as the decision variable.

To compute the likelihoods, it is necessary to determine the underlying distributions that generate the response vectors for any given state of the world. It has been shown that the linearity of the foveation operator and the characteristics of Gaussian white noise led to the response vectors being drawn from fully described multivariate normal distributions (Peterson & Eckstein,
The mean response of face $i$ when face $f$ is the true state of the world is $\mu_{i,f,k} = s_j^T s_{i,k}$, where $s_{i,k}$ is face $i$ filtered twice by the eye filter $E_k$ and $T$ is the transpose operator. The double filtering is a mathematical consequence of cross-correlating two single-filtered images (see Peterson & Eckstein, 2012). The covariance between the $i$th and $j$th responses under the hypothesis that face $f$ is present can be shown to be $\Sigma_{i,j,k} = \sigma_{ex}^2 s_j^T + \sigma_{in}^2 s_{i,k}$. The parameters $\sigma_{ex}^2$ and $\sigma_{in}^2$ are the external and internal noise variances, respectively, while $s_{i,k}$ is face $i$ filtered once by eye filter $E_k$. Here, the external noise variance for the FIO's performance estimation was set to be identical to the variance of the white noise added to the experimental stimuli presented to the human observers. Note that the covariance does not depend on the state of the world.

With the distribution parameters known to the FIO, the likelihood was computed as

$$l_{f,k} = \exp\left(-\frac{1}{2}(r_k - \mu_{f,k})^T \Sigma_k^{-1}(r_k - \mu_{f,k})\right).$$

(3)

The proportion of correct responses at each fixation point is then the average probability that the likelihood for the presented face is greater than that for each other face.

**Model fitting**

The FIO and the associated CSF (Equation 1) are defined by five free parameters. The eccentricity factor $d_0$ is allowed to vary as a function of direction from fixation, an implementation of the well-known eccentricity asymmetries. Here, $d_0$ is divided into three separate variables, one each for the polar angles corresponding to the up direction ($du = d_0(\frac{\pi}{2})$), the down direction ($dd = d_0(-\frac{\pi}{2})$), and the horizontal direction ($dh = d_0(0)$), assuming symmetry across the vertical meridian. Linear interpolation is used to compute the eccentricity factor for any given angle. The next free parameter is the steep roll-off factor $n_0$, which is assumed to be constant in all directions. Finally, the variance $\sigma_{in}^2$ of the internal noise is used to modulate the overall performance.

Four of the five parameter values ($du = 0.0001$, $dd = 0.00024$, $dh = 0.00005$, $n_0 = 5$) were used based on fitting forced-fixation performance profiles from a separate group of observers in a previous study (Peterson & Eckstein, 2012). The four parameter values were used (and left unchanged) for all FIO simulations in the current article. The exception was the internal noise variance parameter $\sigma_{in}^2$, which was allowed to vary to match the overall group performance separately for the four observer–stimulus conditions. The fitted value of the internal noise in each condition was estimated by a weighted least-squares method. It should be noted that fitting the empirical data using different internal noise variances resulted in small variations of the FIO’s optimal fixation positions ($M \pm SEM$ below the eyes): Asian faces: $0.30^\circ \pm 0.05^\circ$; Caucasian faces: $1.24^\circ \pm 0.06^\circ$. Interestingly, this result mirrored our empirical finding that fixations on Caucasian faces were higher than those on facial features.

**Natural systems analysis**

To evaluate whether the common optimal fixation areas can be generalized to larger populations of Asian and Caucasian faces, we fitted the FIO model to a larger sample of faces. This large-scale analysis is known as natural systems analysis (NSA; Geisler, 2008). The larger sample consisted of 100 Asian and 100 Caucasian faces (half of which were male; see supplementary materials for justifications of using 100 faces), including the 10 Asian and 10 Caucasian faces presented to the observers. The Asian faces came from two different databases (the University of Hong Kong Face Database, Visual Cognition Laboratory, the University of Hong Kong, Hong Kong, and Asian Face Image Database PF01, Intelligent Multimedia Laboratory, Pohang University of Science and Technology, South Korea). The Caucasian faces were all taken in-house. Image processing followed the methods used for the psycho-physical stimuli (see Stimuli under General methods). For each face race, the 100 faces were clustered randomly into 10 groups of 10 faces. Each group was composed of 10 distinct faces of the same sex. An FIO fitting was performed on each of these face groups, using the same model parameters as in fitting the human data. In particular, the internal noise variance was chosen as the average of the fitted values across the four empirical observer–stimulus conditions.

**Results**

Figure 9 shows the FIO’s identification performance along the vertical midline (in solid lines) fitted to human identification accuracies for fixating at five different positions of the face in Experiment 5 (in data points). We assessed the goodness of fit of the model in the form of the coefficient of determination $R^2$. For Asian faces, $R^2 = 66.5\%$ and $65.3\%$ for Asian and Caucasian observers, respectively. The corresponding coefficients for Caucasian faces are $93.7\%$ and $86.0\%$. These values indicate that the model explains the majority of variance of human performance, though the fits are better for Caucasian faces, as the FIO parameters, other than the internal noise, were originally fitted on Caucasian faces. Figure 12 shows the NSA’s identification performance, separately for
Asian and Caucasian faces, for all potential fixation points (i.e., all pixels in practice) across the entire face. Table 9 shows statistical comparisons between the model’s prediction of the optimal position and the empirical fixation position (averaged across Experiments 1, 2, and 4 considering all 54 observers) in the form of $t$-tests. For the vertical dimension, the theoretically optimal positions predicted by the NSA were highly consistent with human data, as the model correctly predicted optimal fixations to a common area encompassing the eyes and the nose (the area shaded in red in Figure 12B) for both Asian and Caucasian faces. Interestingly, the model even correctly predicted slightly higher optimal fixation positions on Asian faces than on Caucasian faces ($M \pm SEM$ fixation position below the eyes: Asian faces: $0.44^\circ \pm 0.08^\circ$; Caucasian faces: $0.98^\circ \pm 0.15^\circ$; see Discussion). For the horizontal dimension, the theoretical and empirical horizontal positions agreed with each other only for Asian observers, both of which were close to the vertical midline (model: $0.01^\circ$ to the left of midline, Asians: $0.23^\circ$), but not for Caucasian observers, with a leftward bias of $0.79^\circ$ (Table 9). Yet the FIO predicts that the $0.79^\circ$ bias should only lead to a small performance loss ($0.72\%$) from maximum theoretical identification accuracy. This would suggest that the small horizontal bias represents a minor source of suboptimality in encoding visual information for face identification. Together, the results indicate that individuals, both Asian and Caucasian, generally optimized their first eye movements depending on the race of faces to be discriminated.

**Discussion**

**Similar initial fixations to faces across cultural groups**

There is an ongoing debate about the extent of cultural influences on eye movements to faces. In the current work, we focused on assessing the effect of culture and face race on initial task-related eye movements during face identification. Our findings show that Asians and Caucasians use similar initial eye-movement strategies for face identification, at least for identifying Asian faces among other Asian faces and Caucasian faces among other Caucasian faces. Both
groups fixated a featureless point just below the eyes. Table 10 summarizes the statistical comparisons of the first fixation positions across observer cultures and face races in Experiments 1, 2, 3, and 4 and the associated effect sizes (from all 54 observers who performed part or all of Experiments 1, 2, and 4). Together, our results showed no consistent cultural effects on the first fixation positions in the vertical dimension. For the horizontal dimension, we observed a leftward bias (0.79 \pm 0.8\% of interocular distance) for Caucasian observers after averaging across Experiments 1, 2, and 4 considering all 54 observers. The result agrees with that of Hsiao and Cottrell (2008), who also showed a leftward bias (0.24 \pm 0.7\% of interocular distance of 3.7\%) for Caucasian observers directing their initial fixations to Caucasian faces (see also Bindemann, Scheepers, & Burton, 2009; Butler et al., 2005; Everdell, Marsh, Yurick, Munhall, & Pare, 2007; Leonards & Scott-Samuel, 2005; Mertens, Siegmund, & Grusser, 1993).

Yet the leftward bias (0.23 \pm 0.8\%) was smaller for Asian observers, leading to a cultural difference of 0.56\% (8.1\% of interocular distance) in the horizontal dimension when averaged across experiments. Given such a small magnitude, separate analyses among experiments did not consistently reach significance in the difference (Table 10), except for Experiment 2 (Table 5; Supplementary Table S2) and for combined data from Experiments 1 and 4 (Table 8). Importantly, the horizontal cultural difference found in our study was (a) smaller than the variability within each cultural group (Table 10; effect size, Cohen’s d = 0.7) and (b) approximately one-sixth to one-fourth of the size of the effect (in percentage of interocular distance) found in previous studies for aggregate fixations within stimulus presentations of 1.5 s or longer (see Table 11 for a summary of the magnitudes of the cultural effects and the methodological details in the current and previous studies). These comparisons suggest a rather small horizontal cultural effect in the first fixation. Furthermore, our analysis of the second and subsequent

<table>
<thead>
<tr>
<th>Observer culture</th>
<th>Face race</th>
<th>t</th>
<th>df</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vertical dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian observers</td>
<td>Asian faces</td>
<td>0.73</td>
<td>35</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Caucasian faces</td>
<td>0.28</td>
<td>35</td>
<td>0.78</td>
</tr>
<tr>
<td>Caucasian observers</td>
<td>Asian faces</td>
<td>1.12</td>
<td>35</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Caucasian faces</td>
<td>0.69</td>
<td>35</td>
<td>0.49</td>
</tr>
<tr>
<td><strong>Horizontal dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian observers</td>
<td>Asian faces</td>
<td>−0.40</td>
<td>35</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Caucasian faces</td>
<td>−1.16</td>
<td>35</td>
<td>0.26</td>
</tr>
<tr>
<td>Caucasian observers</td>
<td>Asian faces</td>
<td>−2.89</td>
<td>35</td>
<td>0.007*</td>
</tr>
<tr>
<td></td>
<td>Caucasian faces</td>
<td>−4.05</td>
<td>35</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

Table 9. Separate t-test results on potential variations of the first fixation positions, averaged across Experiments 1, 2, and 4, from corresponding optimal fixation positions estimated by the FIO. Separate statistical tests were conducted for the vertical and horizontal dimensions. Notes: Significant p values (i.e., p < 0.05) are marked with asterisks (*).

Table 10. Comparisons of the first fixation positions in Experiments 1–4 using two statistical methods: mixed-model repeated measures ANOVA and block bootstrapping (Boot). The statistical results from Experiment 4 only consider 16 observers who performed all four stimulus presentation conditions (see text). Notes: Statistical significance (Sig.) is determined by p < 0.05 in ANOVAs, and 95% confidence interval not including 0 in bootstrapping. The Cohen’s d values consider all 54 observers in Experiments 1, 2, and 4. A: Asian. C: Caucasian. NS: Not significant.
fixations did not reveal any statistically significant differences across cultures in the distribution of fixations across the eye, nose, and mouth regions. Similarly, no cultural differences were found when looking at all fixations together. Together, our results suggest similar fixation patterns across Asian and Caucasian observers.

Our results do not agree with previous data showing major cultural differences in eye-movement patterns to faces with extended viewing times of 1.5–5 s (Blais et al., 2008; Kelly et al., 2010; Table 11) but are consistent with preliminary data (Rodger, Blais, & Caldara, 2010) showing that the first fixation is common across cultures during learning and recognition of faces. One possibility for why our results differ from previous findings arises from the differences in experimental parameters between our Experiment 1 and other studies, such as the shorter display time of our stimuli and the lack of color in the face images. However, we presented additional studies for which observers displayed similar initial fixations when asked to discriminate faces with a longer presentation time (1500 ms and 5000 ms in Experiments 2 and 4), and to determine the familiarity of full-color, noise-free faces of famous people (Experiment 3). In these additional conditions, there were again consistently no significant cultural differences in the vertical dimension, with occasionally small horizontal differences. Another possibility might be that the same 10 Asian faces and 10 Caucasian faces were repeatedly presented a large number of times (in Experiments 1, 2, and 4) such that observers might have learned eye-movement strategies specific to these faces. However, we have shown that the positions of the first fixation did not significantly change across trials. In addition, observers had similar fixation patterns with a larger sample of faces (100 Asian, 100 Caucasian) in Experiment 3. Thus our findings of similar initial fixation patterns for Asians and Caucasians likely generalize to more real-world situations.

The general lack of large cultural differences in fixation patterns reported here should not be attributed to an insufficient number of participants or trials in our study. The common fixation distributions were replicated by two separate groups of participants in a series of free-eye-movement tasks (Experiments 1, 2, 3, and 4), totaling 54 participants (27 Asian, 27 Caucasian) who each performed 400–500 trials. In contrast, only 14 or 21 participants per cultural group were recruited, with no more than 84 trials per participant, in earlier studies (Blais et al., 2008; Kelly et al., 2010; Table 11) that found significant cultural differences on fixation distributions.

### Why are the first fixations during face identification similar across cultural groups?

Our results suggest that the critical first eye movement for face identification is more driven by a neural system aimed at maximizing the acquisition of visual information to support successful face identification rather than by cultural influences. We demonstrated that the first fixation to faces played a major role in face identification (Experiment 4; consistent with the result from a learning-and-recognition task on faces by Hsiao & Cottrell, 2008), and after the third fixation, accuracy reached 93.8% of the asymptotic accuracy with 1444-ms face presentations. The forced-fixation study (Experiment 5) confirmed that for individuals from both cultural groups, the position of the first fixation has an important functional role in

---

**Table 11. Cross-study comparisons of fixation differences across observer cultures.** The cultural difference in the current study is the cross-cultural horizontal difference of the mean first fixation positions considering all observers in Experiments 1, 2, and 4 (see text). For previous studies (all using the learning-and-recognition tasks), the cultural difference is estimated by the vertical difference of the positions with the most positive and the most negative fixation biases (which happened to be at the eyes and the nose regions, respectively) in the difference Map (Caucasian Map − Asian Map) considering all fixations (Blais et al., 2008, figure 2; Kelly et al., 2010, figure 2). Notes: N observers: number of observers. A: Asian observers. C: Caucasian observers. N trials: number of trials per face race. N images: number of images per face race. Δx eyes: mean interocular distance of faces. Δy eye–nose: mean eye–nose distance of faces.
perceptual accuracy. Accuracy deteriorated when observers from both cultural groups were forced to identify faces (regardless of race) while fixating away from their preferred points of fixation. The empirically measured preferred and optimal points of fixation were not simple central points of the image and/or display ($p < 0.001$ in all $t$-tests on the differences), which could be consistent with a general central bias (Tatler, 2007). Instead, the preferred and optimal points of fixation broadly agreed with an FIO (with minor suboptimality for Caucasian observers in the horizontal dimension, see Results under FIO). These findings suggest that the preferred first point of fixation is mainly determined by the distribution of information in faces and its interaction with foveated visual processing. Cultural influences in these early fixations might be constrained by the brain’s aim at maximizing perceptual performance in face identification.

### Variations of the first fixations across face races explained by the FIO

We found small but consistent vertical differences in observers’ first points of fixations across face races (except in Experiment 3, which showed horizontal, instead of vertical, differences). If individuals of each cultural group overgeneralized the optimal strategy for faces of their own race to faces of the other race, then we might have found a small difference in the first fixations across cultural groups. However, we did not find evidence for such a hypothesis. Instead, our data suggest that observers of both cultural groups adapted their strategies to each face race and looked at a slightly higher position on Asian faces. Notably, the FIO model correctly predicted a slightly higher fixation to Asian faces as compared to Caucasian faces.

### Variations of the second and subsequent fixations across face races

Analysis of the second, the third, and all fixations (3.65 and 10.89 fixations for 1500-ms and 5000-ms face presentations, respectively; see also Supplementary Figure S4) found similar results to the first fixation: similar spatial distributions of fixations (eye, nose, and mouth regions) across observer cultures but different across face races (Figure 5; Supplementary Figures S2, S3, and S5). For both cultural groups, more fixations were devoted to the eye regions for Asian faces but the nose region for Caucasian faces. This continued with the trend of lower fixations to Caucasian faces than to Asian faces observed in the first fixation. Our results are not consistent with previous face learning-and-recognition studies suggesting that both Asian and Caucasian observers slightly (by less than 5% of total fixations) preferred fixating the eyes to the nose when presented with faces of their own race (regardless of Asian or Caucasian faces) but preferred the nose to the eyes with other-race faces (Fu et al., 2012; Goldinger et al., 2009; Wu et al., 2012). Instead, our data suggest that observers’ fixation strategy is associated with the face race itself, rather than whether the face belongs to their own culture. Nevertheless, the differences in fixation biases across face race were attenuated with prolonged stimulus presentations of 5000 ms (Supplementary Figure S5), possibly due to other face-perception functions taking over face identification in directing eye movements.

### Variability across individuals in preferred and optimal points of fixation

There were differences across individuals in the preferred points of fixation, but such individual differences were similar for the two cultural groups. While most observers initially fixated closer to the eye regions, a few observers from both cultural groups initially fixated towards the nose region (Peterson & Eckstein, 2013), regardless of the race of faces presented. Despite variations among observers, the consistency within individuals across image sets (Figures 2 and 6) suggests that face-looking strategies that support evolutionarily important perceptual tasks, such as face identification, are tuned to the individual rather than a cultural group. Thus, greater variability was shown among people within a cultural group than between groups (see Experiment 1, Results). Our forced-fixation study, which measured identification accuracy as a function of point of fixation (including observer-specific preferred point of the first fixation), showed that the preferred points of fixation reflected individual-specific optimal points of fixation (Figures 9 and 10; see also Peterson & Eckstein, 2013). Together, our results are consistent with the hypothesis that face recognition is optimized based on specific characteristics of one’s own foveated system.

### Possible influences of culture on eye movements to faces

Do our results imply that culture can never have large influences on the initial, critical eye movement to faces? Not necessarily. Our findings suggest that the initial eye-movement behavior can be explained as an optimization strategy for a system defined by a stimulus (the human face) with a specific distribution of visual information and a foveated sensory input (Najemnik & Geisler, 2005, 2008). This distribution of information...
might constrain early cultural effects on eye movements.

Nevertheless, our study focused on assessing the relationship between the initial eye movement and face identification. Subsequent fixations in prolonged face viewing could still be influenced by culture-specific norms, such as the socially acceptable duration of eye contact that might guide eye movements differently across cultural groups (Senju et al., 2013). In addition, for perceptual tasks with consistently large differences between cultures in the distributions of information, the FIO might predict divergent eye-movement strategies for the maximization of performance. For example, culture has been proposed to exert some influence over the facial features and facial movements that represent emotional states (Jack, Garrod, Yu, Caldara, & Schyns, 2012; Khooshabeh, Gratch, Haung, & Tao, 2010). Our theory suggests that if different expressions of emotions across cultures (Jack et al., 2012) lead to varying theoretically optimal points of fixation, variation of the first eye movements to faces across cultural groups should be observed in a face emotion-recognition task. Still, this influence of culture would be conceptually different from that conceived by previous researchers who addressed the culture would be conceptually different from that conceived by previous researchers who addressed the influence over the facial features and facial movements that represent emotional states (Jack, Garrod, Yu, Caldara, & Schyns, 2012; Khooshabeh, Gratch, Haung, & Tao, 2010). Our theory suggests that if different expressions of emotions across cultures (Jack et al., 2012) lead to varying theoretically optimal points of fixation, variation of the first eye movements to faces across cultural groups should be observed in a face emotion-recognition task. Still, this influence of culture would be conceptually different from that conceived by previous researchers who addressed the issue using more general cultural traits, such as individualistic and collectivistic cultures. In our framework, eye-movement differences would be more accurately described as optimal adaptations to culture-specific stimulus information.

**Keywords:** face perception, face identification, other-race effect, eye movement, culture

## Acknowledgments

This work was supported by Grant EY-015925 from the National Institutes of Health to MPE. CCFO was supported by the Postdoctoral Researcher Fellowship from the National Fund for Scientific Research (F.R.S.-FNRS, Belgium). The Asian-face databases were provided by William Hayward of the University of Hong Kong and Intelligent Multimedia Laboratory at Pohang University of Science and Technology. We thank James Berges, Benjamin Chen, Deborah Goldschmidt, Tabitha Harikul, Dayna Isaacs, and Amy Marton for assistance in conducting experiments.

Commercial relationships: none.

Corresponding author: Charles C.-F. Or.

Email: charles.or@uclouvain.be.

Address: Psychological Sciences Research Institute & Institute of Neuroscience, University of Louvain, Louvain-la-Neuve, Belgium.

## References


