Effect of Ensemble Perception on Spatial Heterogeneity in the Perceived Gender of Faces

Shreyas Pasumarthi, Ashley Feng, Winnie Hou, and Nitin Vegesna
Department of Psychology,
University of California, Berkeley
2121 Berkeley Way West, Berkeley, CA 94704

Abstract

The human visual system has many flaws, one of them being spatial heterogeneity, a visual bias that occurs when objects, such as gendered faces, are perceived differently when presented at different locations near the foveal region of the eye (Afraz et al., 2010). Although past studies have documented many benefits of ensemble perception, it has been hitherto unknown whether ensemble perception can benefit the visual system by correcting visual biases like spatial heterogeneity. Therefore, the purpose of this study is to determine the effect of ensemble perception on spatial heterogeneity, particularly in the perceived gender of faces. This study first replicates Afraz et al.'s single face task, then conducts an additional ensemble task. In both tasks, the participants were asked to identify the gender of face stimuli displayed in different spatial locations, and the results were graphed into psychometric curves. From these curves, we found the points of subjective equality (PSE), which were used to determine whether ensemble perception reduced the single face task biases. Results showed significantly less bias in the ensemble task than in the single face task, indicating that ensemble perception does indeed reduce spatial biases. This may occur due to more neural sampling, which activates more cells, and the pooling of noisy samples, which provides a more accurate perception of the gender. This study furthers our understanding of ensemble perception and its benefits to the visual system. We suggest that future studies focus on other visual biases, such as change blindness, that ensemble perception can also potentially correct.

Introduction

Ensemble perception, or ensemble coding, refers to how the visual system implicitly receives an overall impression, or "gist," of a group of items instead of perceiving the individual features of each item (Whitney, Haberman, & Sweeny, 2014). For example, instead of analyzing each and every leaf of a tree, we can instead achieve an overall impression of the characteristics of the leaves. This mechanism aids visual perception because human awareness is necessarily limited by short term memory capacity (Cohen et al, 2016), visual crowding (Balas, Nakano, & Rosenholtz, 2009), and attentional blink (Dux & Marois, 2009). Perceiving a summary representation overcomes these visual capacity limitations (Whitney, Haberman, & Sweeny, 2014) and allows us to process large amounts of information in shorter amounts of time. Numerous studies have shown that ensemble perception works in both low-level features–such as average motion (Watamaniuk, Sekuler, & Williams, 1989; Williams & Sekuler, 1984), average speed (Watamaniuk, et al., 1989; Williams & Sekuler, 1984), average orientation (Dakin, 2001), average position (Hess & Holliday, 1992), and average size of objects (Ariely, 2001)–and high-level social features, such as the average emotion and gender of a crowd of faces (Haberman & Whitney, 2007).

Ensemble perception can provide a more accurate representation of information than the perception of single items because outliers can cancel out each other, creating a mean that is more mathematically accurate (Alvarez, 2011). In addition to providing visual benefits to neurotypical individuals, ensemble perception also benefits patients with visual deficits, such as unilateral neglect (Yamanashi Leib et al. 2012) and simultanagnosia (Demeyere, Rzeskiewicz, & Humphreys, 2008). Using ensemble perception, these patients can still obtain a veridical representation of their environment, even with significant deficits in single item perception, by utilizing a more holistic visual strategy to process the larger visual field.

However, there remains a question of whether ensemble perception can provide a unique benefit to the visual system by reducing specific spatial biases that exist near the foveal region. For instance, research conducted by Afraz et al. (2010) documented substantial bias when observers viewed gendered faces at diverse spatial locations near the foveal region. Specifically, the same face was identified as more male or more female depending on the angle of eccentricity to the fovea. Again, these differences in observer perception persisted even when the faces were identical or neutral (Afraz et al., 2010). According to the study, these biases were a result of sparse neural sampling. Since the receptive fields of cells have a limited spatial extent, they cannot cover the entire visual field; furthermore, smaller stimuli activate fewer cells. Therefore, the purpose of this study is to investigate whether ensemble perception eliminates this specific bias near the foveal region, referred to as spatial heterogeneity.

Our research was divided into two parts. The first part was a replication of the aforementioned study (Afraz et al., 2010), in which participants were asked to identify individual morphed faces presented in different locations around the fovea as male or female. Replicating Afraz et al.'s study, we observed significant spatial heterogeneity, or biases, in each participant. The second part of the experiment was an ensemble task in which *multiple* faces were shown at the locations with the strongest observed gender biases. Because prior research has demonstrated that ensemble perception generally confers an overall visual benefit due to more area for neural sampling and mathematical averaging, which improves accuracy in object identification, (Robitaille & Harris, 2011; Afraz et al., 2010), we hypothesized that ensemble perception will yield a similar benefit by minimizing spatial heterogeneity in the foveal region–essentially eliminating the single face biases originally present in the first task.

Methods

The experiment was composed of two parts: the single face task and the ensemble task. For both tasks, nine FaceGen prototype faces (four male, one neutral, and four female) were used as stimuli. The nine morphed faces (see Figure 1A) were spaced in intervals of 12.5 morph units from -50 to 50mu, with -50mu being an extremely male signal and 50mu being an extremely female signal.

Four adult participants (average age 21.75, one male, and three female) with normal or corrected vision volunteered to participate in the experiment. Participants sat 57 cm from the screen with their chin stabilized using a chinrest, and the experiment was displayed on a 1920x1080 monitor. The Michelson contrast of the monitor's background luminance was 0.993. All experiment code was written and run in the Matlab R2019a Update 3 (9.6.0.1135713) 64-bit application using Psychtoolbox-3.

For the single face task (see Figure 1B), participants were asked to focus on the central red fixation cross for the duration of the experiment. The background color of the experiment was grey. In total there were eight locations equally spaced around a circle (45 degrees apart) at three degrees of eccentricity around a red central fixation point. Each of the 9 faces was displayed 24 times in each of the 8 locations, summing up to 1728 trials in total. For each trial, a randomly chosen face stimulus was flashed for 200ms in a randomized location, and participants had to press key 'q' for a male response and key 'p' for a female response. The face number (1-9), the location number (1-8), the response ('q' or 'p'), and accuracy (0, a miss, or 1, a hit) were recorded for each trial.

After the single face task was completed, the locations with spatial biases were analyzed by plotting a psychometric curve that graphs the proportion of female responses at each of the eight locations. These biased locations were unique for each participant. Then, data points for each location were fitted using a logistic function to calculate the point of subjective quality (PSE), the value of morphing signal at which the number of female and male responses are equal.

For the ensemble task (see Figure 1C), the background color, locations, and central fixation cross were set up the same way as the single face task. Instead of displaying a single face for each trial, however, a *group* of randomly chosen face stimuli were flashed for 200ms at either the male biased location or the female biased location. Furthermore, a mean morph unit of the group of stimuli was used instead of an individual morph unit. In total, there were nine mean morph values ranging from one to nine. 50% of the trials included stimuli in the male biased locations, and the other 50% included stimuli in the female biased locations. These locations were unique to each participant and depended on their responses from the single face task. In the ensemble task, each of the 9 mean morph values was displayed 96 times at each of the 2 locations, summing up to 1728 trials in total. For each trial, participants were asked to press key ‘q’ (for male) or ‘p’ (for female) to indicate the perceived mean gender of the stimuli displayed. The mean morph value, location, response, and accuracy were recorded for each trial.

After plotting the psychometric curves, the PSE values for the ensemble task was compared with the PSE values for the single face task. In each task, all of the single face PSE values from each of the participants were concatenated into a single vector. In the single face task, we created a 1x32 vector, but in the ensemble task, we used a 1x8 vector. A one-sample t-test was then conducted for each task using these vectors in order to determine whether there was a significant bias present. Also, we conducted a t-test in order to determine if the difference between the single face PSE’s and ensemble PSE’s was significant.

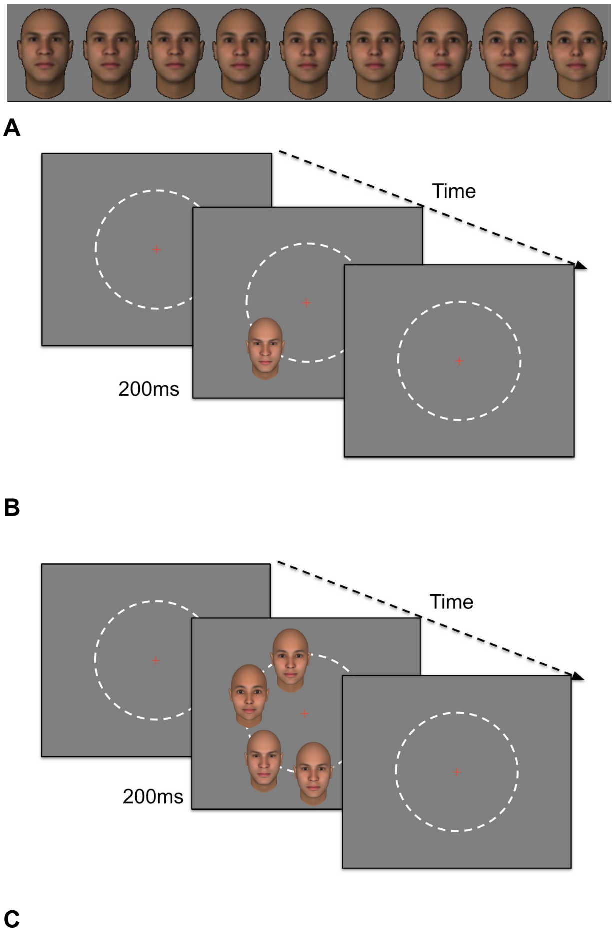


Figure 1.
(A) Stimuli set of morphed faces. Faces are displayed from the most masculine (-50 mu) to the most feminine (50 mu) from left to right.
(B) Single face task. In this task, a single stimulus appeared in one of the eight locations surrounding the center fixation cross in the area indicated by white dotted lines. Each stimulus was shown for 200ms (or 0.2 secs). Participants were then asked to determine whether the face was male or female by pressing ‘q’ or ‘p.’
(C) Ensemble task. In this task, multiple faces appeared in one of the biased locations. These faces had a mean average gender ranging from one to nine, with one being the most male and nine being the most female. The faces were shown for 200ms (or 0.2 secs). Participants were then asked to determine whether the face was male or female by pressing ‘q’ or ‘p.’

*The white dotted lines are used as a visual aid here; they were not present in the actual experiment. The stimuli here are also enlarged for visual purposes; in the actual experiment, the faces were smaller.

Results

In the single face tasks, we found that there was significant bias in each location (see Figure 3 and 4). Note that at each angle of eccentricity, participants' PSE deviates substantially from the midpoint of the x-axis.

In contrast, in the ensemble task, we found a negligible bias at each location (p=0.892). Figure 4 depicts the improvement in spatial heterogeneity for each participant. The darker hued psychometric curves depict the angular locations where the participants exhibited the most extreme biases (red = female, blue = male). The lighter hued psychometric curves depict the ensemble data (magenta = female, cyan = male). Note that the psychometric curves associated with the ensemble are significantly shifted toward the midpoint (0) of the x axis, indicating reduced biases.

To quantify the improvement in the visual bias, we compared the average PSE for the single face locations to the PSE for the ensemble conditions using paired t-tests. When examining the locations as a whole, we found a significant improvement between the average ensemble PSE and average PSE of the single face locations (t(9) = 2.33, p = 0.044). This indicates that viewing the stimuli as an ensemble not only improved the bias, but improved the biases above and beyond what is predicted by pooling the single face locations. We also looked at the ensemble male and female biased locations separately. For both male and female biased locations respectively, we found significant improvements between each ensemble PSE and the average PSE of single face locations. For the male-biased locations, t(9) = 3.145 and p = 0.012,and for the female biased locations, t(9) = -2.244 and p = 0.052. These values illustrate that ensemble perception improved bias for both female and male biased locations and further proves that ensemble perception reduces the bias associated with spatial heterogeneity.

*Please note that figures represent the four subjects run with our experimental code. However, our collaborators ran an identical study. Therefore, data analysis was obtained by pooling the data of all subjects together.

	Average Male Single Face vs Male Ensemble PSE	Average Female Single Face vs Female Ensemble PSE	Average Single Face vs Ensemble PSE's
t(9)	3.145	-2.244	2.33
p	0.012	0.052	0.044

Figure 2.

T-value and p-value table that compares the average PSE's of the single face and ensemble face tasks. The t-value gives the difference between the PSE's of the average single face task and ensemble task. The t-value of the average single face male, average single face female, and the average single face are 3.145, -2.244, and 2.33, respectively. The p-values of the PSE's of the average single face male, average single face female, and the average single face are 0.012, 0.052, and 0.044, respectively. These p-values, which are less than 0.05 for the most part, signify significant differences between the PSE's of all the single face tasks and the PSE's of the ensemble tasks.

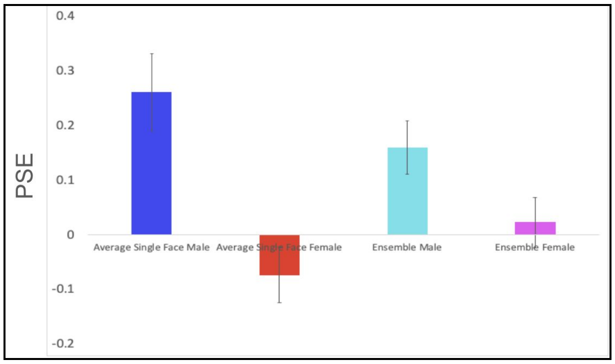
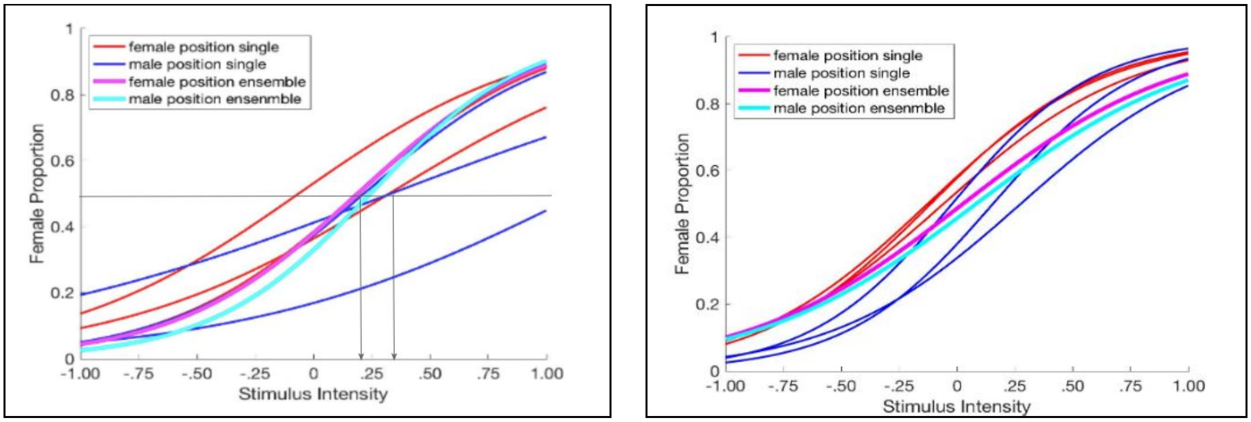


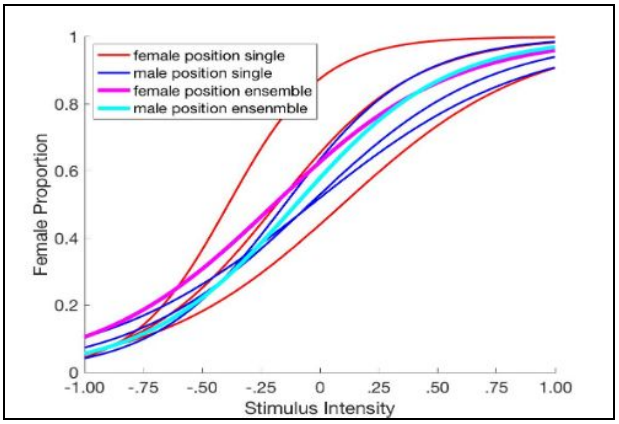
Figure 3.

PSE values for average single face male (blue) and female (red) locations and ensemble male (cyan) and female (magenta) locations. The PSE value of the average single face female shows a female leaning bias, whereas the PSE values for the other graphs show male leaning biases. PSE values for ensemble tasks are closer to zero than the PSE values of single face tasks. This displays the reduction in bias caused by ensemble perception. The black lines represent the standard deviation.

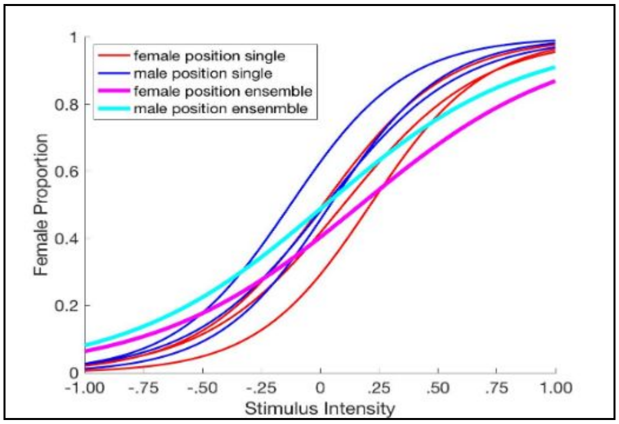


A (Subject 1)

D (Subject 4)



B (Subject 2)



C (Subject 3)

Figure 4. Psychometric curves.
Each participant's proportion of female responses for the single face task and the ensemble task were analyzed using psychometric curves. The x-axis represents the stimulus intensity, or the spectrum ranging from most male (-1.00) to most female (1.00), and the y-axis represents the proportion of female responses. Each psychometric curve displays three female single-face biased locations (red), three male single-faced biased locations (blue), the female ensemble position (magenta), and the male ensemble position (cyan). The single face curves have PSE values further from 0, indicating more bias at those locations. The ensemble curves have PSE values closer to 0, indicating a reduction of biases.

Discussion

This study reveals that ensemble perception does indeed decrease spatial heterogeneity near the foveal region because the ensemble PSE values were closer to 0 than the average single face PSE was. The reason why ensemble perception corrected this bias may be due to the greater neural sampling across the group of faces. As Afraz et al.'s study concluded, the magnitude of spatial heterogeneity depends on the size of the stimulus; as the stimulus size decreases, the heterogeneity index significantly increases (Afraz et al., 2010). This finding explains why a single face stimulus, which covers less area that a group of face stimuli, would create a more significant bias. Their study explained that cells in the visual system are organized retinotopically with a moderate, non-global size, suggesting that the receptive fields of these cells cover only a limited spatial extent and that the cells are susceptible to natural variations and biases (Afraz et al., 2010). It follows that a stimuli that covers more area would engage more groups of cells, causing the variations to average out. This “averaging out” is also a function of ensemble perception, which cancels out the uncorrelated noise of individual items in order to perceive the ensemble as a whole (Alvarez 2011, Galton 1907 & Surowiecki, 2004).

This study concludes that although the human visual system is biased in many ways, it can still utilize ensemble perception to mitigate the effects of biases and allow us to obtain an accurate visual perception of our world. This further advances our understanding of ensemble perception and its far-reaching benefits to our visual system. Future research can replicate this study but use other stimuli, such as Gabor patches with different orientations. Furthermore, future research can investigate whether ensemble perception mitigates other visual biases, such as change blindness.

References

Afraz, A., Pashkam, M. V., & Cavanagh, P. (2010). Spatial heterogeneity in the perception of face and form attributes. Current biology: CB, 20(23), 2112–2116. doi:10.1016/j.cub.2010.11.017

Alvarez GA. 2011. Representing multiple objects as an ensemble enhances visual cognition. Trends Cogn. Sci. 15(3):122–31

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

Balas B, Nakano L, Rosenholtz R. 2009. A summary-statistic representation in peripheral vision explains visual crowding. J. Vis. 9(12):13

Cohen MA, Dennett DC, Kanwisher N. 2016. What is the bandwidth of perceptual experience? Trends Cogn. Sci. 20(5):324–35

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 18, 1016-1026.

Demeyere N, Rzeskiewicz A, Humphreys KA, Humphreys GW. 2008. Automatic statistical processing of visual properties in simultanagnosia. Neuropsychologia 46(11):2861–64

Dux PE, Marois R. 2009. How humans search for targets through time: a review of data and theory from the attentional blink. Atten. Percept. Psychophys. 71(8):1683–700

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

Hess, R. F., & Holliday, I. E. (1992). The coding of spatial position by the human visual-system--effects of spatial scale and contrast. *Vision Research, 32,* 1085-1097.

Robitaille N, Harris IM. 2011. When more is less: extraction of summary statistics benefits from larger sets. J. Vis. 11(12):18

Watamaniuk, S. N. J., & Duchon, A. (1992). The human visual-system averages speed information. *Vision Research, 32,* 931-941.

Whitney, David & Haberman, Jason & Sweeny, Timothy. (2014). From textures to crowds: Multiple levels of summary statistical perception. The new visual neurosciences. 695-710.

Wilkinson, M. O., Anderson, R. S., Bradley, A., & Thibos, L. N. (2016). Neural bandwidth of veridical perception across the visual field. Journal of vision, 16(2), 1. doi:10.1167/16.2.1

Williams, D.W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research, 24,* 55-62.

Yamanashi Leib A, Landau AN, Baek Y, Chong SC, Robertson L. 2012a. Extracting the mean size across the visual field in patients with mild, chronic unilateral neglect. Front. Hum. Neurosci. 6:267