Averaging sets of expressive faces is modulated by eccentricity

Michelle P. S. To

Department of Psychology, Lancaster University, Lancaster, UK
Department of Psychology, University of Hull, Hull, UK ✉

Katherine M. Carvey

Department of Psychology, University of Hull, Hull, UK ✉

Richard J. Carvey

Department of Psychology, University of Hull, Hull, UK ✉

Chang Hong Liu

Department of Psychology Bournemouth University, Poole, UK
Department of Psychology, University of Hull, Hull, UK ✉

Research has shown that participants can extract the average facial expression from a set of faces when these were presented at fixation. In this study, we investigated whether this performance would be modulated by eccentricity given that neural resources are limited outside the foveal region. We also examined whether or not there would be compulsory averaging in the parafovea as has been previously reported for the orientation of Gabor patches by Parkes, Lund, Angelucci, Solomon, and Morgan (2001). Participants were presented with expressive faces (alone or in sets of nine, at fixation or at 3° to the left or right) and were asked to identify the expression of the central target face or to estimate the average expression of the set. Our results revealed that, although participants were able to extract average facial expressions in central and parafoveal conditions, their performance was superior in the parafovea, suggesting facilitated averaging outside the fovea by peripheral mechanisms. Furthermore, regardless of whether the task was to judge the expression of the central target or set average, participants had a tendency to identify central targets' expressions in the fovea but were compelled to average in the parafovea, a finding consistent with compulsory averaging. The data also supported averaging over substitution models of crowding. We conclude that the ability to extract average expressions in sets of faces and identify single targets' facial expressions is influenced by eccentricity.

Introduction

Although the signals from the external environment are abundant and varied, the visual system can recognize its statistical properties and process such signals both effectively and economically. Neighboring features in natural scenes are often similar and to some degree predictable, and this is supported in the sparse coding behavior of V1 neurons that can accurately transmit information about the complex scenes with minimal redundancy and very little spikes (Barlow, 1961; Daugman, 1989; Field, 1987; Olshausen & Field, 1996; To, Baddeley, Troscianko, & Tolhurst, 2011; Vinje & Gallant, 2000).

This idea of compressing repetitive information into a simplified, more tractable ensemble representation has also been supported in behavioral studies (see Alvarez, 2011, and Whitney & Yamanashi Leib, 2018, for reviews). These studies have shown that, when presented with sets of similar objects, participants can accurately extract low-level statistical properties of the set, such as average size (Ariely, 2001; Chong & Treisman, 2003), but are poor at identifying the properties of individual objects. Likewise, Haberman and Whitney (2007, 2009) showed that this ability could also be extended to more complex elements, such as facial expressions. Their study revealed that participants were better at identifying the average emotional expression in a set of expressive faces compared to recognizing a specific set member. Other researchers

have further demonstrated that participants could estimate the average identity in a set of faces (J. de Fockert & Wolfenstein, 2009; J. W. de Fockert & Gautrey, 2013), the average direction of a crowd (Sweeny, Haroz, & Whitney, 2013), and the average lifelikeness of a group of objects (Leib, Kosovicheva, & Whitney, 2016). Given the high degree of repetition in the natural environment, ensemble coding represents an economical approach that enables participants to rapidly gather a general impression of a scene. H. Li et al. (2016) examined the effect of exposure time on the ability to identify faces and their emotions and found that participants were more sensitive to set representations compared to individual representations for short presentations (e.g., 50 ms). However, the difference disappeared for longer durations (500–2,000 ms). This suggests that ensemble coding is mostly relied upon when resources are limited.

Given that neural resources are limited outside central vision (Adams & Horton, 2002, 2003; Horton & Hoyt, 1991; Inouye, 1909; Lister & Holmes, 1916), it should not be surprising to observe averaging of features in parafoveal and peripheral visual areas (e.g., Greenwood, Bex, & Dakin, 2009, 2010; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Averaging in the periphery has been primarily attributed to mechanisms associated with crowding, a visual phenomenon whereby signals of surrounding stimuli can crowd out signals of a single target, therefore, preventing and/or interfering with its identification and recognition (Bouma, 1970, also see Levi, 2008, and Pelli & Tillman, 2008, for reviews). Although crowding effects have been extensively studied, from orientation identification of artificial stimuli (e.g., Greenwood et al., 2009, 2010; Parkes et al., 2001) to feature discriminations in natural scenes (e.g., To, Gilchrist, Troscianko, Kho, & Tolhurst, 2009; To, Gilchrist, Troscianko, & Tolhurst, 2011), its underlying mechanisms remain largely speculative, and theories do not always reflect the known behavior of neurons. A few neurophysiological studies of V1 have compared single neurons with peripheral and central (cat) or foveal (primate) receptive fields. Generally, it has been found that peripheral neurons have larger receptive fields and respond to corresponding lower spatial frequencies of grating (Hubel & Wiesel, 1974; Tolhurst & Thompson, 1981; H. H. Yu et al., 2010). These findings may explain how simple visual acuity falls off with increasing eccentricity, but only subtle differences between foveal and peripheral neural processing have been reported in V1 (H. H. Yu & Rosa, 2014), and these do not seem to explain phenomena, such as veneer acuity or crowding (Levi, Klein, & Aitsebaomo, 1985). Although it used to be well accepted that cortical magnification factors were closely related to retinal sampling density and receptive field size (e.g., Hubel & Wiesel, 1974), many

studies since have reported that foveal V1 is greatly enlarged, even allowing for its smaller receptive fields (Adams & Horton, 2003; Popovic & Sjöstrand, 2001; Tolhurst & Ling, 1988). This implies that foveal neural processing has more dimensions than peripheral, but this finding is not an explanation for crowding or the poverty of peripheral visual perception. We are not aware of any single neuron studies that explicitly investigate responses with stimulus paradigms that cause crowding in human perception. Nonetheless, the effects of crowding have been successfully modeled by spatial uncertainty and substitution (Freeman, Chakravarthi, & Pelli, 2012; Krumhansl & Thomas, 1977; Nandy & Tjan, 2007; Pelli, 1985; Põder & Wagemans, 2007; Popple & Levi, 2005; Strasburger, 2005; Wolford, 1975), imprecise integration (Neri & Levi, 2006; van den Berg, Roerdink, & Cornelissen, 2010), limited field of view (Tjan, 2009), and averaging (also referred to as signal pooling; Freeman et al., 2012; Greenwood et al., 2009, 2010; Parkes et al., 2001) in behavioral studies.

This averaging is not dissimilar to ensemble coding that is reported in central vision. Both processes promote a more efficient statistical representation of global information in a scene while reducing the processing of any redundant or finer-detailed information. Parkes et al. (2001) reported compulsory averaging of features for stimuli presented in the parafovea. Although their participants were unable to identify the orientation of a central target patch, they were able to estimate the average orientation of all patches. Research by Greenwood et al. (2009, 2010) examining the effects of flankers on participants' ability to estimate a target's position and on their ability to detect a change in target appearance lend further support to the averaging of target and flanker information in the periphery (although, in the case of target appearance, target–flanker substitution mechanisms could have also played a part). This signal-averaging mechanism in parafoveal and peripheral vision and the aforementioned ensemble coding reported in central vision are, therefore, quite similar in terms of their function and their effect on viewers' ability to process sets but not individual objects. However, the averaging reported by Greenwood et al. (2009, 2010) and Parkes et al. were confined to low-level features (orientation).

To our knowledge, this is the first study that compares averaging of facial expression in central and parafoveal vision. Our method was largely adopted from Parkes et al. (2001), who made a similar comparison for orientation of Gabor-patch stimuli. In our experiment, we investigated whether participants could extract the average representation of emotional cues from sets of faces presented in central and parafoveal vision and whether their performance would be modulated by eccentricity. We were also interested

in whether or not there was compulsory averaging in the parafovea as was reported in Parkes et al.

We asked participants to complete two tasks. In the central task, they identified the expression of a single central target face. In the average task, they estimated the average emotional expression of a set of faces. We compared participants' performance in the two tasks at different eccentricities. Similar performances throughout would suggest similar mechanisms underlying averaging across the visual field, whereas differences in performances would suggest otherwise. If performance in the average task improved at higher eccentricities, this would suggest that the processing of averages is enhanced in the parafovea. The expressive faces were presented alone (flanker absent) or in sets of nine (flanker present). This manipulation allowed us to disentangle different mechanisms underlying averaging in central and parafoveal areas. The absence of interference by flanking stimuli in central vision suggests that averaging at the fovea differs from peripheral mechanisms.

We then compared participants' responses with the predictions of three simple computational models:

- A central model that predicts responses based on the central target alone.
- An average model that is associated with the aforementioned averaging of signals and that predicts responses based on the set average.
- A substitution model that is associated with spatial uncertainty and target–flanker swap and that predicts responses based on the flankers' properties.

The central and average models allowed us to examine whether participants followed the instructions and completed the assigned task (central vs. average) or whether they were compelled to respond differently. For example, if participants were asked to identify the emotional expression of the central target, the central model should be best at predicting their results as it returns the emotion of the central target. Likewise, if they were asked to estimate the average emotional expression of all faces, then the average model should be the strongest. However, if the "best" models and tasks did not match up, then the data would suggest that the participants were not following instructions and were instead compelled to enter the expression of the central target (in the average task) or the average (in the central task).

We also added the substitution model so that we can compare its performance with that of the average model as this could offer further insight into the mechanisms that underlie crowding in the extrafoveal region. As mentioned above, several mechanisms have been proposed to play a part in crowding in the periphery, two of which are signal averaging (Freeman

et al., 2012; Greenwood et al., 2009, 2010; Parkes et al., 2001) and spatial uncertainty (target–flanker substitution; Freeman et al., 2012; Krumhansl & Thomas, 1977; Nandy & Tjan, 2007; Pelli, 1985; Põder & Wagemans, 2007; Popple & Levi, 2005; Strasburger, 2005; Wolford, 1975). Our interest in ensemble coding and signal averaging propels us to model the averaging of features (average model), but given that recognition of facial expressions is highly complex and holistic, comprising of low- and higher-level features (e.g., D. Yu, Chai, & Chung, 2018), we thought it worthwhile to also consider an unpooled substitution model, which may be more suitable for stimuli whose features are processed as a whole.

Methods

Participants

A total of 18 university students (14 females) participated in this experiment (Age: $M = 23.5$ years, $SD = 6.91$). Informed consent was obtained for all participants, and all had normal or corrected-to-normal vision. This study was approved by the research ethics committee at the University of Hull.

Stimuli and materials

The face stimuli were created from the face database developed at Binghamton University (Yin, Wei, Sun, Wang, & Rosato, 2006). The database contained male and female faces showing neutral and emotional facial expressions (happy, sad, surprised, angry, disgusted, and fearful). All external features, such as hair and glasses, were removed from the faces.

We used three original female faces (from a single model) with neutral, disgusted, and happy expressions and used FantaMorph (2009) to generate a series of faces with intermediate emotions. We selected happy and disgusted faces because these expressed positive and negative emotions, respectively. We chose disgust over sadness because research has shown that happiness and sadness elicit different intensities of arousal, but no such differences have been reported between happiness and disgust (e.g., Schwartz & Davidson, 1997).

The original disgusted and happy faces were at the ends of the face expression continuum and were labeled 100% disgusted and 100% happy, respectively. We created four new faces with various levels of disgust by morphing the 100% disgusted expression to a neutral expression: 80% disgusted with 20% neutral, 60% disgusted with 40% neutral, 40% disgusted with 60%
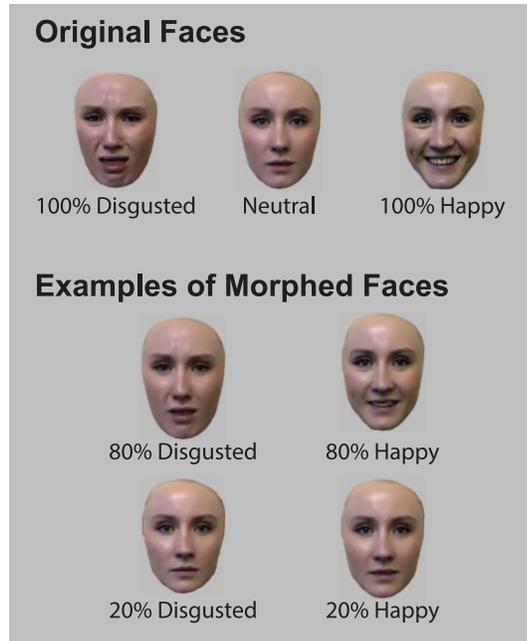
Figure 1. Original faces and examples of morphed faces. The original 100% disgusted, neutral, and 100% happy faces (top three) were taken from the BU-3DFE database (permission to use can be found at http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html). These were morphed using FantaMorph (2009) to generate intermediate facial expressions (bottom four).
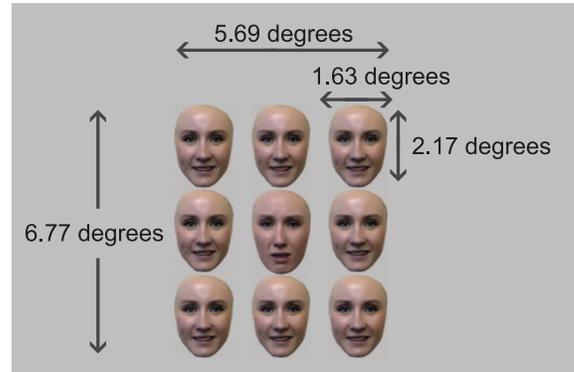

Figure 2. Example of faces presented in a set. This set shows a central target (disgusted) face surrounded by eight flanker (happy) faces. Each face subtended $1.63° \times 2.17°$ of visual angle, and was located within 0.40° horizontally and 0.15° vertically of one another. The whole set covered an area of $5.69° \times 6.77°$. The original 100% disgusted, neutral, and 100% happy faces were taken from the BU-3DFE database and permission to use can be found at http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html.

neutral, and 20% disgusted and 80% neutral. Similarly, we repeated the same procedure to create four levels of happy faces. Figure 1 presents the original faces as well as an example of morphed faces. There were 11 faces in total: three originals (100% disgust, neutral, and 100% happy) and the eight morphed variants listed above.

All faces were displayed on a 22-in. monitor (IiyamaProlite E2200WS). The display resolution of the experiment was set to $1,024 \times 768$ for the experiment, and the background color of the screen was set to gray (Hex: #C0C0C0). E-Prime 1.1 (PST, 2003) was used to run the experiment, display the stimuli, and the record responses.

Design

We examined the effect of three within-subject factors: flankers, eccentricity, and task. Each of these factors were manipulated as follows:

Flankers (absent or present)

In order to study the effect of flankers on participants' ability to identify facial expression(s), a central face was either presented alone (flanker absent) or surrounded by eight faces (all sharing the same

expression) in a 3×3 configuration (flanker present; see Figure 2). The faces in the flanker present condition were located within 0.40° horizontally (2.03° center to center) and 0.13° vertically (2.30° center to center) of each other and the central target, giving the set an overall set dimension of $5.69° \times 6.77°$. According to Bouma (1970), crowding at $\varphi°$ in eccentricity occurs when flankers are located within $\varphi/2°$ of the targets. Although the faces are located 2.30° center to center (beyond 1.5° if $\varphi = 3$), they are nonetheless within 0.4° of each other side to side and 1.21° from center to side of the next horizontal face. All 11 variants of expressions were presented as central targets in the flanker-absent condition and as central targets and/or flankers in the flanker-present conditions. The example in Figure 2 shows an 80% disgusted central face surrounded by eight identical 80% happy flanker faces.

The order of the flanker conditions was counterbalanced such that half the participants started with the flanker-absent condition and the other half started with one of the flanker-present conditions.

Eccentricity (fovea, left, or right)

The parafoveal target sets in Parkes et al.'s (2001) paper were presented at 2.5° to the left and right of fixation. In addition, Calvo, Nummenmaa, and Avero (2010) showed that observers were able to recognize happy and disgusted faces that were presented for 150 ms at 2.5° although their faces were larger ($8.4° \times 6.4°$) and were presented in isolation. A separate study by Goren and Wilson (2006) showed that observers could recognize happy expressions that were presented for

110 ms at 5.5°–5.8° in eccentricity although, again, the faces were larger than ours (6.9° × 9.1°) and were presented without flankers. Based on these studies, we have chosen to present faces foveally and at 3° parafoveally either to the left or to the right of a central fixation to control for any hemispheric bias. In the flanker-present condition, the central face was centered at 3° eccentricity, a location that satisfies the traditional bounds for crowding (Bouma, 1970). The three positions of eccentricity were randomized to avoid successive left or right presentations that might increase the likelihood of participants foveating toward the faces ahead of them appearing.

Task (central or average)

Participants were asked to either identify the expression of the central faces (central) or to estimate the average expression of the sets of faces (average). Although the former task was completed in both flanker-absent and -present conditions (in randomly ordered blocks), the latter task was only conducted in the flanker-present conditions (as there was no point asking participants to estimate the average of only one face).

Procedure

Each participant was tested individually and was asked to complete the two tasks (central and average). An adjustable headrest was used to fix the participant's eye height and viewing distance, which was set at 57 cm away from the screen.

The experiment was divided into three blocks: (a) identify central target in flanker-present conditions, (b) estimate average expression in flanker-present condition, and (c) identify central target in flanker-absent condition. The order of the blocks was counterbalanced for each participant. Before each testing block, the experimenter instructed participants of the task they had to complete (central or average) and asked them to complete 10 practice trials. These were then followed by the test trials. In the flanker-present conditions, each of the 11 faces was presented as central targets and as flankers for each of the three locations, resulting in 363 trials (11 faces as central targets × 11 faces as flankers × 3 locations) in blocks 1 and 2. In the flanker-absent condition, in order to match the number of times each central target face was presented in the flanker-present condition, each of the 11 central faces was presented 11 times at each location, resulting in a total 363 trials (11 faces × 11 times × 3 locations). So participants completed 1,089 trials in total (363 trials/block × 3 blocks).

On each trial, a fixation screen (with a cross in the middle) was presented for 500 ms. Then a second screen presented either a single central face (flanker absent) or a central face surrounded by eight flankers (flanker present). The faces were presented at the fovea or at 3° to the left or right of the fixation cross for 100 ms. Afterward, a third response screen appeared, prompting the participants to enter their responses (following the task they were given) on a scale from zero (disgusted) to 10 (happy) with five corresponding to neutral. This response screen was displayed until a response was entered. Figure 3 shows the fixation screen (top left), examples from flanker absent (fovea and right, top left and right inset), and flanker present (fovea and right, bottom left and right inset) blocks and the response screen.

Measuring participants' performance

To examine participants' ability to identify the expressions of a central target face or the average expressions of sets of faces, we computed their overall mean squared error (MSE) for each condition by first subtracting participants' response from the actual expression(s) presented on each trial, squaring the differences between each "actual − measured" pair, summing the squared differences, and then dividing the whole by the total number of responses/trials:

$$MSE = \sum_{i=1}^{n} \frac{(\text{Participant's Response}_i - \text{Actual Expression}_i)^2}{n}, \quad (1)$$

where i refers to the trial number and n is the total number of responses. Lower MSE values, smaller error, correspond to stronger performances. We selected MSE values to describe participants' performance because our aim was to identify how much participants' responses deviated away from actual expressions and not how these could be biased toward happy or disgusted faces.

Modeling

To examine whether participants were more compelled to only rate the central target or estimate averages in foveal and parafoveal vision and to determine which model is best at describing any potential effects of crowding of emotional features, we compared the performance of three models: the central, average, and substitution models. The central model returns the facial expression of the central face alone and is unaffected by the presence of the flankers. The
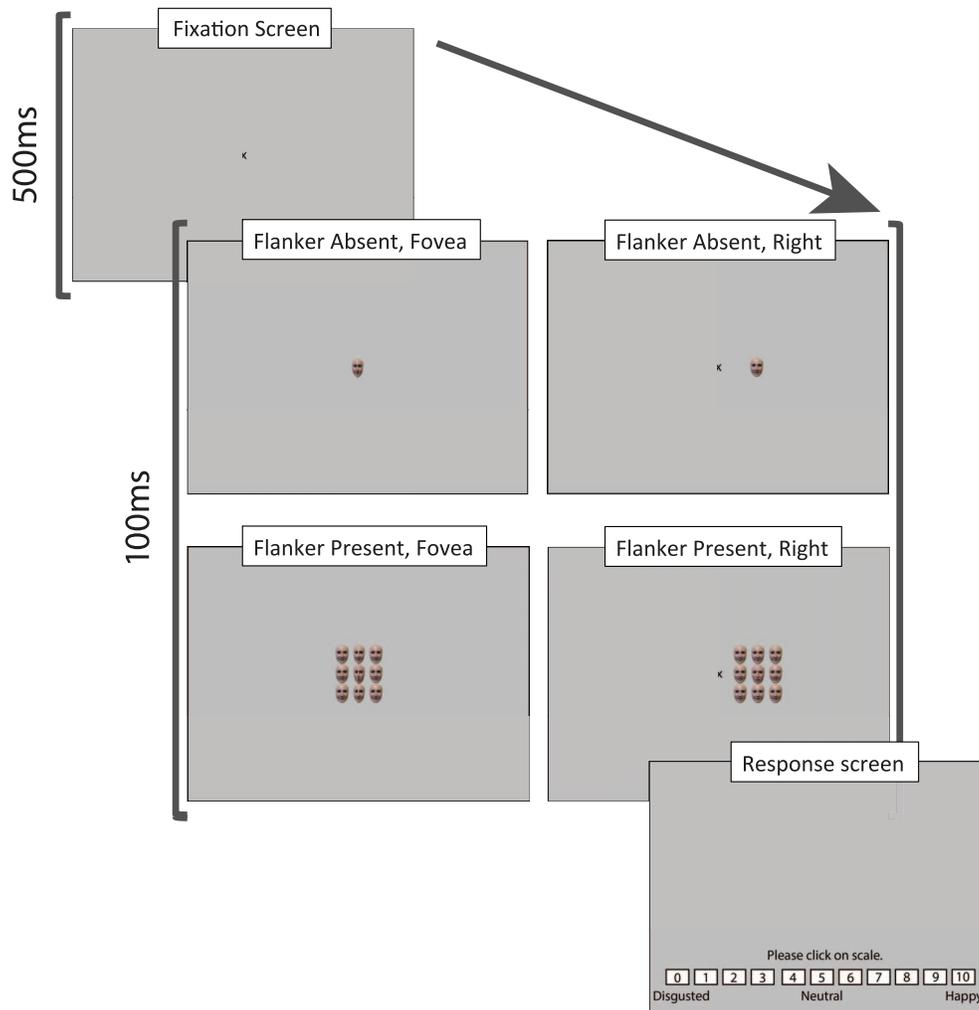
Figure 3. Sample trials from the experiment. The fixation screen was first presented for 500 ms. Then participants were presented with a single face (top row inset) or with a set of nine faces (bottom row inset) for 100 ms; the faces were displayed in the fovea or with the central face at 3° to the left or right of fixation. This was followed by the response screen. The original 100% disgusted, neutral, and 100% happy faces were taken from the BU-3DFE database and permission to use can be found at http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html.

average model computes the average facial expression of all the faces in the set and is, therefore, significantly affected by the presence of flankers. And the substitution model returns the facial expression of the flanker faces (which are all identical) and is unaffected by the expression of the central target. Please note that the three models are not independent as the average model is a composite of the central and substitution models, and its output equals

$$\frac{(\text{Central Model output} + 8 \times \text{Substitution Model output})}{9}. \quad (2)$$

This means that a superior performance in the average model would reflect a combined contribution of the central and substitution models.

We then considered model performance by looking at the errors between model predictions and participants' measured responses. The MSE was calculated by subtracting model predictions from participants' responses, squaring the differences, summing these together and then dividing the total by the number of responses:

$$MSE = \sum_{i=1}^{n} \frac{(\text{Model Prediction}_i - \text{Participant's Response}_i)^2}{n}, \quad (3)$$

where i refers to the trial number and n is the total number of responses. Stronger model performances are reflected by lower MSE values.

Results

Measuring error

We used MSE to describe the deviation between participants' responses from the actual expressions or correct answers. However, to ascertain whether the measured MSEs were indicative of performance above chance levels, i.e., participants accomplished the tasks with some degree of error versus participants were unable to complete the tasks and were forced to enter random values, we used a Monte Carlo simulation to generate 180,000 simulated data sets and then to calculated the MSEs for the different conditions, which averaged 20.00 with standard error values under 0.006. We set this value as the chance performance level. None of the MSEs measured were equal or beyond 18, so we concluded that participants could perform at above-chance levels and that the measured MSEs were reasonable estimates of accuracy and performance.

Effect of flankers

To examine the effects of flankers on participants' ability to identify expressions of a target face across eccentricities, we compared how they performed in rating the central targets' expression when these were presented alone (flanker absent) and among eight other faces (flanker present). In addition, we also considered performance when participants were asked to estimate the average expression from a set of faces. Performance on the tasks was evaluated using MSE (see Methods), so the better the performance, the lower the MSE values.

The data are presented in Figure 4. We ran a 2×3 within-subject ANOVA with two factors: flankers (absent vs. present) and eccentricity (foveal, left vs. right). There was a significant effect of flanker, $F(1, 17) = 38.58$, $p < 0.001$, $\eta^2 = 0.69$. When asked to identify the central targets' expressions, participants were more accurate when target faces were presented alone ($MSE = 7.24$, $SD = 0.37$) compared to when they were presented among flankers ($MSE = 10.62$, $SD = 0.52$).

There was a significant effect of eccentricity on participants' performance, $F(2, 34) = 34.05$, $p < 0.001$, $\eta^2 = 0.67$. Responses were more accurate in the foveal condition ($MSE = 6.15$, $SD = 0.43$) compared to parafoveal, left vs. right ($MSE = 10.42$, $SD = 0.55$ and $MSE = 10.20$, $SD = 0.50$, respectively). Bonferroni post hoc tests revealed significant differences between foveal and left ($d = 0.93$) and foveal and right ($d = 0.74$) but none between left and right ($d = 0.06$).

There was also a significant interaction between flankers and eccentricity, $F(2, 34) = 23.67$, $p < 0.001$, η^2
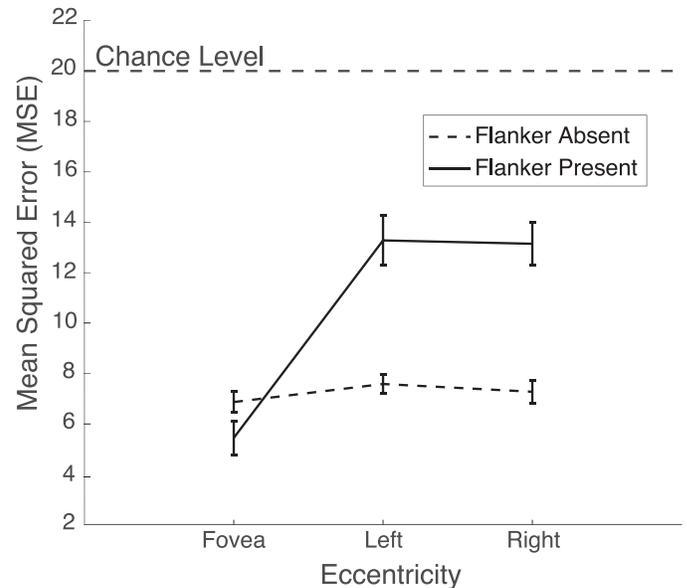

Figure 4. The MSE in the flanker-absent and -present conditions (in solid and dashed lines, respectively) across the three eccentricities. Error bars represent ± 1 *SEM*.

$= 0.58$. To determine the cause(s) of this interaction, we examined two sets of simple main effects. First, we compared the effect of flankers at each eccentricity. In foveal vision, participants' performances were similar in the flanker-present and -absent conditions, $F(1, 17) = 3.84$, $p > 0.05$ ($MSE = 6.88$, $SD = 0.41$ and $MSE = 5.44$, $SD = 0.68$, respectively). However, in the parafovea, flankers had a detrimental effect on overall performance: Errors were significantly higher in the flanker-present condition compared to the flanker-absent condition: $F(1, 17) = 32.78$, $p < 0.01$, $\eta^2 = 0.66$ and $F(1, 17) = 40.15$, $p < 0.01$, $\eta^2 = 0.70$ for left and right, respectively. At left and right, the data points along the solid line (flanker present) are higher compared to the points on the dashed line (flanker absent); Cohen's d values for left and right conditions were 1.35 and 1.50, respectively.

Second, we compared the effect of eccentricity on the flanker-absent and -present conditions. When participants were asked to identify the central targets' expressions without flankers (dashed black line in Figure 4), their performance was the same across the different eccentricities, $F(2, 34) = 2.75$, $p = 0.08$, $\eta^2 = 0.21$. When participants were asked to identify the central targets' expressions among flankers (solid black line in Figure 6), there was a significant effect of eccentricity, $F(2, 34) = 30.52$, $p < 0.001$, $\eta^2 = 0.64$. Bonferroni post hoc analyses revealed that performance was better in the fovea ($MSE = 5.44$, $SD = 0.68$) compared to the left ($MSE = 13.27$, $SD = 0.97$, $d = 2.04$) and right ($MSE = 13.14$, $SD = 0.85$, $d = 1.36$). There was no difference between the left and right conditions
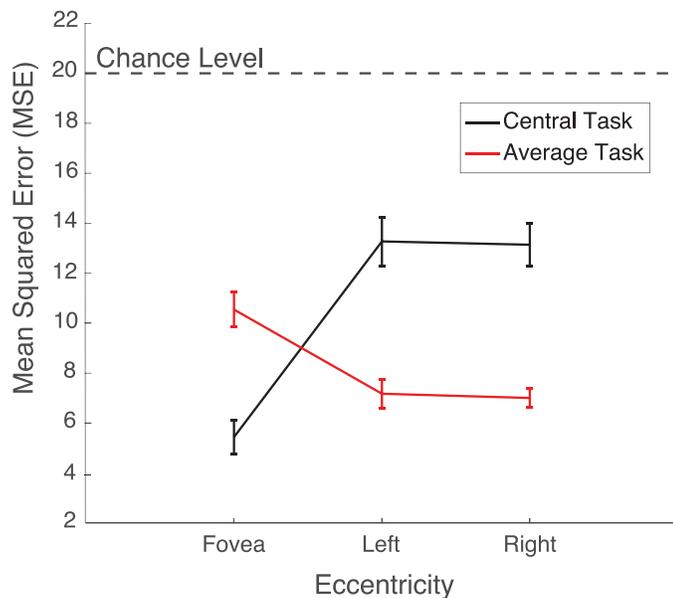
Figure 5. The MSE in performance in the central and average task conditions (in black and red, respectively) across the three different eccentricities, foveal, left, and right. Error bars represent ± 1 *SEM*.

($d = 0.02$). This indicates that both factors flanker and eccentricity affected each other.

Central versus average tasks

Next, we compared participants' ability to identify the expressions of central targets among flankers (central) and their ability to estimate the average expression of a set of faces (average). We then ran a 2×3 within-subject ANOVA with two factors: task (central vs. average) and eccentricity (foveal, left vs. right). The data is presented in Figure 5.

We found a significant effect of task, $F(1, 17) = 12.53$, $p < 0.005$, $\eta^2 = 0.42$, where participants were more accurate on the average condition ($MSE = 8.24$, $SD = 0.41$) compared to the central condition ($MSE = 10.62$, $SD = 0.52$). There was a significant effect of eccentricity, $F(2, 34) = 14.73$, $p < 0.001$, $\eta^2 = 0.46$: Participants' performance was superior in foveal vision ($MSE = 7.99$, $SD = 0.40$) compared to left and right ($MSE = 10.22$, $SD = 0.44$ and $MSE = 10.07$, $SD = 0.43$, respectively; $d = 0.33$ and 0.27, respectively). There were no significant differences between the two parafoveal conditions ($d = 0.04$).

There was also a significant interaction between task and eccentricity, $F(2, 34) = 31.38$, $p < 0.001$, $\eta^2 = 0.65$. To determine the cause(s) of this interaction, we examined two sets of simple main effects. First, we compared the effect of task at each eccentricity. When faces were presented in the fovea, participants were better at rating the central expression ($MSE = 5.44$, $SD = 0.68$) compared to estimating the average of the sets of faces ($MSE = 10.54$, $SD = 0.70$), $F(1, 17) = 20.81$, $p < 0.001$, $\eta^2 = 0.55$. When faces were presented in the parafovea, participants were poorer at rating the central expression ($MSE = 13.27$, $SD = 0.70$ and $MSE = 13.12$, $SD = 0.85$ for left and right, respectively) compared to estimating the average of the sets of faces ($MSE = 7.18$, $SD = 0.57$ and $MSE = 7.01$, $SD = 0.37$ for left and right, respectively), $F(1, 17) = 21.01$, $p < 0.001$, $\eta^2 = 0.55$ and $F(1, 17) = 37.48$, $p < 0.001$, $\eta^2 = 0.67$ for left and right.
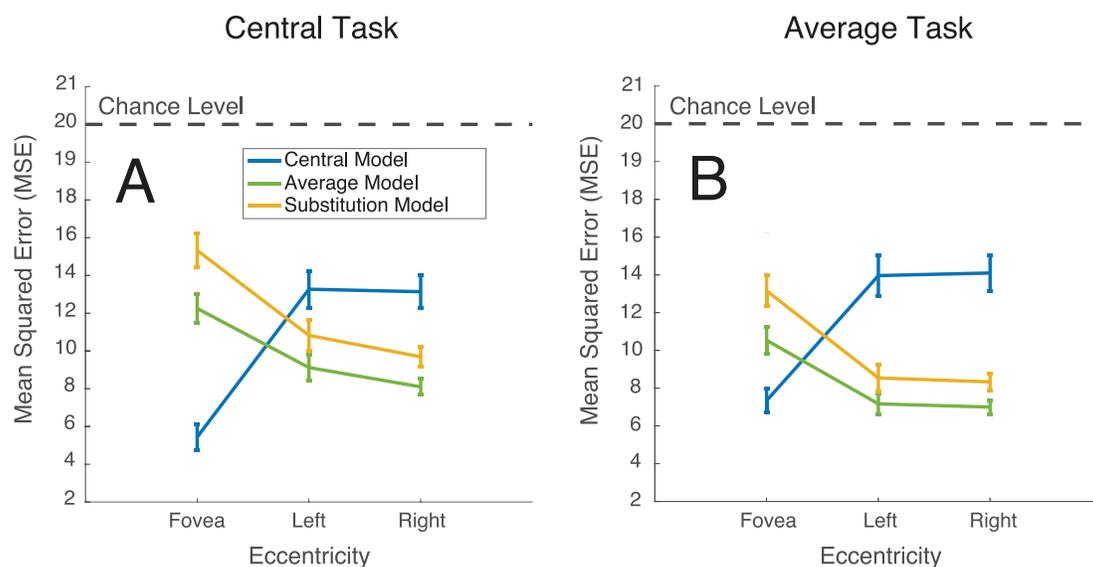

Figure 6. Model performances for the central and average tasks (panels A and B, respectively) across the three different eccentricities, foveal, left, and right. The MSE for the central, average, and substitution models are presented in blue, green, and yellow, respectively. Error bars represent ± 1 *SEM*.

Second, we compared the effect of eccentricity on each task. When participants were asked to identify the expressions of central targets that were presented among flankers (black line in Figure 5), their performance was significantly better in the fovea ($MSE =$ 5.44, $SD = 0.68$) compared to left and right ($MSE =$ 13.27, $SD = 0.97$ and $MSE = 13.14$, $SD = 0.85$, respectively), $F(2, 34) = 30.52$, $p < 0.001$, $\eta^2 = 0.64$. When participants were asked to estimate the average expression of a group of faces (red line in Figure 5), their performance was significantly better in the left and right ($MSE = 7.18$, $SD = 0.57$ and $MSE = 7.01$, $SD = 0.37$, respectively) compared to the fovea ($MSE =$ 10.54, $SD = 0.70$), $F(2, 34) = 18.04$, $p < 0.001$, $\eta^2 = 0.52$. This indicates that both factors task and eccentricity affect each other.

Comparing central, average, and substitution models

We examined the performance of three models (central, average, and substitution) in predicting participants' responses in the two different tasks. For each task, we ran a 3×3 within-subject ANOVA with factors model (central, average, and substitution) and eccentricity (foveal, left, and right).

Central task

The data is presented in Figure 6A. There was a significant model $\times$ eccentricity interaction, $F(4, 68) =$ 31.30, $p < 0.001$, $\eta^2 = 0.65$. First, we compared the effect of model at each eccentricity. In the fovea, there was a significant effect of model, $F(2, 34) = 47.55$, $p <$ 0.001, $\eta^2 = 0.74$. Bonferroni post hoc tests revealed that the central model was the most accurate in predicting participants' responses ($MSE = 5.44$, $SD = 0.68$) and was superior to the average and substitution models ($MSE = 12.25$, $SD = 0.75$ and $MSE = 15.33$, $SD = 0.88$, $d = 1.33$ and 1.72, respectively). The average model, in turn, was significantly better than the substitution model ($d = 4.73$). In the parafovea, model also had a significant effect, $F(2, 34) = 5.84$, $p = 0.007$, $\eta^2 = 0.26$ and $F(2, 34) = 15.49$, $p < 0.001$, $\eta^2 = 0.48$, for left and right, respectively. In the left, the average model ($MSE = 9.12$, $SD = 0.68$) was superior to both the central and substitution models ($MSE = 13.27$, $SD = 0.97$ and $MSE = 10.83$, $SD = 0.82$, respectively; $d = 0.68$ for average vs. central, and 2.23 for average vs. substitution), and there were no differences between the central and substitution models ($d = 0.66$). However in the right, the average model ($MSE = 8.10$, $SD = 0.43$) was superior to both the central and substitution models ($MSE = 13.14$, $SD = 0.85$ and $MSE = 9.69$, $SD = 0.52$, respectively; $d =$ 1.09 for average vs. central, and 2.74 for average vs.

substitution), and the substitution model was better than the central model ($d = 0.66$).

Second, we compared the effect of eccentricity for each model. For the central model (in blue in Figure 6A), there was a significant effect of eccentricity, $F(2, 34) = 30.52$, $p < 0.001$, $\eta^2 = 0.64$: Performance was significantly stronger in the fovea ($MSE = 5.44$, $SD = 0.68$) compared to the left and right conditions ($MSE = 13.27$, $SD = 0.97$ and $MSE = 13.14$, $SD = 0.85$, respectively; $d = 2.04$ and 1.36, respectively). Results of the left and right conditions were comparable ($d = 0.02$). For the average model (in green in Figure 6A), there was a significant effect of eccentricity, $F(2, 34) = 21.60$, $p < 0.001$, $\eta^2 = 0.56$: Performance was significantly weaker in the fovea ($MSE = 12.25$, $SD = 0.75$) compared to the left and right conditions ($MSE = 9.12$, $SD = 0.68$ and $MSE = 8.10$, $SD = 0.43$, $d = 1.33$ and 1.30, respectively). No difference was found between left versus right ($d = 0.36$). For the substitution model (in yellow in Figure 6A), there was a significant effect of eccentricity, $F(2, 34) = 24.86$, $p < 0.001$, $\eta^2 = 0.59$: As was in the case of the average model, performance was significantly weaker in the fovea ($MSE = 15.23$, $SD = 0.86$) compared to the left and right conditions ($MSE = 10.83$, $SD = 0.82$ and $MSE = 9.69$, $SD = 0.52$, $d = 1.49$ and 1.37, respectively). No difference was found between results of left versus right, $d = 0.30$.

Average task

There was a significant model $\times$ eccentricity interaction, $F(4, 68) = 26.37$, $p < 0.001$, $\eta^2 = 0.61$. The data is presented in Figure 6B. First, we compared the effect of model at each eccentricity. In the fovea, there was a significant effect of model, $F(2, 34) = 17.73$, $p < 0.001$, $\eta^2 = 0.74$. Bonferroni post hoc tests revealed that the central model was the most accurate in predicting participants' responses ($MSE = 7.35$, $SD = 0.62$) and was superior to the average and substitution models ($MSE = 10.54$, $SD = 0.70$ and $MSE = 13.17$, $SD = 0.82$, $d = 0.66$ and 1.06, respectively). The average model, in turn, was significantly better than the substitution model ($d = 4.26$). In the parafovea, model also had a significant effect, $F(2, 34) = 15.45$, $p < 0.001$, $\eta^2 = 0.48$ and $F(2, 34) = 32.54$, $p < 0.001$, $\eta^2 = 0.66$ for left and right, respectively. In the left, the average model ($MSE = 7.18$, $SD = 0.57$) was superior to both the central and substitution models ($MSE = 13.97$, $SD = 1.08$ and $MSE = 8.55$, $SD = 0.73$, $d = 1.06$ and 1.67, respectively), and the substitution model was better than the central model ($d = 0.75$). Likewise, in the right, the average model ($MSE = 7.01$, $SD = 0.37$) was superior to both the central and substitution models ($MSE = 14.11$, $SD = 0.94$ and $MSE = 8.34$, $SD = 0.45$, $d = 1.77$ and 2.25,

respectively), and the substitution model was again better than the central model ($d = 1.56$).

Second, we compared the effect of eccentricity for each model. For the central model (in blue in Figure 6B), there was a significant effect of eccentricity, $F(2, 34) = 27.91$, $p < 0.001$, $\eta^2 = 0.62$: Performance was significantly stronger in the fovea ($MSE = 7.35$, $SD = 0.62$) compared to the left and right conditions ($MSE = 13.97$, $SD = 1.08$ and $MSE = 14.11$, $SD = 0.94$, $d = 1.59$ and 1.31, respectively). Results for the left versus right were comparable ($d = 0.04$). For the average model (in green in Figure 6B), there was a significant effect of eccentricity, $F(2, 34) = 18.04$, $p < 0.001$, $\eta^2 = 0.52$: Performance was significantly weaker in the fovea ($MSE = 10.54$, $SD = 0.70$) compared to the left and right conditions ($MSE = 7.18$, $SD = 0.57$ and $MSE = 7.01$, $SD = 0.37$, $d = 1.01$ and 1.15, respectively). Results for left versus right were similar ($d = 0.09$). For the substitution model (in yellow in Figure 6B), there was a significant effect of eccentricity, $F(2, 34) = 20.41$, $p < 0.001$, $\eta^2 = 0.55$: As was in the case of the average model, performance was significantly weaker in the fovea ($MSE = 13.17$, $SD = 0.82$) compared to the left and right conditions ($MSE = 8.55$, $SD = 0.73$ and $MSE = 8.34$, $SD = 0.45$, $d = 1.08$ and 1.18, respectively), no difference was found between results of left versus right ($d = 0.09$).

The analyses showed that model performance was affected by eccentricity. Although the central model dominated in the fovea, the average model was strongest in the left and right conditions. The performance of the substitution model was generally significantly weaker than the average model although it too was stronger than the central model in the parafovea. These results could suggest that, although participants were given two tasks (central and average), they had a natural tendency to identify the central target in the fovea and to estimate the average in the parafovea regardless of the instructions.

Discussion

We presented participants with expressive faces in foveal and parafoveal vision and asked them to either identify the expression of a central face (central task) or to estimate the average expression from a set of faces (average task). We found that participants were able to complete both tasks, but their performance was modulated by eccentricity: Although they were better at identifying the emotion of the central face in the fovea, they were better at estimating the average expression of faces in the parafovea. We then compared the performance of three models (central, average, and substitution) in predicting participants' data and

showed that the central and average models were the best predictors of participants' responses in central and parafoveal vision, respectively, irrespective of the task. The performance of the substitution model was weakest overall.

First, we considered the effects of flankers on participants' ability to identify the expression of single central targets. The baseline performances showed that observers' ability to identify a central target (central task) was identical across eccentricities when it was presented alone (flanker absent), suggesting that, in isolation, the targets were equally visible at the different locations. However, with the addition of flankers (flanker present), the performance in the central task remained the same in the fovea but deteriorated in the parafoveal left and right. These results revealed that flankers did not affect participants' ability to identify the target's expression in the fovea but did worsen participants' performance in the left and right parafoveal locations (see Figure 4), which is consistent with research on crowding that demonstrates how objects in the vicinity of a target can interfere with its identification (Bouma, 1970, but also see Levi, 2008, and Pelli & Tillman, 2008, for reviews). The absence of interference by neighboring stimuli in central vision would suggest that averaging at the fovea was primarily driven by ensemble coding and was largely independent of peripheral crowding mechanisms, whereas averaging in the parafovea may involve both central and peripheral averaging mechanisms.

Our analyses also showed that participants were able to extract an average expression from a group of faces, and, interestingly, their performance was superior in parafoveal vision compared to foveal vision (Figures 5 and 6B). In fact, their performance in estimating averages was comparable to when they were asked to identify the expression of a single central target in the fovea. Here the performance in the fovea was inferior compared to the baseline (central task, flanker absent), and the performance in the left and right was improved but still comparable to the baseline, so the trends were not influenced by any flooring or ceiling effects. These results also showed that overall information could be captured quickly, not only at fixation (Ariely, 2001; Chong & Treisman, 2001, 2003; Haberman & Whitney, 2009), but also beyond it. In addition, our findings supported feature averaging in the parafovea and near periphery as was previously reported in studies that have used homogenous sets of stimuli, such as crosses (Greenwood et al., 2009) or Gaussian patches (Greenwood et al., 2010; Parkes et al., 2001). These stimuli are characterized by repetitive low-level features, which can be averaged though peripheral mechanisms that compute statistical means of corresponding points. The faces in our experiment were also repetitive, but they were more complex and contained low- and high-level

information (emotion). Estimating the average emotional expression of a set of faces may, therefore, involve more than simply computing the mean of corresponding low-level features.

Calvo, Nummenmaa, and Hyönä (2008) demonstrated higher-level gist processing mechanisms in the processing of complex natural scenes in the periphery. They showed that participants presented with emotional scenes in their peripheral visual field could not describe the specific emotional content but could, nonetheless, gather a general "average" impression of pleasantness or unpleasantness. Such complex naturalistic scenes could not simply be averaged through low-level mechanisms, which suggests the contribution of some higher-level mechanism at the semantic level in the periphery. The enhanced ability to estimate average facial expressions presented in the parafovea in our study could, therefore, also be driven by a similar ensemble-coding mechanism in combination with the peripheral mechanisms. Future research may offer further insight into how different averaging mechanisms interact across the visual field by presenting carefully designed stimuli across different eccentricities. For example, isolated facial features (by decomposing expressive faces) could tap into the role of lower-level averaging, and complex natural scenes that contain expressive faces of different sizes, colors, and identities could examine the role of higher-level ensemble coding.

When considering the data from the foveal condition alone, participants were more accurate in identifying the expression of a central target compared to estimating the average emotion of the set. This is inconsistent with Haberman and Whitney's (2009) study in which participants were better at processing the average expression of a group of faces compared to the expression of a single face. This difference may arise from the fact that, in the present experiment, we asked participants to identify the expression of a central target face, which was always displayed at the same location, whereas in Haberman and Whitney's (2009) study, participants were exposed to an array of face images, then presented with two faces and asked to identify which of the two they had previously seen. Participants in their study were required to store the emotional expression of four faces (only one of which was later presented for identification), whereas our participants only had to remember and retrieve the emotion of a single central face, which is, therefore, a far less demanding task. Another possible reason for more superior performance in judging the central target in our study is that the flanking faces in our experiment were identical, resulting in a more homogeneous background (of both low- and high-level features) that could be more easily ignored. The difference between our stimuli could have led the central face to pop out, leading to its dominance across all foveal conditions. In contrast, Haberman and Whitney (2007, 2009) used different images with varying levels of emotional expressions in each set of faces. Repeating our study with variable flanking faces would be necessary to ascertain whether our result could be influenced by the increased homogeneity of the flankers.

We tested the performance of three models on predicting participants' responses: the central model, which returned the expression of the central target; the average model, which returned the average expression of a set; and the substitution model, which returned the expression of a flanker. The central and average models enabled us to consider how well participants completed their assigned task (central vs. average) and to determine when they might be compelled to respond differently. If participants were instructed to identify the central targets' expressions, the central model should be best at predicting their results. Similarly, if they were instructed to estimate the average expressions, then the average model should generate the best predictions. So, in principle, model performance should be determined by the participants' task.

However, the present data showed that model performance was not influenced by the task assigned to the participants but rather by location of the stimuli. This mismatch between model performance and task suggests that participants were not always completing the assigned task and/or were sometimes compelled to do otherwise. In the foveal condition, the central model outperformed the average model in both task conditions: Participants' responses were closest to the central target expression regardless of whether they were asked to identify the central target expression or to estimate the average expression. In the parafoveal conditions, the average model was consistently stronger: Participants' responses were always nearer to the average expression irrespective of the task they were given. In foveal vision, the results showed that the facial expression of the central target was dominant over the averaged expression of the set. This could be explained by the fact that central targets are processed by a higher proportion of the neural substrates compared to the peripheral flankers (Adams & Horton, 2002, 2003; Horton & Hoyt, 1991), and so their signals should be stronger and not so easily suppressed or ignored. On the other hand, in the parafovea, the average expression was dominant over the central target expression, supporting Parkes et al.'s (2001) "compulsory" averaging of features in the periphery. We also considered the average and substitution models to examine which mechanisms may be involved in the processing of sets of faces and facial expressions, particularly in the parafovea. Our analyses revealed that the average model consistently generated stronger predictions for participants' responses on both tasks and at all eccentricities. This could further support averaging

over substitution as a model for crowding (Freeman et al., 2012; Greenwood et al., 2009, 2010; Parkes et al., 2001). Some might argue that the tested models are not independent, which may limit the understanding of the mechanisms operating at the fovea and peripheral locations. However, despite the average model being a composite of the central and substitution models, the results still suggest that it is superior compared to either one individually.

The present study considered participants' ability to identify facial expressions of a single target face or extract the average expression from a set of faces. We found that, consistent with crowding literature, in the parafovea, the ability to recognize the expression from a single face was compromised in the presence of flankers. We demonstrated that participants were able to process the average expression from a set of faces in central and parafoveal vision, but their performance was actually better in the parafovea. We also compared the performance of three models (central, average, and substitution), and participants were compelled to respond with the expression of the central target in foveal vision and the average expression in the parafovea regardless of the instructions. In addition, we found further evidence to support averaging as a superior model for crowding compared to unpooled substitution. Our results have raised some questions regarding the contribution of central and peripheral and low- versus high-level averaging mechanisms and the involvements of the single and crowd face-processing pathways in participants' ability to estimate averages of facial expressions at different eccentricities. Future experiments, using complex natural scenes containing multiple expressive faces of different sizes, colors, and identities, shown from various angles, could be used to explore the role of higher-level ensemble coding, and presenting such scenes in central, parafoveal, and more peripheral visual areas may be able to shed light on these outstanding issues.

Acknowledgments

Corresponding author: Michelle P. S. To.
Email: m.to@lancaster.ac.uk.
Address: Department of Psychology, Lancaster University, Lancaster, UK.

References

Adams, D. L., & Horton, J. C. (2002, October 18). Shadows cast by retinal blood vessels mapped in primary visual cortex. *Science*, *298*(5593), 572–576.

Adams, D. L., & Horton, J. C. (2003). A precise retinotopic map of primate striate cortex generated from the representation of angioscotomas. *Journal of Neuroscience*, *23*(9), 3771–3789.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Barlow, H. (1961). Possible principles underlying the transmission of sensory messages. In W. D. Keidel, U. O. Keidel, M. E. Wigand, & W. A. Rosenblith (Eds.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.

Bouma, H. (1970, April 11). Interaction effects in parafoveal letter recognition. *Nature*, *226*(5241), 177–178.

Calvo, M. G., Nummenmaa, L., & Avero, P. (2010). Recognition advantage of happy faces in extra-foveal vision: Featural and affective processing. *Visual Cognition*, *18*(9), 1274–1297.

Calvo, M. G., Nummenmaa, L., & Hyönä, J. (2008). Emotional scenes in peripheral vision: Selective orienting and gist processing, but not content identification. *Emotion*, *8*(1), 68–80.

Chong, S. C., & Treisman, A. (2001). Representation of statistical properties. *Journal of Vision*, *1*(3):54, https://doi.org/10.1167/1.3.54. [Abstract]

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

Daugman, J. G. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Transactions on Biomedical Engineering*, *36*(1), 107–114.

FantaMorph. (2009). FantaMorph 4. [Computer software]. Abrosoft Co. Retrieved from http://www.fantamorph.com

Field, D. J. (1987). Relations between the statistics of

natural images and the response properties of cortical cells. *Journal of Optical Society of America A*, *4*(12), 2379–2394.

de Fockert, J., & Wolfenstein, C. (2009). Short article: Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, *62*(9), 1716–1722.

de Fockert, J. W., & Gautrey, B. (2013). Greater visual averaging of face identity for own-gender faces. *Psychonomic Bulletin & Review*, *20*(3), 468–473.

Freeman, J., Chakravarthi, R., & Pelli, D. G. (2012). Substitution and pooling in crowding. *Attention, Perception, & Psychophysics*, *74*(2), 379–396.

Goren, D., & Wilson, H. R. (2006). Quantifying facial expression recognition across viewing conditions. *Vision Research*, *46*(8–9), 1253–1262.

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences, USA*, *106*(31), 13130–13135.

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2010). Crowding changes appearance. *Current Biology*, *20*(6), 496–501.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753.

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.

Horton, J. C., & Hoyt, W. F. (1991). The representation of the visual field in human striate cortex: A revision of the classic Holmes map. *Archives of Ophthalmology*, *109*(6), 816–824.

Hubel, D. H., & Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology*, *158*(3), 267–293.

Inouye, T. (1909). Die Sehstörungen bei Schussverletzungen der kortikalen Sehsphäre nach Beobachtungen an Verwundeten der letzten Japanischen Kriege. W. Engelmann, Leipzig. English translation by M. Glickstein and M. Fahle under the title: Visual disturbances following gunshot wounds of the cortical visual area, *Brain* (Supplement) *123*, Oxford University Press, 2000.

Krumhansl, C. L., & Thomas, E. A. (1977). Effect of level of confusability on reporting letters from briefly presented visual displays. *Perception & Psychophysics*, *21*(3), 269–279.

Leib, A. Y., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*: 13186.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654.

Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, *25*(7), 963–977.

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.

Lister, W. T., & Holmes, G. (1916). Disturbances of vision from cerebral lesions, with special reference to the cortical representation of the macula. *Proceedings of the Royal Society of Medicine*, *9*, 57–96.

Nandy, A. S., & Tjan, B. S. (2007). The nature of letter crowding as revealed by first-and second-order classification images. *Journal of Vision*, *7*(2):5, 1–26, https://doi.org/10.1167/7.2.5. [PubMed] [Article]

Neri, P., & Levi, D. M. (2006). Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vision Research*, *46*(16), 2465–2474.

Olshausen, B. A., & Field, D. J. (1996, June 13). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of Optical Society of America A*, *2*(9), 1508–1532.

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, *11*(10), 1129–1135.

Põder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, *7*(2):23, 1–12, https://doi.org/10.1167/7.2.23. [PubMed] [Article]

Popovic, Z., & Sjöstrand, J. (2001). Resolution, separation of retinal ganglion cells, and cortical magnification in humans. *Vision Research*, *41*(10–11), 1313–1319.

Popple, A. V., & Levi, D. M. (2005). The perception of spatial order at a glance. *Vision Research*, *45*(9), 1085–1090.

PST. (2003). E-Prime 1.1. Sharpsburg, PA: Psychology Software Tools.

Schwartz, G. E., & Davidson, R. J. (1997). Neuroanatomical correlates of happiness, sadness, and disgust. *The American Journal of Psychiatry*, *154*(7), 926–933.

Strasburger, H. (2005). Unfocussed spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision*, *5*(11):8, 1024–1037, https://doi.org/10.1167/5.11.8. [PubMed] [Article]

Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 329–337.

Tjan, B. S. (2009). Three essential ingredients of crowding. *Journal of Vision*, *9*(8):988, https://doi.org/10.1167/9.8.988. [Abstract]

To, M. P. S., Baddeley, R. J., Troscianko, T., & Tolhurst, D. J. (2011). A general rule for sensory cue summation: Evidence from photographic, musical, phonetic and cross-modal stimuli. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*, 1365–1372.

To, M. P. S., Gilchrist, I. D., Troscianko, T., Kho, J., & Tolhurst, D. J. (2009). Perception of differences in natural-image stimuli: Why is peripheral viewing poorer than foveal? *ACM Transactions on Applied Perception*, *6*(26), 1–9.

To, M. P. S., Gilchrist, I. D., Troscianko, T., & Tolhurst, D. J. (2011). Discrimination of natural scenes in central and peripheral vision. *Vision Research*, *51*(14), 1686–1698.

Tolhurst, D. J., & Ling, L. (1988). Magnification factors and the organization of the human striate cortex. *Human Neurobiology*, *6*(4), 247–254.

Tolhurst, D. J., & Thompson, I. D. (1981). On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *213*(1191), 183–199.

van den Berg, R., Roerdink, J. B., & Cornelissen, F. W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS Computational Biology*, *6*(1), e1000646.

Vinje, W. E., & Gallant, J. L. (2000, February 18). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, *287*(5456), 1273–1276.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, *69*, 105–129.

Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review*, *82*(3), 184–199.

Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. J. (2006, April). A 3D facial expression database for facial behavior research. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition* (pp. 211–216). Washington, DC: IEEE Computer Society.

Yu, D., Chai, A., & Chung, S. T. (2018). Orientation information in encoding facial expressions. *Vision Research*, *150*, 29–37.

Yu, H. H., & Rosa, M. G. (2014). Uniformity and diversity of response properties of neurons in the primary visual cortex: Selectivity for orientation, direction of motion, and stimulus size from center to far periphery. *Visual Neuroscience*, *31*(1), 85–98.

Yu, H. H., Verma, R., Yang, Y., Tibballs, H. A., Lui, L. L., Reser, D. H., & Rosa, M. G. (2010). Spatial and temporal frequency tuning in striate cortex: Functional uniformity and specializations related to receptive field eccentricity. *European Journal of Neuroscience*, *31*(6), 1043–1062.