

Supplemental Information

Culture shapes 7-month-olds' perceptual strategies in discriminating facial expressions of emotion.

Elena Geangu, Hiroko Ichikawa, Junpeng Lao, So Kanazawa, Masami K. Yamaguchi, Roberto Caldara, Chiara Turati

Supplemental Data

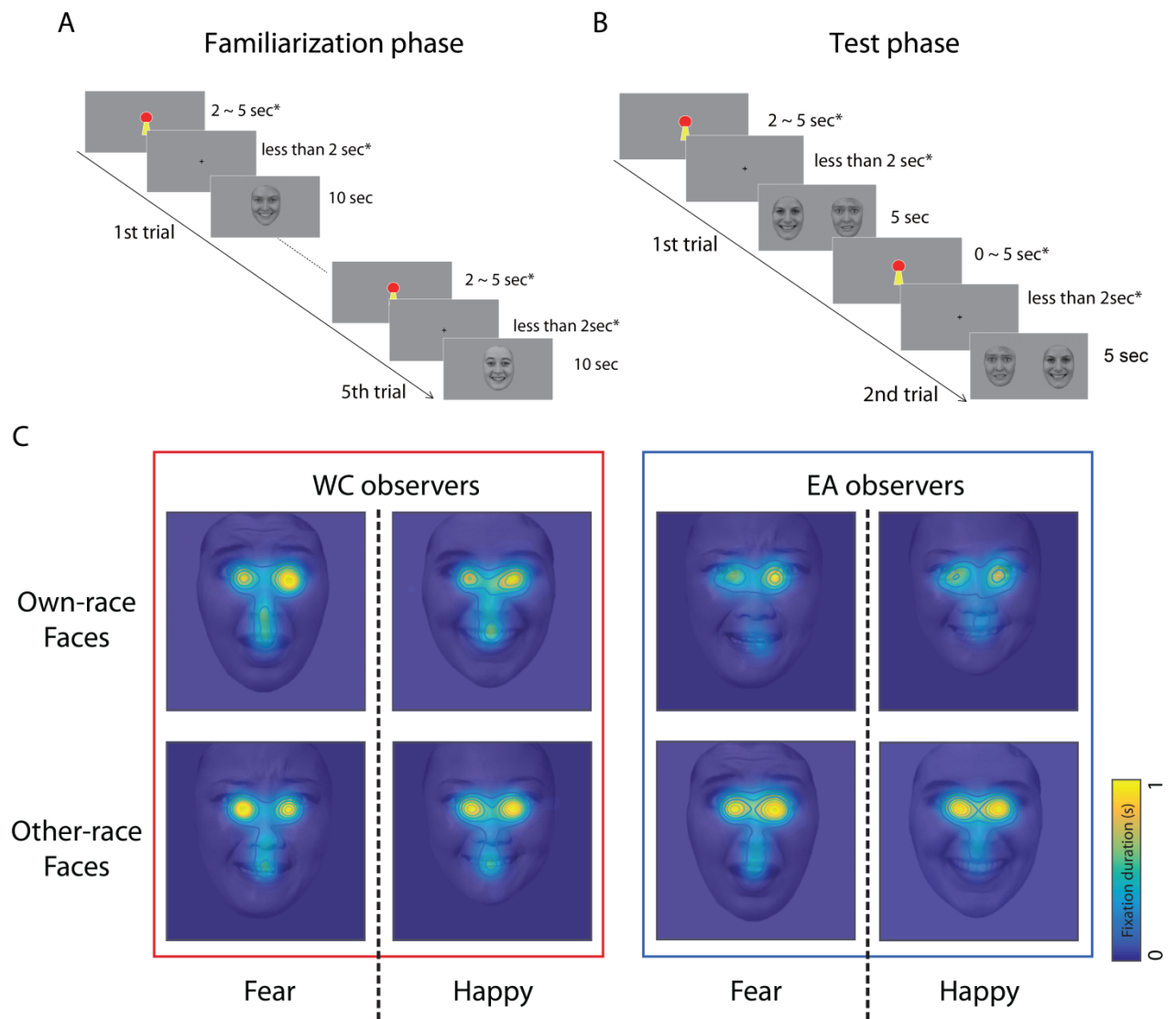


Figure S1 (related to Figure 1): Experimental paradigm (A & B) and eye movement analyses (C).

(A) The experimental procedure of *familiarization* phase (KDEF stimulus ID: AF28HAS, AF14HAS).

(B) The experimental procedure of *test* phase (KDEF stimulus ID: AF07HAS, AF22AFS).

(C) Descriptive iMap4 output of the spatial fixation mapping of the *familiarization* phase. Color maps show the mean fixation duration map for the 2x2x2 design.

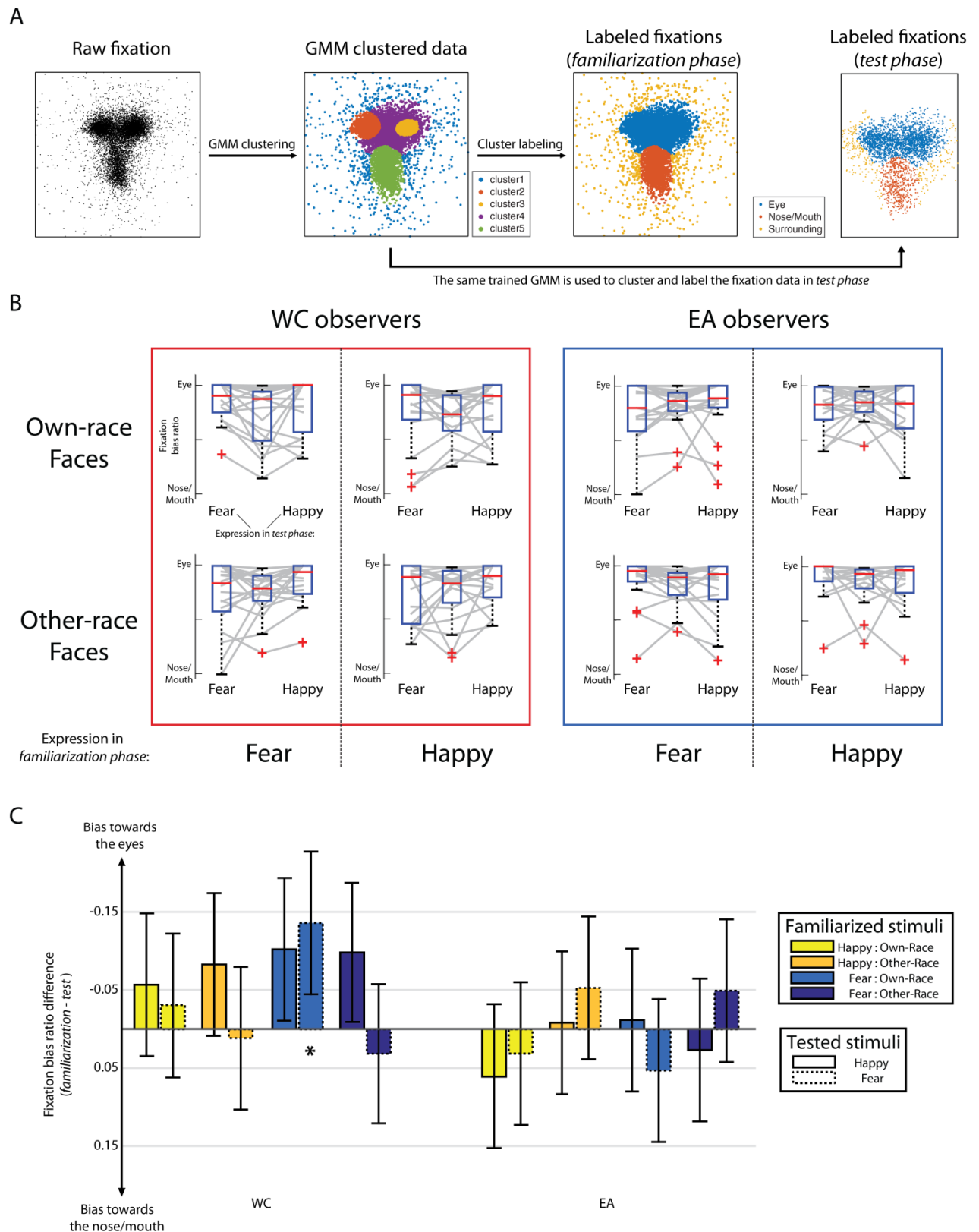


Figure S2 (related to Figure 1): Fixation pattern changes between *familiarization* and *test* phase.

(A) Unsupervised spatial clustering using Gaussian mixture model (GMM) on both *familiarization* and *test* phase fixation data. We trained the GMM on the data from *familiarization* phase based on the spatial mapping results using *iMap4*. Fixation clusters correspondent to the eye region and nose/mouth region were visually identified and labeled. The labeled GMM was then used to fit the fixations from *test* phase to cluster them into the eye region and the nose/mouth region.

(B) Individual line plot and boxplot of the fixation bias ratio computed on the fixation clusters from the GMM. The eight categorical conditions from our 2x2x2 design are displayed separately in different subplots. The central box bar indicates the fixation bias ratio in the *familiarization* phase; whereas the box bar beside shows the two expressions each infant participant saw during the *test* phase.

(C) Bar plot of fixation bias ratio difference between *familiarization* and *test* phase. The negative value represented an eye bias in the *test* phase; whereas a positive value represented a nose/mouth bias. Statistical significance is indicated by the * sign. Error bars report 95% CI.

Supplemental Experimental Procedures

Participants

A total of 153 7-month-old infants ($M_{\text{age}} = 226$ days, $SD = 8.23$ days) from both cultures (WC: $N = 77$; $M_{\text{age}} = 225$ days, $SD = 8.78$ days; EA: $N = 76$; $M_{\text{age}} = 226$ days; $SD = 7.62$ days) were included in the final analysis. An additional 41 infants were tested, but the data was not included in the analysis due to: fussiness (8 WCs, 5 EAs), eye-tracker calibration failure (6 WCs, 9 EAs), failure to attend both stimuli in each of the two test trials (7 WCs, 6 EAs). The WC infants were recruited from a North-West UK urban area, while the EA infants were from a metropolitan area of Japan. This study was approved by the University Ethical Committees in both countries. Participants' parents gave written informed consent before testing. The experiment was conducted according to the Declaration of Helsinki.

Stimuli

The face stimuli were extracted from the Karolinska Directed Emotional Faces database (KDEF; [S1]) and the Japanese Female Facial Expression Database (JAFFE, [S2]). A total of 28 grayscale images of female faces were used, including 7 unique identities from each race. Different face identities were presented during the *familiarization* and *test* phase. An equal number of face race (WC or EA) and facial expressions (HAPPY or FEAR) were displayed. All faces have been validated for their emotional content and intensity to ensure they had the highest intensity for the target emotion and the lowest intensity for other emotions [S3]. The inclusion of only two emotional expressions was motivated by the specific constraints of the age of the population tested here (i.e., duration of the experiment, paradigm characteristics). During the *familiarization* phase, the stimuli ($\sim 22^\circ \times 28^\circ$ of visual angle) were presented centred and aligned at the level of the eyes and face vertical midline (i.e., nose top and centre of the lips) (Figure S1A). For the *test* phase, the two face images ($\sim 20^\circ \times 28.8^\circ$ of visual angle) were presented simultaneously, aligned at the level of the eyes and the centre of lips (Figure S1B).

Apparatus

All stimuli were displayed on a 21.5-inch wide LCD monitor (1920 x 1080 pixels), while participants' gaze was recorded binocularly with a Tobii X120 eye-tracker (Tobii Technology, Sweden) at 60Hz, with .5 degrees accuracy and a drift of $< .3$ degrees. An infant-specific 5-points calibration procedure was used. Infants sat on their parents' lap, ~ 60 cm away front of the monitor.

Preferential looking task

The current task consisted of two consecutive parts: a *familiarization* phase followed by visual-paired comparison *test* phase. Infants were first familiarized to one emotional (either happy or fear) facial expression – *familiarization* phase. This was followed by the simultaneous presentation of the familiarized expression paired with a novel facial expression in order to test for the visual discrimination of the two expressions – *test* phase. We applied a between-subject design. Each infant was presented with either WC or EA face identities (*Face Race*), and familiarized to either fearful or happy facial expressions. In the *test* phase, both fearful and happy expressions were shown simultaneously, keeping the face race consistent with the *familiarization* phase.

Familiarization phase. Each familiarization trial lasted 10 seconds and was preceded by the presentation of an animated audio-visual non-social attention grabber in the center of the screen, then followed by a central fixation cross. Progress to the next trial was only made when the infant gazed to the fixation cross in order to ensure that all participants begin the face exploration from the same location (Figure S1A).

Test phase. This phase consisted of two trials; each lasting 5 seconds long (Figure S1B). Each test trial was preceded by the presentation of a central fixation cross. The position of the novel emotional expression (left or right side of the display) was reversed between the two trials. The position of the novel emotional expression in the first trial was randomly counterbalanced across the infants.

During the experiment, infants viewed the face presentation on the monitor without any active task.

Supplemental data analysis

This section is organized in three parts. Raw data and analysis codes are available upon request.

- Data pre-processing:
 1. Eye movement data pre-processing.
- Supplemental details of the analysis presented in the main text. Please note that our main conclusions are based on the analyses S2 and S3:
 2. *iMap4* analysis of the *familiarization* phase. We applied spatial mapping to assess the cultural differences.
 3. Visual preference analysis of the *test* phase. Linear models were applied to assess 7-month-old infants' ability to discriminate between expressions.
 4. Fixation pattern analysis of the *test* phase. We used unsupervised learning to quantify the stability of the fixation patterns at the individual level.
- Confirmatory analysis of the cultural fixation contrasts:
 5. Replication of the *iMap4* results using GMM clustering and the fixation bias ratio

1. Eye movement data pre-processing

Since our experiment is directly comparing two groups of observers, it was important to ensure that any observed eye movement difference between the two groups was not due to *inherent* differences in terms of precision/accuracy of the eye movement recordings. In fact, some physiological factors can affect the quality of the eye tracking data recordings, because of the infrared camera technology used to track eye movements in remote eye trackers (such as the Tobii X120 used here). For instance, the narrow eyes of some Asian participants might lead to a reduction in the accuracy and precision [S4]; data recorded from participants with a light/bluish eye color (a color variation of the iris only present in Western participants) are more difficult to track, less precise compared to dark eye color [S5 p. 42; S6]. We took several approaches to take into account all these potential technical problems:

- 1) Infants presenting the physical variations of the eyes (i.e., narrow eyes, light/bluish eye color, etc.) that make eye tracking difficult were detected at the beginning of the experiment, as the eye tracker failed to adequately calibrate those participants. As already reported above, the data from these participants (6 WCs, 9 EAs) were not included in the analyses.
- 2) We used the up-to-date Tobii eye tracker algorithm that includes a Bright Pupil/Low Diode Visibility option that compensates for the variations in terms of pupil color.
- 3) Eye tracker signal drifts were identified at a single-subject level during the pre-processing of the data, and corrected manually when necessary (5.9% of the participants - see below).
- 4) In the *iMap* analysis, we applied a larger size of smoothing compared to the one used in adults studies (1.5 FWHM instead of 1 FWHM) in order to minimize the intrinsic inter-infants eye movement variability [S7].
- 5) The cultural differences in fixation patterns we are reporting here could be also reproduced with approaches that are less sensitive to the precision/accuracy of the movement data (GMM clustering and fixation bias ratio, see below). Therefore, the lack of precision/accuracy in terms of eye movement data of infants cannot be a valid explanation for the cultural-contrast in eye movement strategies reported here.

Raw eye movement recordings were filtered into eye movement events (e.g., fixations and saccades) using the Tobii I-VT fixation filter algorithm, as implemented in Tobii Studio. Fixation information (*x*-, *y*- coordinates and fixation duration) recorded during the experiment were exported from Tobii Studio data files. Data analysis was then performed in Matlab2014b (Mathworks) using custom scripts and various Matlab toolboxes. Fixations landing outside of the stimuli were discarded from the analysis. The fixations from the *familiarization* phase were manually aligned using a custom script across all participants to correct for the drift during eye movement recording. For each subject, we visually identified the local density centers of the fixation data around the eye region (i.e., left and right eye). In case of a large drift ($> 2.5^\circ$ degrees of visual angle), we aligned the mid-point of the two local density peaks to the intersection of the central line of the face and the horizontal line of the eyes. After this linear transformation the local density peaks on both eyes have an equal distance to the center of the face and the local density peaks are on the same horizontal line with the eyes after correction. Those individuals without clear fixation cluster on both eyes were left unchanged. It is worth noting that this procedure affected only 5.9% of the participants (6 WCs, 3 EAs) and that the cultural differences in the fixation pattern revealed with *iMap4* are preserved when these participants are excluded from the sample, reproducing a similar pattern of results as the one reported in Figure 1A.

2. *iMap4* analysis of the *familiarization* phase

In order to assess whether WC and EA infants differ in terms of strategies to extract visual information during the *familiarization* phase, spatial mapping of the fixation pattern was conducted using *iMap4* [10]. This is a data-driven analysis framework for statistical fixation mapping [S7]. Given that only for the *familiarization* phase we had enough

data to ensure an appropriate statistical power for spatial modelling, the *test* phase data were analysed by using another approach (see below).

*i*Map4 implements a Linear Mixed Model (LMM) and non-parametric statistics based on resampling and spatial clustering [10]. We used *i*Map4 to project the duration of the fixations into two-dimensional space according to their *x*- and *y*- coordinates at the single-trial level. The sparse fixation duration maps were computed by smoothing the fixations with a 2D Gaussian Kernel function of full width at half maximum (FWHM) around 1.5° visual angle (Figure S1B). To model the spatial distribution of the fixation pattern, we normalized the fixation map by dividing the single-trial smoothed map with the sum of all the pixel value within each map. The resulting 3D normalized fixation duration matrix (trials x *x*-Size x *y*-Size) was then modeled in *i*Map4 as the response variable. To quantify the spatial bias in the fixation pattern of our 2x2x2 design, the value of each pixel in the smoothed duration map was fitted with a full model using the following formula (Eq. s1):

$$\begin{aligned} \text{Normalized Fixation duration}_{(x,y)} \sim & 1 + \text{Culture} + \text{Expression} + \text{Face Race} + \text{Culture} * \text{Expression} \\ & + \text{Culture} * \text{Face Race} + \text{Expression} * \text{Face Race} + \text{Culture} * \text{Expression} * \text{Face Race} \\ & + (1 | \text{Subject}), 1 \leq x \leq x\text{Size}, 1 \leq y \leq y\text{Size} \end{aligned}$$

Thus, the normalized fixation duration at different spatial locations (e.g., eyes, nose, or mouth) was considered as a function of *Culture* (UK or Japanese infant), *Expression* (Fear or Happy), *Face Race* (WC or EA face stimuli), and their *interactions*. The subject predictor was treated as random intercept to control for the variation across individuals and account of the repeated measurement (i.e., trials). *i*Map4 made use of the *LinearMixedModel* class from the Statistics Toolbox™ in Matlab for model fitting and to perform hypothesis testing such as ANOVA and linear contrasts. The linear mixed model coefficients were estimated using maximal likelihood (ML) with the default *i*Map4 settings. A bootstrap spatial clustering procedure threshold on the cluster size was applied for null hypothesis significance testing and multiple comparison corrections [10].

To control for false positives from pixel-wise hypothesis testing, *i*Map4 applies a bootstrap spatial clustering method as statistics that simultaneously consider multiple comparison corrections. The original parametric statistical values were thresholded at $p < .05$. *i*Map4 computes the size of the significant clusters and later compares them with a bootstrap distribution. To construct the bootstrap null distribution, *i*Map4 removed the conditional mean from each categorical condition and randomized without replacement to further disrupt possible covariation. This procedure ensures the null hypothesis, in which no difference in the mean is true for all coefficients and linear contrasts while keeping the variance intact. *i*Map4 then bootstraps the subject with replacement to create 1000 null response matrices, and performs the same modelling (Eq. s1) and linear contrasts (ANOVA on Eq. s1). For each bootstrap, *i*Map4 computes the maximum cluster size of the F-value map at $p < .05$ and saves it in a vector. The bootstrapped cluster distribution under H0 is the sorted vector resulting from each hypothesis testing. Statistical significance of the original F test is the cluster with size value larger than the 95th bootstrapped cluster size in the null distribution.

The *i*Map4 results are reported in Figure 1A. Pixel-wise ANOVA on the model coefficients of the linear mixed model (Eq. s1) revealed a significant main effect of *Culture* on the mouth and the nose area and a significant main effect of *Culture* around the eye region. By performing a linear contrast between WC and EA infants, our data show that WC infants fixated more on the mouth compared to EAs (local maximum within the significant cluster: $\beta_{WC} = 2.49$ [1.992, 2.981], $\beta_{EA} = 1.24$ [0.747, 1.743], $F(1, 757) = 12.07$; local minimum: $\beta_{WC} = 1.54$ [1.279, 1.809], $\beta_{EA} = 1.17$ [0.899, 1.434], $F(1, 757) = 3.86$, $p < 0.05$ cluster corrected; brackets show 95% confidence interval). In comparison, EA infants showed a bias towards the eye region and displayed longer fixation duration on the eye than WCs (local maximum within the significant cluster: $\beta_{WC} = 1.72$ [1.360, 2.081], $\beta_{EA} = 2.81$ [2.451, 3.176], $F(1, 757) = 17.59$; local minimum: $\beta_{WC} = 0.26$ [0.160, 0.351], $\beta_{EA} = 0.39$ [0.295, 0.487], $F(1, 757) = 3.86$, $p < 0.05$ cluster corrected).

3. Visual preference analysis of the *test* phase

In order to assess whether infants discriminate between facial expressions of emotion, we quantified the viewing bias towards fear or happy facial expressions in the group and at individual level, by estimating the total fixation duration (viewing duration) on the face under different conditions. Importantly, the viewing length of the familiarized and novel expression are correlated within each trial, nested under each individual observer. To account for the multivariate effect of visual preference, we applied multivariate generalized linear model on viewing length of familiarized and novel expression (Eq. s2):

$$[Duration_{familiarized}, Duration_{novel}] \sim 1 + Culture + Expression + Face\ Race + Culture * Expression + Culture * Face\ Race + Expression * Face\ Race + Culture * Expression * Face\ Race$$

Thus, the multivariate effect of viewing duration on familiar or novel faces was analyzed as a function of *Culture* (UK or Japanese infant), *Expression* in the *familiarization* phase (Fear or Happy), *Face Race* (WC or EA face stimuli), and their *interactions*. We performed linear contrasts on the model coefficient of Eq. s2 for hypothesis testing. The between subject effects (*Culture*, *Expression*, and *Face Race*) are tested on the difference ($Duration_{familiarized} - Duration_{novel}$). The within subject effects (viewing duration bias towards familiarized or novel stimuli and viewing duration bias towards fearful or happy faces) were shown as a smoothed 2D histogram using Plotly Matlab API [S8] (Plotly Technologies Inc, see Figure 1B & 1C).

Importantly, because infant observers viewed two face stimuli simultaneously during the *test* phase, a complex interaction effect with the stimuli they familiarized with during the first part of the experiment may occur. To better estimate for the multivariate effect of viewing duration bias during the *test* phase, we introduced a novel bootstrap procedure based on 2D histogram. The 2D surface could be considered as the empirical likelihood function of the viewing duration bias. The distance between the density peak of the 2D surface and the diagonal line represents the magnitude of the viewing bias effect. To demonstrate the effect of familiarity or expression, we estimated the most likely location of the density peak using a bootstrap procedure. We randomly resample subject with replacement for 10,000 times and projected the bootstrapped data into 2D. We then applied an equal eight bins for both x-axis and y-axis and find the maximum bin location in each bootstrap. The most likely occurred area of the peak location is defined as the bin location contains the highest counts across all bootstraps (the most likely occurred area of the peak is displayed on top of the 2D contour map as the white square in Figure 1B & 1C).

4. Fixation pattern analysis of the *test* phase

To isolate statistical differences between the fixation patterns during the *test* phase, and visualize the fixation strategy changes between *familiarization* and *test* phases, we applied an unsupervised machine learning using Gaussian mixture model to cluster the fixation location and categorize the individual fixation strategy as being different fixation strategies. Based on the spatial mapping results from *iMap4*, we then first fitted a multivariate 2D Gaussian mixture model (GMM) with five components on *all* the fixation location data across participants during *familiarization* phase. The GMM was fitted using *gmdistribution* class in Matlab with iterative Expectation-Maximization (EM) algorithm and 10 replications to ensure robustness. The number of components was chosen to maximize the component posterior probability accounting for the crucial regions of the face: four face features (two eyes, nose, and mouth) and the surrounding area (i.e., outside of the face). This is a data-driven alternative of the region of interested (ROI) analysis. We identified the model components corresponding to the eye region and nose/mouth region. The procedure of the fixation strategy classification is shown in Figure S2A. The individual fixation strategy was computed as the following fixation bias ratio:

$$Ratio_{familiarization} = \frac{fixation_{eye}}{fixation_{eye} + fixation_{mouth}}$$

This ratio ranges within [0, 1], with 0 indicating the infant only fixates on the nose/mouth and 1 indicating the infant only look at the eye region of the face. The fixation bias ratio for all participant during *familiarization* and *test* phases could be found in Figure S2B. To ensure the fixation labelling using GMM and the strategy classification is valid, we replicated the *iMap4* result using both the GMM clustering and the fixation bias ratio.

We then used the trained GMM to cluster the fixation data during *test* phase into fixation in the eye region and nose/mouth region on an image-by-image basis. The fixation bias ratio was calculated similarly for each infant for each expression presented in the *test* phase (i.e., fear and happy). To investigate whether the fixation strategy (i.e., fixation bias ratio) predicts the viewing preference, we fitted a simple linear regression (Eq. s3) independently for each of the categorical predictors:

$$(Duration_{fear} - Duration_{happy}) \sim 1 + Ratio_{familiarization} + Ratio_{test-fear} + Ratio_{test-happy} + Interactions$$

In addition, we modeled the fixation strategy changes between *familiarization* and *test* phase using a linear mixed model (Eq. s4):

$$(Ratio_{familiarization} - Ratio_{test}) \sim 1 + Culture + Expression + Expression_{test} + Face\ Race + Interactions + (1|Subject),$$

This is a full design with all the main effects (*Culture*: UK or Japan, *Expression*: Fear or Happy in the *familiarization* phase, *Expression_test*: Fear or Happy in the *test* phase, *Face Race*: WC or EA face stimuli) and all their interactions. We considered the subject as the random effect to take into account the between-subject effect of *Expression_test* in the *test* phase. An ANOVA was then performed on the Eq. s4 and linear contrasts were conducted as post-hoc analysis (see Figure S2C).

This analysis is performed to explore the relationship between the differential fixation strategy (i.e., eye-nose/mouth bias) and the viewing duration bias for the fearful expression stimuli. Multiple linear regressions between viewing duration bias and the fixation bias ratio (Eq. s3) did not return any significant effect after multiple comparison correction ($t_{\max}(11) = -2.90$, *n.s.* at $p < .05$ Bonferroni corrected). The fixation pattern (in both *familiarization* and *test* phase) does not predict the discrimination ability: both groups performing comparably well. Moreover, the ANOVA on Eq. s4 showed that the fixation strategy is overall consistent between *familiarization* and *test* phase for the infant observers, as the fixed intercept of Eq. s4 is not significantly different from zero (Ratio differences = $-.025$ [$-.0515$, $.0019$], $t(290) = -1.83$, $p = .0680$). Pair comparisons performed for each condition showed that infants do not change their fixation strategy between the *familiarization* and *test* phase, except in one condition (WC infants familiarized with the own-race fearful expression looked more to the eye region of the (non-familiarized) happy expression during the *test* phase, $\text{Mean}_{\text{diff}} = -.106$ [$-.197$, $-.015$], $t(290) = -2.30$, $p = .0022$ significant after Bonferroni correction).

5. Replication of the iMap4 results using GMM clustering and the fixation bias ratio

We used a five-component Gaussian mixture model to cluster the fixation in *familiarization phase* into eye region or nose/mouth region. We applied a linear mixed model analysis on the fixation duration in these two regions. Total fixation duration of each trial for each participant within the region was fitted in a linear mixed model similar to Eq. s1.

$$\begin{aligned} \text{Fixation Duration} \sim & 1 + \text{Culture} + \text{Expression} + \text{Face Race} + \text{Culture} * \text{Expression} + \text{Culture} * \text{Face Race} \\ & + \text{Expression} * \text{Face Race} + \text{Culture} * \text{Expression} * \text{Face Race} \\ & + (\text{Viewing duration} | \text{Subject}) \end{aligned}$$

Thus, the total fixation duration on the eye region and nose/mouth region was considered as a function of *Culture* (UK or Japanese infant), *Expression* (Fear or Happy), *Face Race* (WC or EA face stimuli), and their *interactions*. Also, the total fixation durations for each trial were treated as random effects, and were nested under each subject to control for the variation across individuals and account for the repeated measurement (i.e., trials). We found the same significant main effect of *Culture* in the eye region ($F(1, 757) = 6.94$, $p = .0086$), with also a significant interaction between *Culture* and *Face Race* ($F(1, 757) = 5.40$, $p = .0204$). The linear contrast showed that EA infants fixated longer than WCs in the eye region (EA: $M_{\text{viewing duration}} = 2.80\text{s}$ [2.597 , 3.004]; WC: $M_{\text{viewing duration}} = 2.35\text{s}$ [2.085 , 2.617]). Moreover, EAs fixated longer on the eyes when they were presented with other-race faces than own-race faces (Other-race: $M_{\text{viewing duration}} = 3.16\text{s}$ [2.830 , 3.498], Own-race: $M_{\text{viewing duration}} = 2.44\text{s}$ [2.204 , 2.670]), $t(757) = -3.50$, $p = 4.888\text{e-}04$). As a comparison, the difference between own-race and other-race faces in WC infants does not reach the significance ($t(757) = .25$, $p = .8052$). Similarly, the main effect of *Culture* is significant in the mouth area (EA: $M_{\text{viewing duration}} = 0.75\text{s}$ [0.543 , 0.964]; WC: $M_{\text{viewing duration}} = 1.30\text{s}$ [1.041 , 1.566], $F(1, 757) = 10.32$, $p = .0014$). In addition, we fitted the above equation with the fixation bias ratio computed from the GMM clustering as response variable. The results also revealed a significant *Culture* effect ($F(1, 757) = 7.44$, $p = .0065$). Altogether, our main results using iMap4 and the supplementary results presented here confirm the robustness of the cultural difference in fixation patterns in 7-month-old infants during the discrimination of facial expressions of emotion.

Supplemental References

- S1. Lundqvist, D., Flykt, A., and Ohman, A. (1998). The Karolinska directed emotional faces—KDEF. Stockholm, Sweden: Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet.
- S2. Lyons, M., Akamatsu, S., Kamachi, M., and Gyoba, J. (1998). Coding facial expressions with gabor wavelets. In *Automatic Face and Gesture Recognition*, 1998. Proceedings. pp. 200-205
- S3. Goeleven, E., De Raedt, R., Leyman, L., and Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cogn. Emot.* 22, 1094-1118.
- S4. Blignaut, P., and Wium, D. (2014). Eye-tracking data quality as affected by ethnicity and experimental design. *Behav. Res. Methods.* 46, 67-80.

- S5. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. (OUP Oxford).
- S6. Nyström, M., Andersson, R., Holmqvist, K., and van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behav. Res. Methods.* 45, 272-288.
- S7. Caldara, R., and Miellet, S. (2011) iMap: a novel method for statistical fixation mapping of eye movement data. *Behav. Res. Methods.* 43, 864-878.
- S8. Plotly Technologies Inc. Collaborative data science. Montréal, QC, 2015. <https://plot.ly>.