Foveal vision determines the perceived emotion of face ensembles

Yu Dandan, Luyan Ji, Yunping Song, Bilge Sayim

Foveal vision determines the perceived emotion of face ensembles

Yu R. Dandan[1], Luyan Ji[2], Yunping Song[3], and Bilge Sayim[1,4]

[1] Sciences Cognitives et Sciences Affectives (SCALab), CNRS, UMR 9193, University of Lille, Lille, France

[2] Center for Brain and Cognitive Sciences, Department of Psychology, Faculty of Education, Guangzhou University, Guangzhou, China

[3] School of Education, Soochow University, Suzhou, China

[4] Institute of Psychology, University of Bern, Bern, Switzerland

Abstract

People can extract summary statistical information from groups of similar objects, an ability called ensemble perception. However, not every object in a group is weighted equally. For example, in ensemble emotion perception, faces far from fixation were weighted less than faces close to fixation. Yet the contribution of foveal input in ensemble emotion perception is still unclear. In two experiments, groups of faces with varying emotions were presented for 100 ms at three different eccentricities (0°, 3°, 8°). Observers reported the perceived average emotion of the group. In two conditions, stimuli consisted of a central face flanked by eight faces ('flankers') ('central-present' condition) and eight faces without the central face ('central-absent' condition). In the central-present condition, the emotion of the central face was either congruent or incongruent with that of the flankers. In Experiment 1, flanker emotions were uniform (identical flankers); in Experiment 2 they were varied. In both experiments, performance in the central-present condition was superior at 3° compared to 0° and 8°. At 0°, performance was superior in the central-absent (i.e., no foveal input) compared to the central-present condition. Poor performance in the central-present condition was driven by the incongruent condition where the foveal face strongly biased responses. At 3° and 8°, performance was comparable between central-present and central-absent conditions. Our results showed how foveal input determined the perceived emotion of face ensembles, suggesting that ensemble perception fails when salient target information is available in central vision.

Keywords: ensemble emotion perception, foveal input bias, peripheral vision

Introduction

Ensemble perception is the visual system's ability to extract summary statistical information from groups of similar objects (Dakin & Watt, 1997; Haberman & Whitney, 2007, 2009; Whitney & Yamanashi Leib, 2018). For example, observers are able to extract the average size of a group of objects without inspecting each individual object. Ensemble perception has been shown not only for a large range of 'low-level' features such as size (Chong & Treisman, 2003, 2005), orientation (Dakin & Watt, 1997; Parkes et al., 2001), and motion (Watamaniuk et al., 1989), but also 'high-level' features such as the gaze of crowds (Sweeny & Whitney, 2014), emotion (Haberman & Whitney, 2007, 2009), gender (Haberman & Whitney, 2007), attractiveness (Luo & Zhou, 2018) and identity of faces (Jung et al., 2017). Representing features of ensembles by summary statistics is an efficient way to represent complex stimuli under limited capacity (Alvarez, 2011). Importantly, not all items in a group contribute equally to the perception of the ensemble (Dakin, 2001; Haberman & Whitney, 2010; Solomon, 2010; Allik et al., 2013; Hubert-Wallander & Boynton, 2015). For example, it was shown that observers tend to integrate only about the square-root of the number of items ($\sqrt{N}$) during ensemble coding (see, e.g., Whitney & Yamanashi Leib, 2018). Besides, the feature distribution of the stimuli in the group also matters (Haberman & Whitney, 2010; Michael et al., 2014; Kanaya et al., 2018; Cant & Xu, 2020): Outliers - for example a strongly tilted line among weakly tilted lines (e.g., Epstein et al., 2020) - are often weighted less than the majority of items that are more similar in regard to the measured feature.

When presented with a set of faces varying in emotional states, observers were capable of accurately estimating the average emotion of faces (Haberman & Whitney, 2007, 2009, 2010). This capacity to extract emotional states from groups of faces has been shown for short presentation times (as short as 50ms; Li et al., 2016), large sets (up to 24 faces; Wolfe et al., 2015), and even for Mooney faces (Han et al., 2021). However, when multiple faces are integrated into an ensemble representation, not all faces are necessarily weighted equally (Whitney & Yamanashi Leib, 2018; Hubert-Wallander & Boynton, 2015). For example, previous studies found eccentricity-based weighting of ensemble face representations (e.g., To et al., 2019; Jung et al., 2017; Ji et al., 2014). Several studies showed a fovea-bias in ensemble face perception: faces that were close to fixation ('foveal faces'; at about 2° of visual angle around fixation in Atkinson & Smithson, 2013; Jung et al., 2017) were weighted more than more peripheral faces (Atkinson & Smithson, 2013; Ji et al., 2014; Jung et al., 2017). In Ji et al.'s study (2014), stimuli consisted of 16 faces with varying facial expressions. The stimuli were divided into two subsets: the 4 central faces (occupying 3.98 × 4.02 degrees of visual angle) were considered as foveal input and the other 12 faces extrafoveal input. The emotional valence of the foveal and extrafoveal input was either congruent (both positive or negative) or incongruent (one positive and one negative subset). Participants were asked to judge the face set's average emotion which was always the same as the emotional valence of the extrafoveal input (observers were not informed about this). It was found that the ensemble performance was better in the congruent than in the incongruent condition. The results indicated that the foveal input weighted more than extrafoveal input in ensemble emotion perception. At the same time, some studies suggested that foveal input was not required for ensemble emotion perception (Haberman et al., 2009; Wolfe et al., 2015; To et al., 2019). For example, Wolfe and colleagues (2015) found participants' ensemble performance was unaffected when there was no foveal input. In their study, stimuli consisted of 24 faces with different levels of happy, sad, and angry expressions, presented for 1500 ms (participants were allowed to make eye movements). In the condition without foveal input, a gaze-contingent occluder was used to occlude a circular foveal region of 2.6 degrees of visual angle. After stimulus presentation, participants adjusted a probe face to match the perceived average emotion of the face set. No difference between the conditions with occluded and non-occluded foveal input was found. Hence, foveal input has been shown to be unnecessary (Wolfe et al., 2015), and to bias responses (Ji et al., 2014; Jung et al., 2017). In a recent review it was proposed that – consistent with these studies - foveal information might not be necessary for ensemble coding, however, once there is foveal input, it may bias individuals' averaging estimation (Whitney & Yamanashi Leib, 2018).

When presenting a face set either in the fovea or the parafovea, To et al. (2019) found a parafovea averaging advantage. In their experiment, a set of 9 faces was presented either at fixation (fovea) or at 3° eccentricity (parafovea). Participants were asked to judge the average emotion of the face set. Ensemble judgments were more accurate in the parafovea than in the fovea, showing a parafovea averaging advantage. Importantly, participants' responses in the foveal condition were biased by the central face, indicating that observers were not able to equally weight foveal and parafoveal faces. However, as there was no condition without foveal input, the exact role of the foveal face on ensemble emotion perception remained unclear. In the current study, we directly compared performance in conditions with and without a foveal face, and with the same stimuli at different eccentricities. Unlike most previous studies that investigated the contribution of foveal input to ensemble perception (e.g., Ji et al., 2014; Jung et al., 2017; Wolfe et al., 2015), only a single face was used as the foveal input. Stimuli consisted of a 3×3 matrix of faces with a central face either present ('central-present' condition) or absent ('central-absent' condition), and eight surrounding faces ('flankers'). Additionally, a single face was presented at the central face location ('single face' condition). The emotions of the flankers were either congruent (either all faces happy or disgusted) or incongruent (happy central face and disgusted flankers or vice versa) with the central face's emotion (differences between the congruent, incongruent, and central-absent conditions would show the bias induced by the central face; see below). The face sets and the single face were presented at three different eccentricities: 0°, 3°, 8°. Observers were asked to indicate the average emotion (positive or negative) of the entire set. In the single face condition, observers reported whether the face was positive or negative. At 0°, the center of the face set was presented at fixation, enabling us to measure the contribution of the foveal face to ensemble perception by comparing the central-present and central-absent conditions, and thereby estimating the foveal input bias**.** The face set as at 0° was used at two peripheral locations, with the central faces centered at 3° or 8°. No (or a much weaker) bias by the central face was expected in the two peripheral conditions compared to the foveal condition. In Experiment 1, all flanker emotions were the same (‘uniform’ condition). To test to what extent the grouping of the flankers by similarity – and correspondingly, ungrouping of the central face from the flankers – played a role in ensemble emotion perception, we varied flanker emotions in Experiment 2 (‘varied’ condition).

Taken together, we tested whether – and to what extent - foveal input would bias ensemble emotion perception by comparing the performance in the congruent, incongruent, and central-absent conditions: If the foveal face biased ensemble emotion perception, observers' ensemble performance would be expected to be impaired in the incongruent condition when the foveal face was present compared to when it was absent. By contrast, in the congruent condition, a bias to respond with the foveal emotion would yield correct responses. If there was no foveal input bias, the ensemble performance would be expected to be similar in the conditions with and without the foveal face (averaging either eight or nine faces), as well as in the congruent and incongruent conditions. Furthermore, we presented the face set at three different eccentricities to compare the possible bias by the central face in the foveal location (i.e., foveal input bias) and in the periphery. At 3° and 8°, neither a difference between the central-present and central-absent conditions, nor between the congruent and incongruent conditions was expected. Varying eccentricity also allowed us to test whether the parafoveal averaging advantage in ensemble emotion perception could be explained by the foveal input bias. If the parafovea averaging advantage mentioned above was a result of the foveal input bias, participants' ensemble performance would be expected to be better at 3° than at 0° in the central-present, incongruent condition but not in the central-absent and congruent conditions. Finally, the flanker homogeneity manipulation was designed to test whether (un)grouping of the central face and the flankers was driving the foveal input bias: Ungrouping of the central face from the flankers in

Experiment 1 was expected to modulate the foveal input bias. In particular, uniform (Experiment 1) compared to varied flankers (Experiment 2) could have resulted in either a weaker foveal input bias – because the 'ungrouped' foveal item could be ignored and its contribution to ensemble estimates lessened (or corrected) more easily, or a stronger foveal input bias – because access to the 'ungrouped' flankers could be hindered. Taken together, the main goal of the current study was to investigate the role of foveal input in ensemble emotion perception by testing if – and to what extent -- it biased estimates of the ensemble.

To preview our results, we found a strong foveal input bias at 0°. Performance was superior when the foveal input was absent than when it was present. The deterioration of performance with a foveal face present (central-present condition) was driven by the incongruent condition where the emotion of the foveal face strongly biased responses. At 3° and 8°, no bias by the central face was observed. Performance in the central-present condition was better at 3° compared to 0° and 8°. However, in the central-absent condition – where no foveal face was presented at 0° - the ensemble performance was superior at 0° compared to 3° and 8°, suggesting that foveal input biases could play an important role in the parafoveal averaging advantage in ensemble emotion perception. The pattern of results was similar with identical (Experiment 1) and varied (Experiment 2) flankers, indicating that (un)grouping of the central face with (from) the flankers due to flanker homogeneity did not underlie the foveal input bias observed in Experiment 1. Taken together, by directly comparing observers' discriminability to average facial expressions in the presence and absence of a foveal face, as well as in the fovea and periphery, our results revealed a strong foveal input bias in ensemble emotion perception. Importantly, the very low discriminability when the emotion of the foveal face was incongruent with that of the flankers suggests that ensemble perception may fail when salient target information is available in central vision.

Experiment 1: Uniform flankers

Method

Participants

In Experiment 1, 17 observers participated (18-25 years, 12 females, 5 males). The number of participants was based on an a priori power analysis based on the smallest effect size from a previous investigation using a similar paradigm (To et al., 2019, $\eta^2 = 0.21$), with α at 0.05. A sample size of 8 was needed to achieve a power of 0.95 (1-β). All participants had normal or corrected-to-normal vision and were naïve to the purpose of the experiment. They provided informed consent approved by the Institutional Review Board at Soochow University and got paid after the experiment.

Stimuli

The face stimuli were created using three images of the same individual with happy, disgusted, and neutral expressions from the NimStim database (Tottenham et al., 2009). All external features, such as hair, neck and ears, were removed from the faces by using GIMP software (Version 2.10). Fantamorph (Version 5) was used to create 11 different emotional valences by morphing the happy and disgusted expressions, respectively, with the neutral expression, yielding the following percentages: 100% happy/disgusted, 80% happy/disgusted and 20% neutral, 60% happy/disgusted and 40% neutral, 40% happy/disgusted and 60% neutral, 20% happy/disgusted and 80% neutral, and 100% neutral. Stimuli were presented on a gray background (85 cd/m^2). There were three conditions: a single face (the 'single-face' condition), a face set containing 9 faces (i.e., a central face and eight surrounding faces, i.e., 'flankers'; the 'central-present' condition), and a face set without the central face (i.e., only the eight flankers; the 'central-absent' condition) (Figure 1a). The face set (or a single face) was presented centered at three different eccentricities: 0°, 3°, 8°. Each face subtended 1.49° $\times$ 2.21° of visual angle and was separated by 0.30° horizontally and 0.15° vertically from neighboring faces (edge-to-edge distance). The whole face set subtended a visual angle of 5.07° × 6.93°. Flankers' emotions were either identical in a given stimulus (Experiment 1) or varied (Experiment 2).

All stimuli were presented using E-prime 3.0 (Psychology Software Tools, Pittsburgh, PA) on a 19-in. LCD monitor (E196FP, DELL) with a refresh rate of 60 Hz and a resolution of 1280 × 1024. The viewing distance was kept constant at 57 cm using a chin-rest.

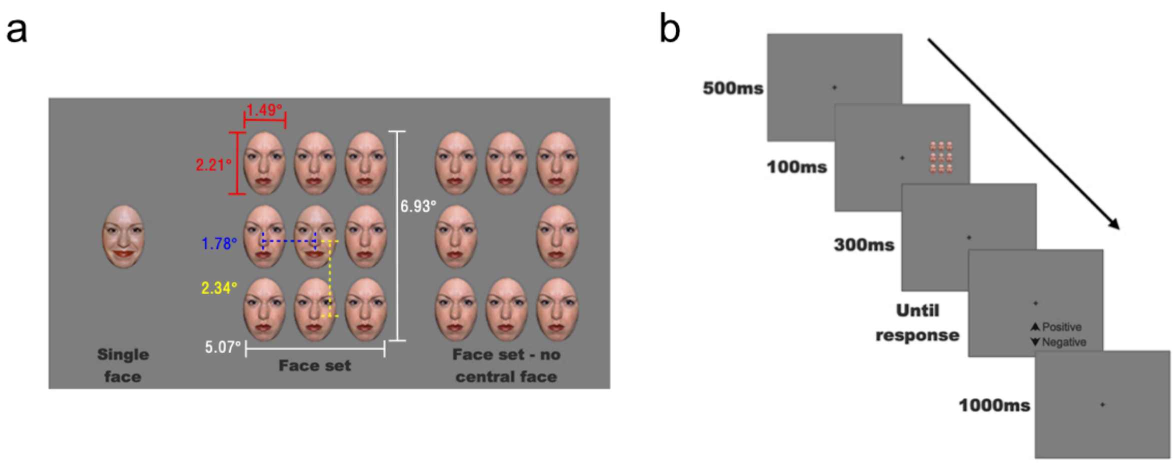
Design and Procedure

Participants were asked to report the average emotion of the face set. There were six blocks (2 (central face: present vs. absent) × 3 (eccentricity: 0°, 3°, 8°)) with 121 trials per block (11 emotions of the central face × 11 emotions of the flankers). In the central-present condition, each of the 11 faces was presented as central faces and as flankers, and there were two different congruency conditions: (1) the 'congruent' condition where the emotion of the central face and the flankers were the same (either both happy or both disgusted); (2) the 'incongruent' condition where the emotion of the central face and flankers were different (the central face happy and the flankers disgusted or vice versa). In the central-absent condition, the same stimuli as in the central-present condition were presented without the central face. In the 'single face' condition, a single face was presented centered at the three different eccentricities (0°, 3°, 8°). There were 3 blocks (one block per eccentricity) in each of which each of the 11 faces was presented in 11 trials (resulting in 121 trials per block). Hence, there were 1089 trials per observer (observers also performed a crowding task with the same stimuli in the same session; results not reported here). In Experiment 1, the eight flankers of a given stimulus were identical. Before the experiment, participants completed 12 practice trials in which a face set containing 9 faces was presented at fixation (i.e., central-present, 0° condition) and participants were required to report the average emotion of the face set.

On each trial, a black fixation cross was presented for 500 ms, followed by the stimulus (a single face, a face set containing eight or nine faces). Stimuli were presented for 100 ms either centered at 0° or randomly to the left or right of fixation at 3° or 8° eccentricity (eccentricity was kept constant throughout each block). After stimulus offset, a blank screen was presented for 300 ms, followed by the response screen. Participants were asked to judge whether the whole face set's average emotion (or the emotion of the single face) was positive or negative. After participants' responses, an inter-trial interval of 1000 ms was inserted before the next trial (Figure 1b).

Figure 1

Face Stimuli and Experimental procedure



(**a**) Stimuli of Experiment 1: a single face, a face set containing 9 faces (central-present condition), and a face set without the central face (central-absent condition). The 'flankers' in the two face set examples consist of 40% disgusted faces, and the central face in the central-present condition shows a 100% happy face. (**b**) General procedure of the study. Participants judged the emotion of the ensemble face (or the single face) by indicating 'positive' or 'negative'.

Analysis

To determine the discriminability and response bias, we used signal detection theory (SDT, Macmilian & Creelman, 2004) in our primary analyses, defining disgusted face sets reported as negative as “hits”, disgusted face sets reported as positive as “misses”, happy face sets reported as positive as “correct rejections”, and happy face sets reported as negative as “false alarms”. We calculated discriminability (d') and the criterion (c), using the following formula:

$$d' = z\,(\text{Hit}) - z\,(\text{False alarm})$$

$$c = -0.5 \times (z\,(\text{Hit}) + z\,(\text{False alarm}))$$

where z (Hit) and z (False alarm) are the z transforms of Hit and False alarm, respectively.

A criterion value of zero indicated no bias, a negative value represented a bias to report the face set as negative, and a positive value represented a bias to report the face set as positive. A repeated-measures ANOVA was used to analyze the discriminability and criterion data (see Figure 2, Figure 3). Heatmaps with the emotion of the central face plotted against the emotion of the flankers (11 × 11 matrices) to provide a visualization of the responses for each combination of central face and flankers are shown in Figure 4.

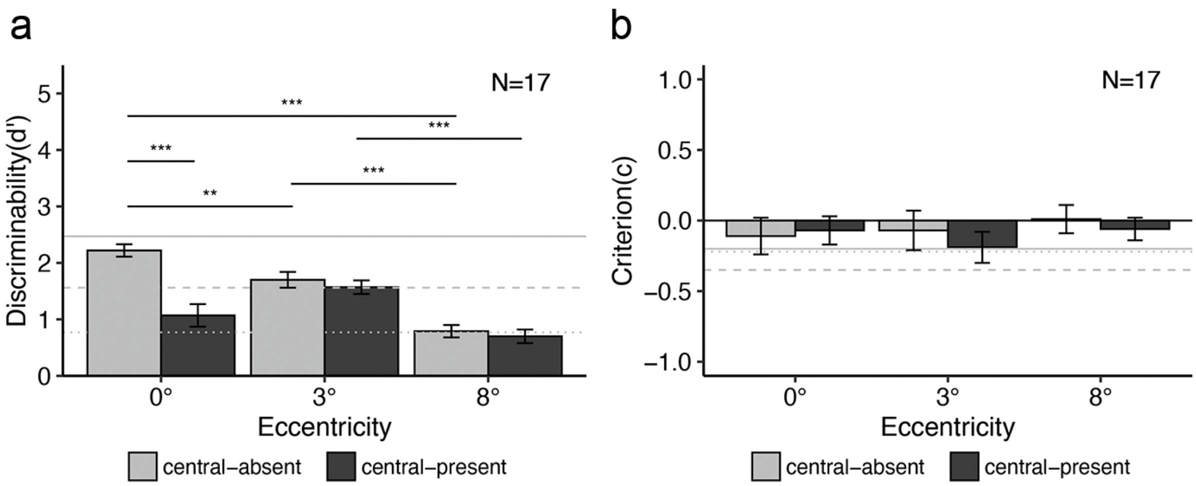
Results

Discriminability *and criterion*

We compared participants’ discriminability (d′) to identify the average emotion in the central-present and central-absent condition at the three different locations (0°, 3°, and 8°; Figure 2a). A repeated-measures ANOVA with the two factors Central Face (central-present vs. central-absent) and Eccentricity (0°, 3°, 8°) revealed significant main effects of Central Face, $F(1, 16) = 41.79$, $p < 0.001$, partial $\eta^2 = 0.72$, and Eccentricity, $F(2, 32) = 38.79$, $p < 0.001$, partial $\eta^2 = 0.71$, as well as an interaction between Central Face and Eccentricity, $F(2, 32) = 14.42$, $p < 0.001$, partial $\eta^2 = 0.47$. Participants’ ensemble performance was better in the central-absent condition compared to the central-present condition at 0° ($p < 0.001$), but not at 3° ($p = 0.36$) and 8° ($p = 0.50$). In the central-present condition, discriminability was higher at 3° (1.57 +/- 0.51) than at 8° (0.70 +/- 0.48) ($p < 0.001$), and there was a trend for higher discriminability at 3° (1.57 +/- 0.51) compared to 0° (1.07 +/- 0.83) ($p = 0.13$). There was no difference between 0° and 8° ($p = 0.24$). In the central-absent condition, discriminability was best at 0° (2.22 +/- 0.46), and decreased with eccentricity: Discriminability was higher at 0° compared to 3° (1.70 +/- 0.57; $p < 0.01$) and 8° (0.79 +/- 0.46; $p < 0.001$), and higher at 3° than 8° ($p < 0.001$). The average criterion (-0.08 +/- 0.11) was close to zero in all conditions, with a slight trend for a negative bias (i.e., judging the face set as negative; Figure 2b). A repeated-measures ANOVA on the criterion yielded no main effect of Central Face, $F(1, 16) = 0.95$, $p = 0.34$, partial $\eta^2 = 0.06$, no main effect of Eccentricity, $F(2, 32) = 0.89$, $p = 0.42$, partial $\eta^2 = 0.05$, and no Central Face × Eccentricity interaction, $F(2, 32) = 0.94$, $p = 0.40$, partial $\eta^2 = 0.06$.

Figure 2

Discriminability and Criterion for emotion recognition of Face ensembles in Experiment 1



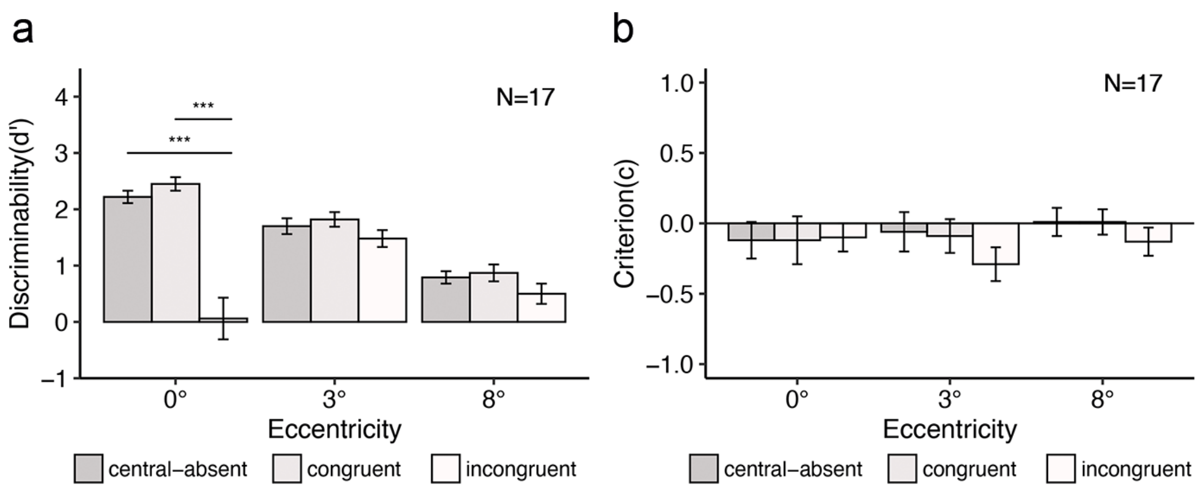
Results of Experiment 1. Discriminability (a) and criterion (b) separated for face sets with and without central face. The gray horizontal lines represent discriminability and criterion in the single face condition at 0° (solid line), 3° (dashed line) and 8° (dotted line). Asterisks indicate significance with alpha levels of 0.01 (**), and 0.001 (***). Error bars represent ± 1 *SEM*.

Congruency

To investigate the influence of the central face on ensemble perception, we calculated d' and c separately for congruent (the same emotion of the central face and the flankers), and incongruent (different emotions of the central face and the flankers), comparing the congruent, incongruent, and central-absent conditions. A repeated-measures ANOVA with two factors (Congruency × Eccentricity) was conducted. The results showed main effects of Congruency, $F(2, 32) = 56.17$, $p < 0.001$, partial $\eta^2 = 0.78$, and Eccentricity, $F(2, 32) = 33.69$, $p < 0.001$, partial $\eta^2 = 0.68$, and an interaction between Congruency and Eccentricity, $F(4, 64) = 14.47$, $p < 0.001$, partial $\eta^2 = 0.48$ (Figure 3). In the 0° condition, participants' averaging performance was similar in the congruent (2.45 +/- 0.48) and central-absent (2.22 +/- 0.46) conditions ($p = 0.30$), and worse in the incongruent (0.06 +/- 1.51) condition (congruent > incongruent: $p < 0.001$; central-absent > incongruent: $p < 0.001$). At 3° eccentricity, averaging performance was comparable in the three conditions (congruent (1.82 +/- 0.52) vs. central-absent (1.70 +/- 0.57): $p = 0.77$; incongruent (1.48 +/- 0.60) vs. central-absent: $p = 0.55$). However, there was a clear trend for lower discriminability in the incongruent compared to the congruent condition ($p = 0.05$). The pattern of results was similar at 8° as at 3° (congruent (0.87 +/- 0.61) vs. central-absent (0.79 +/- 0.46): $p = 0.93$; congruent vs. incongruent (0.50 +/- 0.74): $p = 0.23$; incongruent vs. central-absent: $p = 0.43$). As noted above, performance in the central-absent condition was best at 0°, worse at 3°, and worst at 8° (0° > 3°: $p < 0.01$; 0° > 8°: $p < 0.001$; 3° > 8°: $p < 0.001$). The pattern of results was similar in the congruent as in the central-absent condition (0° > 3°: $p < 0.001$; 0° > 8°: $p < 0.001$; 3° > 8°: $p < 0.001$). In the incongruent condition, however, higher discriminability was found at 3° compared to 0° and 8°, and there was no significant difference between 0° and 8° (3° > 0°: $p < 0.01$; 3° > 8°: $p < 0.001$; 0° vs. 8°: $p = 0.61$). The criterion analysis (ANOVA) showed that there was no main effect of Congruency, $F(2, 32) = 2.43$, $p = 0.10$, partial $\eta^2 = 0.13$, no main effect of Eccentricity, $F(2, 32) = 0.76$, $p = 0.48$, partial $\eta^2 = 0.05$, and no interaction between the two factors, $F(4, 64) = 1.00$, $p = 0.41$, partial $\eta^2 = 0.06$.

Figure 3

Discriminability and Criterion for congruent, incongruent and central-absent stimuli



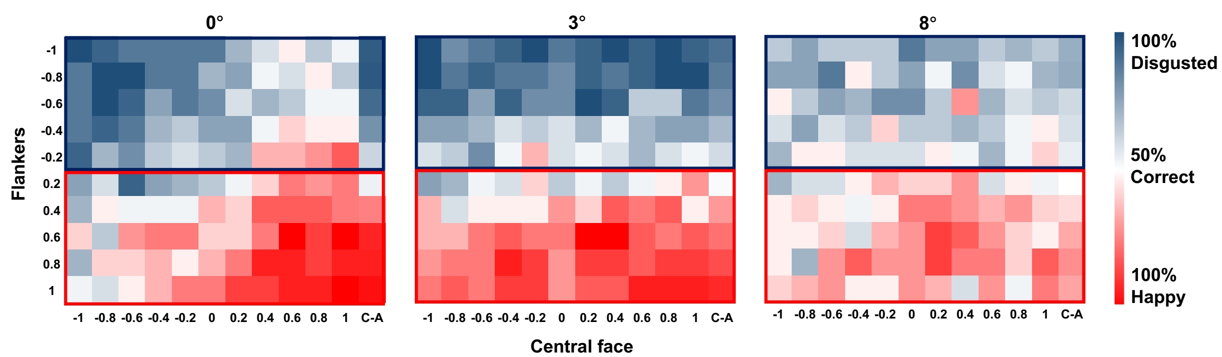
Congruency results of Experiment 1. Discriminability (a) and criterion (b) separated for face sets with congruent and incongruent central faces and flankers, and without a central face (central-absent condition). Asterisks indicate significance with alpha levels of 0.001 (***). Significance is only indicated for the comparisons of the three conditions (central-absent, congruent, incongruent) at each eccentricity. Error bars represent ± 1 *SEM*.

Proportion correct for different combinations of central face and flanker emotions

To illustrate the different contributions of the central face and the flankers to the ensemble judgments in the central-present condition, we plotted the proportion correct for all combinations of central face emotions and flanker emotions (Figure 4). The correct answer always corresponded to the emotion of the flankers. At 0°, participants' averaging performance was strongly biased by the emotion of the central face: When the central face's emotion was positive, participants judged the average emotion as positive even though the flankers were negative (and vice versa). At 3°, participants' ensemble judgment was mostly consistent with the emotion of the flankers regardless of the emotion of the central face. However, with slightly happy flankers (that required a happy response), there was a trend to respond with the central, negative face (see also the trend for better discriminability in the congruent compared to the incongruent condition). At 8°eccentricity, participants' ensemble performance was overall strongly impaired, and there was no bias from the central face.

Figure 4

Heatmap: proportion correct for all combinations of central face and flankers



Heatmap showing the results of Experiment 1. Each cell in the matrix represents participants' proportion correct with different combinations of central face and flankers. The x-axis represents the emotion of the central face and the y-axis represents the emotion of the flankers. A value of -1 represents 100% disgusted; 0 represents neutral; +1 represents 100% happy; "C-A" represents the central-absent condition. The blue and red rectangles surrounding the upper and lower part of the graphs correspond to the correct response (blue: "disgusted"; red: "happy").

Overall, we found a strong foveal input bias in the 0° condition. Participants' performance was better when the foveal input was absent than present. This effect was driven by the incongruent condition: When the emotion of the foveal face was different from that of the flankers, performance was strongly impaired compared to the condition where the emotion of the foveal face and the flankers was the same. In the central-present condition, we found a trend for better performance at 3° than 0°. In the central-absent condition, performance was best at 0°, worse at

3°, and worst at 8°. The flankers were identical in each given stimulus which could have caused or enhanced the foveal input bias. In Experiment 2, we sought to investigate the role of flanker homogeneity by varying the valence of flanker emotions.

Experiment 2: Varied flanker emotions

The strong foveal input bias we found might have (partly) been driven by presenting identical flankers. In particular, grouping of the flankers due to similarity – and, correspondingly, ungrouping of the flankers from the central face - could have made the central face stand out from the flankers, biasing responses. To investigate whether the foveal input bias found in Experiment 1 was due to the homogeneity of the flankers, we varied flanker emotions in Experiment 2. If the homogeneity of the flankers was a (major) reason for the foveal input bias, then the bias would be reduced or abolished with varying flankers.

Method

Participants

Eighteen new observers (18-23 years, 13 females) participated in Experiment 2. All reported normal or corrected-to-normal vision and provided informed consent approved by the Institutional Review Board at Soochow University and got paid after the experiment.

Stimuli and Procedure

Compared to Experiment 1 in which flankers were identical in each trial, flankers were varied in the current experiment. Average emotions of the face sets were the same as in Experiment 1 (i.e., 0%, 20%, 40%, 60%, 80%, 100% happy/disgusted). For each average emotion (except 0% and 100%, see below), we iteratively selected faces that maximized the number of different emotions within the set. To obtain high levels of variability, none of the face sets contained more than four faces of the same emotional valence. There were 11 unique stimuli per average emotion level. The emotion of the central face varied from 100% disgusted to 100% happy (i.e., in total of 11 levels). Each of the 11 emotions was presented as central face in the 60%, 40%, and 20% conditions (as Experiment 1). Note that in the 80% average emotion conditions, there were only three possible face combinations. In the 100% and the 0% average conditions, there was only one face combination (i.e., all the faces were the same). These stimuli were repeated in a block to match the number of trials with the other average emotion values (11 trials). As in Experiment 1, each block consisted of 121 trials (11 averages × 11 face combinations). The procedure was the same as in Experiment 1.

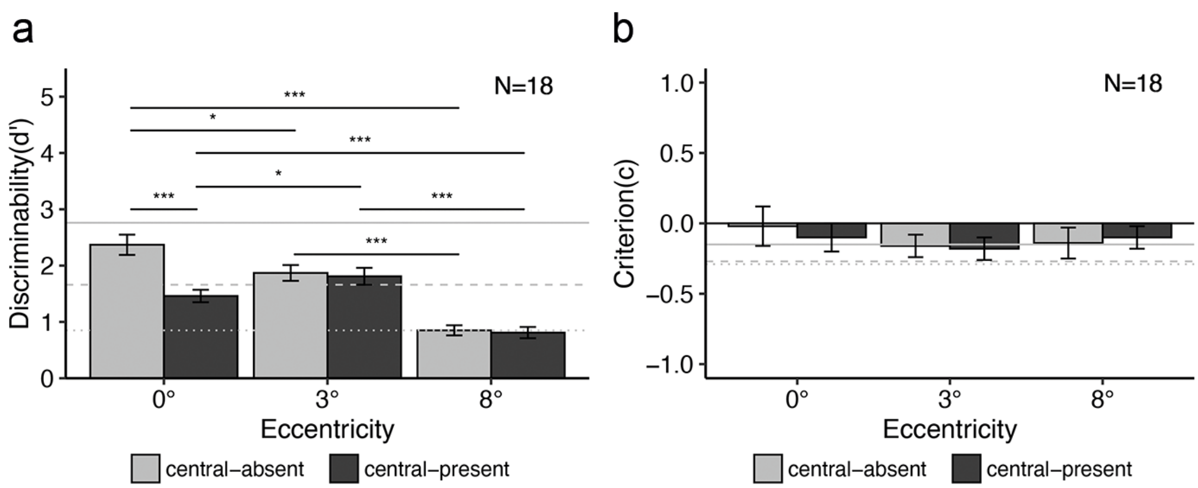
Results

Discriminability *and criterion*

The analysis of d′ revealed main effects of Central Face, F (1, 17) = 22.04, $p < 0.001$, partial $\eta^2 = 0.57$, and Eccentricity, F (2, 34) = 59.05, $p < 0.001$, partial $\eta^2 = 0.78$, and a Central Face × Eccentricity interaction, F (2, 34) = 13.04, $p < 0.001$, partial $\eta^2 = 0.43$. As in Experiment 1, d′ differed between the central-absent and central-present conditions only at 0°, with higher discriminability in the central-absent than in the central-present condition (Figure 5a; 0°: $p < 0.001$; 3°: $p = 0.62$; 8°: $p = 0.72$). In the central-present condition, performance was best at 3° (1.81 +/- 0.63), followed by 0° (1.46 +/- 0.45), and 8°(0.81 +/- 0.44; 3° > 0°: $p < 0.05$; 3° > 8°: $p < 0.001$; 0° > 8°: $p < 0.001$). In the central-absent condition, performance was best at 0° (2.37 +/- 0.76), worse at 3° (1.87 +/- 0.61), and worst at 8° (0.85 +/- 0.40; 0° > 3°: $p < 0.05$; 0° > 8°: $p < 0.001$; 3° > 8°: $p < 0.001$). As in Experiment 1, the average criterion (-0.12 +/- 0.1) was close to zero in all conditions with a slight trend for a negative bias (Figure 5b). A repeated-measures ANOVA on the criterion yielded no significant main effect of Central Face, F(1, 17) = 0.16, $p = 0.69$, partial $\eta^2 = 0.01$, no main effect of Eccentricity, F(2, 34) = 1.25, $p = 0.30$, partial $\eta^2 = 0.07$, and no Central Face × Eccentricity interaction, F(2, 34) = 0.60, $p = 0.56$, partial $\eta^2 = 0.03$.

Figure 5

Discriminability and Criterion for emotion recognition of face ensembles in Experiment 2



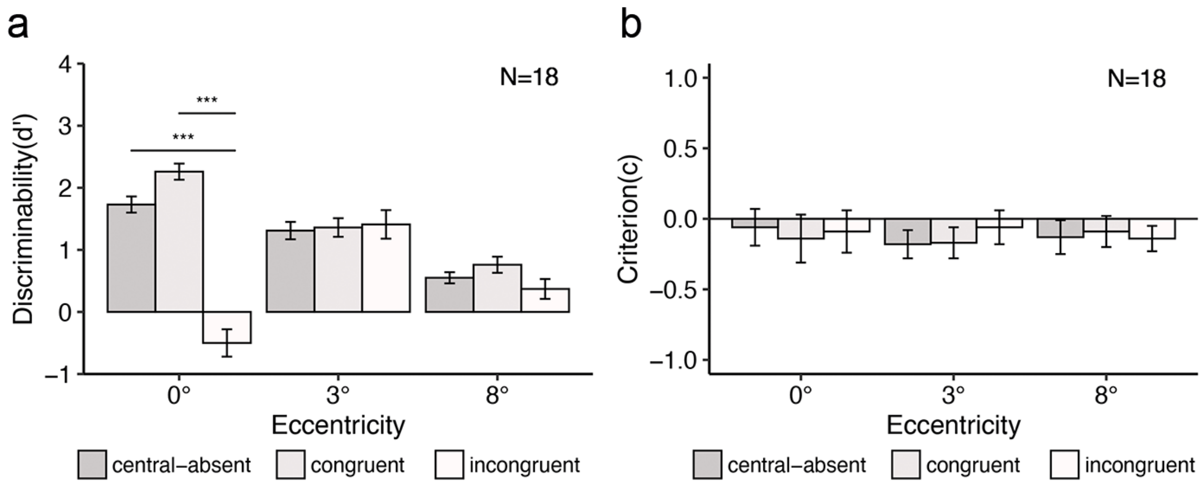
Results of Experiment 2. Discriminability (a) and criterion (b) separated for central-absent and central-present conditions. The gray horizontal lines represent discriminability and criterion in the single face condition at 0° (solid line), 3° (dashed line) and 8° (dotted line). Asterisks indicate significance with alpha levels of 0.05 (*), and 0.001 (***). Error bars represent ± 1 *SEM*.

Congruency

To investigate the role of congruency between the central face and the flankers, we compared congruent and incongruent trials as in Experiment 1. As the central face was always congruent with the flankers in the trials where the average emotion was 80% and 100%, we excluded these trials. The results of the congruency analysis showed a strong foveal input bias (Figure 6). A repeated-measures ANOVA with d′ as the dependent variable showed main effects of Congruency, $F(2, 34) = 37.70$, $p < 0.001$, partial $\eta^2 = 0.69$, Eccentricity, $F(2, 34) = 34.19$, $p < 0.001$, partial $\eta^2 = 0.67$, as well as an interaction between Congruency and Eccentricity, $F(4, 68) = 29.43$, $p < 0.001$, partial $\eta^2 = 0.63$. Similar as in Experiment 1, ensemble performance was worse in the incongruent compared to the congruent and central-absent condition at 0° (congruent (2.26 +/- 0.57) > central-absent (1.73 +/- 0.54): $p < 0.001$; congruent > incongruent (-0.50 +/- 0.94): $p < 0.001$; central-absent > incongruent: $p < 0.001$). At 3° eccentricity, there were no differences between the three conditions (congruent (1.36 +/- 0.65) vs. central-absent (1.31 +/- 0.57): $p = 0.99$; congruent vs. incongruent (1.41 +/- 0.99): $p = 0.10$; incongruent vs. central-absent: $p = 0.96$). At 8°, discriminability was overall low and the three conditions did not differ (congruent (0.76 +/- 0.56) vs. central-absent (0.55 +/- 0.36): $p = 0.17$; congruent vs. incongruent (0.37 +/- 0.68): $p = 0.19$; incongruent vs. central-absent: $p = 0.74$). In the central-absent condition, performance was best at 0°, worse at 3°, and worst at 8° (0° > 3°: $p < 0.001$; 0° > 8°: $p < 0.001$; 3° > 8°: $p < 0.001$). The pattern of results was the same in the congruent as in the central-absent condition (0° > 3°: $p < 0.001$; 0° > 8°: $p < 0.001$; 3° > 8°: $p < 0.001$). In the incongruent condition, however, performance was best at 3°, worse at 8°, and worst at 0° (3° > 0°: $p < 0.001$; 3° > 8°: $p < 0.01$; 8° > 0°: $p < 0.01$). Again, there was a tendency to report emotions as negative (M = -0.12 +/- 0.12). There were no main effects of Congruency, $F(2, 34) = 0.24$, $p = 0.79$, partial $\eta^2 = 0.01$, or Eccentricity, $F(2, 34) = 0.11$, $p = 0.89$, partial $\eta^2 = 0.01$, and no interaction, $F(4, 68) = 0.54$, $p = 0.70$, partial $\eta^2 = 0.03$.

Figure 6

Discriminability and Criterion for congruent, incongruent and central-absent stimuli



Congruency results of Experiment 2. Discriminability (a) and criterion (b), separated for face sets with congruent and incongruent central face and flankers, and without central face (central-absent condition). Asterisks indicate significance with alpha levels of 0.001 (***). Significance is only indicated for the comparisons of the three conditions (central-absent, congruent, incongruent) at each eccentricity. Error bars represent ± 1 *SEM*.

General Discussion

The current study investigated whether and to what extent foveal input biased responses in ensemble emotion perception. To test this, we compared the ensemble performance when presenting a foveal face with the performance when not presenting a foveal face. Experiment 1 showed that participants' ensemble performance was worse when there was foveal input (central-present) compared to no foveal input (central-absent). The poor performance in the central-present condition was due to the incongruent condition where the central face and the flankers required opposite responses. Experiment 2 used varying flankers and replicated the pattern of results of Experiment 1 (where flankers were uniform). The same pattern of results with uniform and with varying flankers indicated that ungrouping between the target and the flankers did not underlie the results. In both experiments, we presented the face set at different eccentricities (0°, 3°, 8°). An increase in eccentricity yielded the expected decrease of ensemble performance in all conditions without a central face. However, with a central face, performance at 0° was worse than at 3° in both experiments. At 8°, performance was poor with and without the central face. The pattern of results demonstrates that the foveal input strongly biased the ensemble performance when it was incongruent with that of flankers.

Overall, discriminability was similar for all conditions at each given eccentricity (except central-present at 0°). In particular, at 3°, discriminability was similar for the central-absent, central-present and single face condition, replicating typical findings in previous studies (e.g., Haberman & Whitney, 2007, 2009, Li et al., 2016). The same pattern of results was found at 8°, however, with clearly lower discriminability compared to 0° and 3°. Interestingly, even at 8°, performance was above chance level (63% percent correct in Exp.1; 65% percent correct in Exp.2), showing that facial expressions of single faces and groups of faces can still be extracted at relatively large eccentricities where visual resolution is reduced and crowding is strong. Consistent with previous studies that found an anger bias in the evaluation of crowd emotions (Becker et al., 2007; Neta et al., 2009; Mihalache et al., 2021), our criterion results showed a small trend to report the emotion of the face set as negative (Figure 2b, 5b).

The foveal input bias in the current study is consistent with prior demonstrations that foveal input weighs more in ensemble perception (Atkinson & Smithson, 2013; Ji et al., 2014; Jung et al., 2017). For example, Jung et al. (2017) found that foveal input was more strongly weighted in ensemble face race perception. In their study, a set of 12 faces (a 3 × 4 matrix subtending visual angles of 12° × 13°) was presented for 250 ms, and participants were required to adjust a probe face to the average race of the face set. The two central faces of the matrix were regarded as the foveal input. The results showed that the two faces presented foveally weighed more than the faces presented peripherally, suggesting that foveal (or close-to-foveal) input biased ensemble face race perception. Jung and colleagues (2017) suggested that participants could not scrutinize the faces in the face set consciously due to the short presentation durations and high number of stimuli. Rather, participants were unconsciously biased by the faces they were looking at directly. Unlike the study by Jung et al. (2017), we presented only a single face in the foveal location. Participants were required to fixate the very same location in which the foveal face was presented, ensuring that only one face was fixated directly. Presentation time was 100 ms, and thereby sufficiently short to prevent eye movements from the initially fixated (foveal) face to other faces. Hence, there was a clear distinction between fixated face and surrounding faces, making it more likely to notice the different capacities to extract information from the foveal and peripheral faces. Noticing this difference could have led to a strategy to give less weight to the foveal face when judging the ensemble. However, our results suggest that observers did not compensate for the prominent position of the foveal face, but judged the average emotion strongly biased by the foveal's face emotion. While it is unclear whether they did so unconsciously (Jung et al., 2017), we showed in a recent study that observers were capable to disregard the foveal input (at least to a large extent) and accurately estimate the emotion of the surrounding faces when they were asked to ignore the foveal face (Yu et al., 2021). Hence, it seems that while the foveal input bias is very strong without further instructions as in the current study, it is not ubiquitous but can be modulated by voluntary control.

More generally, the current results support weighted averaging in ensemble perception (e.g., Kanaya et al., 2018; Choi & Chong, 2020; Pascucci et al., 2021). According to weighted averaging, the relative contributions of members of the group are not equal when integrated into an ensemble. For instance, it has been shown that salient stimuli (Kanaya et al., 2018; Iakovlev & Utochkin, 2020; Goldenberg et al., 2020), attended stimuli (de Fockert & Marchant, 2008; Li & Yeh, 2017; Choi & Chong, 2020), and the stimuli seen first or last (Hubert-Wallander & Boynton, 2015) contributed more to the ensemble. One explanation of the foveal input bias is that attention increased the contribution of the foveal input (Wolfe et al., 2015; Jung et al., 2017). For instance, in deFockert & Marchant (2008), observers were required to report the average size of items while also locating either the largest or smallest item in the set. Observers' averaging judgments were shifted towards the sizes of the attended items, suggesting that greater statistical weights were assigned to them than to less attended items. Here, by presenting only a single face at fixation, instead of two or more faces, we sought to maximize attention to the foveal location. The results showed a pronounced foveal input bias, suggesting that attention to a single face at fixation strongly interferes with ensemble emotion perception.

Similar to Wolfe et al. (2015), our results showed that the foveal input was not necessary for ensemble emotion perception. In their study, observers freely viewed face stimuli for 1.5 seconds either with a central occluder that prevented viewing the faces foveally, or without any occluder. Observers indicated the average emotion of the entire set (24 faces). The results showed no difference between the two conditions, suggesting that foveal information was not necessary to extract the average emotion of the group. Interestingly, a recent study showed that observers overestimated the average emotion of a group of faces. This crowd-emotion-amplification effect (Goldenberg et al., 2021), was proposed to be due to attentional biases by faces with strong emotions which were fixated longer than less emotional faces. As the presentation time in the study by Wolfe and colleagues (2015) was not sufficient to fixate all faces, a similar effect as the crowd-emotion-amplification effect could have been expected in their study as well, resulting in stronger average emotion reports in the unoccluded condition. However, with the large number of faces, possible temporal dependencies (e.g., perception of emotional expressions, Libermann et al., 2018; perceived age of face stimuli, Manassi & Whitney, 2022), and the degree of emotional variance (e.g., separate stimuli for positive and negative emotions in Goldenberg et al., 2021, and mixed positive and negative emotions in Wolfe et al., 2015) of the presented faces, several factors could have modulated the averaging process, yielding different results. The basic foveal bias effect found here is consistent with the crowd-emotion-amplification effect: The emotions of fixated faces weighed more than those of faces that were not fixated.

How the foveal input bias manifests itself in more natural settings, such as social interactions, is an open question. In the current study, brief presentation times (i.e., 100 ms) assured that participants could not fixate multiple faces of the stimulus. This was similar in related studies using short presentation times where multiple faces were presented in the foveal region without the possibility to fixate more than one face directly (e.g., Ji et al., 2014). Jung et al. (2017) presented stimuli for 250 ms and asked participants to indicate the average race of the set of 12 faces varied in race. There were two faces in the center – possibly allowing the fixation of both of them, at least in some trials. It was found that ensemble face race judgments were biased by the average of the two foveal faces. However, as eye movements were not recorded, it remains unclear how the foveal input bias varied under different ways of fixating the stimulus (e.g., one or two faces, in between the faces). With longer presentation times that allow eye movements during stimulus presentation, multiple faces of the presented ensemble can be fixated. Recently, Ueda (2022) presented highly natural (i.e., color photographs of faces with external features) emotional (happy or angry) and neutral facial expressions for 1000 ms, and asked participants to report which expression appeared more frequently. The results showed that centrally presented faces weighed more than peripheral faces, suggesting a foveal input bias with multiple faces (interestingly, this was only the case when emotional, but not when neutral faces were presented in the foveal location; see also Yu et al., 2021). However, how fixation patterns interacted with the observed bias is not clear as no eye movements were recorded. Goldenberg et al. (2021) presented face sets consisting of 12 faces for 1000 ms, allowing participants to fixate multiple faces. Participants were asked to report the average emotion of the face set. Eye movements and fixations were recorded. The results showed that fixated faces weighed more than non-fixated faces, showing a clear foveal input bias with multiple fixated faces. When successively fixating multiple faces of a face set consisting of simultaneously presented, spatially distributed faces, some faces are fixated before others. To investigate how the order of fixated emotional facial expressions influenced ensemble judgments, Goldenberg et al. (2022) sequentially presented single faces with varying expressions and set sizes (e.g., 1-12 faces). It was found that ensemble judgments were less accurate with more (fixated) faces. Importantly, faces that were presented later in the stream weighed more strongly in the ensemble, revealing a recency effect in ensemble emotion perception (see also Hubert-Wallander & Boynton, 2015). Hence, it seems that to predict the perceived emotion of a group of faces, it is not only key to know which faces were fixated but also when they were fixated. Taken together, these results suggest that the foveal input bias is similar with multiple as with one fixated face(s), and that it can be modulated by factors such as the temporal order of fixated faces, the emotionality of the foveal face, and – as discussed above -- voluntary control.

We varied the eccentricity of our stimuli, presenting them at 0°, 3° and 8°. The presence or absence of the central face had different effects on performance at different eccentricities. At 0°, the central face resulted in the strong foveal input bias; at 3° and 8°, there was no effect of the central face. When no central face was presented, the face set (flankers) had an average eccentricity of about 2.5° in the 0° condition. Performance was superior in this condition compared to 3° (with or without a central face). This advantage could be due to several factors. In particular, at 3°, faces were presented randomly to the left or right, hence, shifts of attention between the two visual fields were necessary. Also, the eccentricities of the faces varied more strongly at 3° than at 0°. However, the face closest to fixation was positioned at 1.78° from fixation in the 0° condition, and closer - at 1.22° degrees - in the 3° condition. In Experiment 1, where the flankers were all identical, reporting the emotion of a single face was an accurate response for the ensemble. Hence, a strategy to report the emotion of the face closest to fixation would have yielded good performance. Nevertheless, performance was better at 0° where the closest face to fixation was farther away than at 3°. Importantly, in Experiment 2, where the flankers were heterogeneous, the same pattern of results was observed: A large discriminability difference between central face present and absent at 0°, no difference at 3° (and 8°), and better performance without a central face at 0° than at 3°. In contrast to Experiment 1, a strategy to report the emotion of the face closest to fixation would have been less advantageous as the average emotion could strongly deviate from individual faces in the set. Hence, it is unlikely that participants adopted a strategy to make ensemble judgments based on one single face's emotion. Note that the inward-outward asymmetry of crowding, with items on the side farther from fixation (outward) exerting stronger crowding than items at the closer side (inward) suggest that the face closest to fixation was crowded more strongly than the face farthest from fixation (Bouma, 1973; Petrov & Meleshkevich, 2011; Rummens & Sayim, 2021). Hence, a strategy to report the emotion of the face farthest from fixation – with a corresponding reduction of visual resolution - seems equally possible. The reasons outlined above for (not) using the innermost face remain the same.

Varying the flanker emotions in Experiment 2 also showed that the foveal input bias was not due to flanker homogeneity. The foveal input bias in Experiment 1 could have been due to the ungrouping between the uniform flankers and the unique central face. Grouping of items in the fovea (Malania et al., 2007; Sayim et al., 2008, 2010), in the periphery (Sayim & Cavanagh, 2013; Saarela et al., 2009; Manassi et al., 2012, 2013) and between the periphery and the fovea (Sayim et al., 2014) has been shown to strongly modulate performance in crowding paradigms (Herzog et al., 2015). Usually, strong grouping between a target and the flankers deteriorates performance compared to weak grouping (Banks et al., 1979; Malania et al., 2007; Sayim et al., 2010; Livne, 2010; Manassi et al., 2012). However, recently, strong target-flanker grouping has also been shown to improve performance compared to weak grouping when emergent features of target-flanker configurations contained target-relevant information (Melnik et al. 2018, 2020; Rummens & Sayim, 2022). In the present study, ungrouping would have made the central face stand out from the flankers, in particular in the incongruent conditions. Both, an improvement or a deterioration of performance, could be expected under strong ungrouping compared to weak ungrouping (at all three eccentricities). Improvement would be expected if the ungrouping enabled easier prioritizing of the flankers as overall, reporting the average flanker emotion was more accurate than reporting the central face's emotion. Deterioration would be expected if ungrouping reduced access to the flankers. Ungrouping and the "standing out" of the central face could underlie the foveal input bias. However, in Experiment 2, we found the same pattern of results as in Experiment 1. Because of their heterogeneity, grouping among the flankers - while still possible to some extent based on the arrangement of them - was not possible based on flanker identity, as the flankers' emotions varied (in contrast to Experiment 1). Hence, the results of Experiment 2 showed that (un)grouping of central face and flankers does not explain the foveal input bias. The same pattern of results was also found in Experiments 1 and 2 at the two eccentricities 3° and 8°, indicating that flanker homogeneity did not play any important role for averaging performance in the periphery. However, there was a trend for higher discriminability in the congruent compared to the incongruent condition at 3° eccentricity in Experiment 1, suggesting that ungrouping of the central face and the flankers could have led to reduced access to the flankers or prioritization of the central face, at least to some extent. Hence, ungrouping of the central face from the flankers might play a minor role in the periphery, however, the potential effect seems negligible.

Conclusion

The current study investigated if foveal input biased ensemble emotion perception. The results showed that the foveal input strongly biased participants' emotion perception of face ensembles. At 0°, performance was better when no face was presented at fixation (central-absent condition) compared to when a face was presented (central-present condition), showing a strong foveal input bias. The poor performance with foveal input was driven by the incongruent condition where the emotion of the foveal face strongly biased responses. We found interactions between eccentricity and central face absent/present conditions: A strong effect of the central face was only observed at 0°, but not at 3° and 8° eccentricity. Ungrouping of the central face from surrounding (identical) faces played – if at all – only a very minor role. Our results suggest that ensemble emotion perception may fail when salient target information is available in central vision.

Declarations

Funding

This work was supported by the CNRS.

Conflicts of interest/Competing interests

Part of this study was previously presented at the Annual Meeting of the Vision Sciences Society (VSS) 2021. We have no conflicts of interest to disclose.

Ethics approval

The approval was granted by the Ethics Committee of Soochow University.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

Participants have consented to the publication of their data.

Availability of data and materials

Data and additional online materials are available at the project's Open Science Framework page (https://osf.io/bzs2m/).

Code availability

The custom R analysis scripts generated during the study are available upon request.

Authors' contributions

B.S., L.J., and Y.R.D. designed the study and prepared the manuscript. Y.S. collected the data, and Y.R.D. analysed the data. All authors reviewed the manuscript.

References

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision research*, *83*, 25-39. https://doi.org/10.1016/j.visres.2013.02.018

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences*, *15*(3), 122–131. https://doi.org/10.1016/j.tics.2011.01.003

Atkinson, A. P., & Smithson, H. E. (2013). Distinct contributions to facial emotion perception of foveated versus nonfoveated facial features. *Emotion Review*, *5*(1), 30-35. https://doi.org/10.1177/1754073912457226

Banks, W. P., Larson, D. W., & Prinzmetal, W. (1979). Asymmetry of visual interference. *Perception & Psychophysics*, *25*(6), 447–456. https://doi.org/10.3758/BF03213822

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology, 92*(2), 179–190. https://doi.org/10.1037/0022-3514.92.2.179

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision research*, *13*(4), 767-782. https://doi.org/10.1016/0042-6989(73)90041-2

Cant, J. S., & Xu, Y. (2020). One bad apple spoils the whole bushel: The neural basis of outlier processing. *NeuroImage*, *211*,116629. https://doi.org/10.1016/j.neuroimage.2020.116629

Choi, Y. M., & Chong, S. C. (2020). Effects of Selective Attention on Mean-Size Computation: Weighted Averaging and Perceptual Enlargement. *Psychological Science*, *31*(10), 1261-1271. https://doi.org/10.1177/0956797620943834

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. https://doi.org/10.1016/S0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900. https://doi.org/10.1016/j.visres.2004.10.004

Dakin, S. C., & Watt, R. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192. https://doi.org/10.1016/S0042-6989(97)00133-8

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *JOSA A*, *18*(5), 1016-1026. https://doi.org/10.1364/JOSAA.18.001016

de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, *70*(5), 789-794. https://doi.org/10.3758/PP.70.5.789

Epstein, M. L., Quilty-Dunn, J., Mandelbaum, E., & Emmanouil, T. A. (2020). The outlier paradox: The role of iterative ensemble coding in discounting outliers. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(11), 1267–1279. https://doi.org/10.1037/xhp0000857

Goldenberg, A., Sweeny, T. D., Shpigel, E., & Gross, J. J. (2020). Is this my group or not? The role of ensemble coding of emotional expressions in group categorization. *Journal of Experimental Psychology: General*, *149*(3), 445. http://dx.doi.org/10.1037/xge0000651

Goldenberg, A., Weisz, E., Sweeny, T. D., Cikara, M., & Gross, J. J. (2021). The crowd-emotion-amplification effect. *Psychological science*, *32*(3), 437-450. https://doi.org/10.1177/0956797620970561

Goldenberg, A., Schöne, J., Huang, Z., Sweeny, T. D., Ong, D. C., Brady, T. F., ... & Gross, J. J. (2022). Amplification in the evaluation of multiple emotional expressions over time. *Nature Human Behaviour*, *6*(10), 1408-1416. https://doi.org/10.1038/s41562-022-01390-y

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753. https://doi.org/10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734. https://doi.org/10.1037/a0013899

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of vision*, *9*(11):1, 1-13. https://doi.org/10.1167/9.11.1

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics*, *72*(7), 1825–1838. https://doi.org/10.3758/APP.72.7.1825

Han, L., Leib, A. Y., Chen, Z., & Whitney, D. (2021). Holistic ensemble perception. *Attention, Perception, & Psychophysics*, *83*(3), 998-1013. https://doi.org/10.3758/s13414-020-02173-1

Herzog, M. H., Sayim, B., Chicherov, V., & Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of vision*, *15*(6):5, 1-18. https://doi.org/10.1167/15.6.5

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of vision*, *15*(4), 5-5. https://doi.org/10.1167/15.4.5

Iakovlev, A. U., & Utochkin, I. S. (2020). Roles of saliency and set size in ensemble averaging. *Attention, Perception, & Psychophysics*, *83* : 1251-1262. https://doi.org/10.3758/s13414-020-02089-w

Ji, L., Chen, W., & Fu, X. (2014). Different roles of foveal and extrafoveal vision in ensemble representation for facial expressions. In D. Harris (Ed.), International conference on engineering psychology and cognitive ergonomics (pp. 164–173). Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-07515-0_17

Jung, W., Bülthoff, I., & Armann, R. G. (2017). The contribution of foveal and peripheral visual information to ensemble representation of face race. *Journal of Vision*, *17*(13):11, 1-12. https://doi.org/10.1167/17.13.11

Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1879), 20172770. https://doi.org/10.1098/rspb.2017.2770

Liberman, A., Manassi, M., & Whitney, D. (2018). Serial dependence promotes the stability of perceived emotional expression depending on face similarity. *Attention, Perception, & Psychophysics*, *80*(6), 1461-1473. https://doi.org/10.3758/s13414-018-1533-8

Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X. (2016). Processing of individual items during ensemble coding of facial expressions. *Frontiers in psychology*, *7*, 1332. https://doi.org/10.3389/fpsyg.2016.01332

Li, K. A., & Yeh, S. L. (2017). Mean size estimation yields left-side bias: role of attention on perceptual averaging. *Attention, Perception, & Psychophysics*, *79*(8), 2538-2551. https://doi.org/10.3758/s13414-017-1409-3

Livne, T., & Sagi, D. (2010). How do flankers' relations affect crowding? *Journal of Vision*, *10*(3):1, 1–14. https://doi.org/10.1167/10.3.1

Luo, A. X., & Zhou, G. (2018). Ensemble perception of facial attractiveness. *Journal of vision*, *18*(8), 7-7. https://doi.org/10.1167/18.8.7

Macmillan, N. A., & Creelman, C. D. (2004). Detection theory: a user's guide. Mahwah: Lawrence Erlbaum.

Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *Journal of Vision*, *7*(2):1, 1-7. https://doi.org/10.1167/7.2.1

Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, *12*(10):13, 1–14. https://doi.org/10.1167/12.10.13

Manassi, M., Sayim, B., & Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *Journal of Vision*,*13*(13):10, 1–10. https://doi.org/10.1167/13.13.10

Manassi, M., & Whitney, D. (2022). Illusion of visual stability through active perceptual serial dependence. *Science advances*, *8*(2), eabk2480. 10.1126/sciadv.abk2480

Melnik, N., Coates, D. R., & Sayim, B. (2018). Emergent features in the crowding zone: When target–flanker grouping surmounts crowding. *Journal of vision*, *18*(9):19, 1-15. https://doi.org/10.1167/18.9.19

Melnik, N., Coates, D. R., & Sayim, B. (2020). Emergent features break the rules of crowding. *Scientific reports*, *10*(1), 1-10. https://doi.org/10.1038/s41598-019-57277-y

Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the national academy of sciences*, *111*(21), 7873-7878. https://doi.org/10.1073/pnas.1308674111

Mihalache, D., Lamer, S. A., Allen, J., Maher, M., & Sweeny, T. D. (2021). Anger bias in the evaluation of crowds. *Journal of Experimental Psychology: General, 150*(9), 1870–1889. https://doi.org/10.1037/xge0001025

Neta, M., Norris, C. J., & Whalen, P. J. (2009). Corrugator muscle responses are associated with individual differences in positivity-negativity bias. *Emotion*, *9*(5), 640 – 648. https://doi.org/10.1037/a0016819

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739–744. https://doi.org/10.1038/89532

Petrov, Y., & Meleshkevich, O. (2011). Locus of spatial attention determines inward–outward anisotropy in crowding. *Journal of Vision*, *11*(4), 1-11. https://doi.org/10.1167/11.4.1

Pascucci, D., Ruethemann, N., & Plomp, G. (2021). The anisotropic field of ensemble coding. *Scientific Reports*, *11*(1), 1-10. https://doi.org/10.1038/s41598-021-87620-1

Rummens, K., & Sayim, B. (2021). Broad attention uncovers benefits of stimulus uniformity in visual crowding. *Scientific Reports*, *11*(1), 1-13.

Rummens, K., & Sayim, B. (2022). Multidimensional feature interactions in visual crowding: When configural cues eliminate the polarity advantage. *Journal of Vision*, *22*(6):2, 1-19. https://doi.org/10.1167/jov.22.6.2

Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision*, *9*(2):5, 1–11.https://doi.org/10.1167/9.2.5

Sayim, B., Westheimer, G., & Herzog, M. H. (2008). Contrast polarity, chromaticity, and stereoscopic depth modulate contextual interactions in vernier acuity. *Journal of Vision*, *8*(8):12, 1–9. https://doi.org/10.1167/8.8.12

Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science*, *21*(5), 641–644. https://doi.org/10.1177/0956797610368811

Sayim, B., Cavanagh P (2013). Grouping and Crowding Affect Target Appearance over Different Spatial Scales. *PLoS ONE*, *8*(8): e71188. https://doi.org/10.1371/journal.pone.0071188

Sayim, B., Greenwood, J. A., & Cavanagh, P. (2014). Foveal target repetitions reduce crowding. *Journal of Vision*, *14*(6):4, 1–12. https://doi.org/10.1167/14.6.4

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, *10*(14):19,1-16. https://doi.org/10.1167/10.14.19

Sweeny, T. D., and Whitney, D. (2014). Perceiving crowd attention ensemble perception of a crowd's gaze. *Psychol. Sci*, *25*, 1903–1913. https://doi.org/10.1177/0956797614544510

To, M. P., Carvey, K. M., Carvey, R. J., & Liu, C. H. (2019). Averaging sets of expressive faces is modulated by eccentricity. *Journal of vision*, *19*(11):2, 1-14. https://doi.org/10.1167/19.11.2

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Ueda, Y. (2022). Understanding Mood of the Crowd with Facial Expressions: Majority Judgment for Evaluation of Statistical Summary Perception. *Attention, Perception, & Psychophysics*, *84*(3), 843-860. https://doi.org/10.3758/s13414-022-02449-8

Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision*, *15*(4):11, 1–13. https://doi.org/10.1167/15.4.11

Watamaniuk, S. N., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: the integration of direction information. *Vision Research*, *29*(1), 47–59. https://doi.org/10.1016/0042-6989(89)90173-9

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, *69*, 105– 129. https://doi.org/10.1146/annurev-psych-010416-044232

Yu, D., Song, Y., & Sayim, B. (2021). Voluntary control eliminates the fovea bias in ensemble emotion perception. *Perception*, *50*, 136-137.