Foveal input bias in ensemble emotion perception

A thesis submitted
by
Dandan Yu

for the degree
of
Philosophy Doctor in Psychology

to the
Ecole Doctorale Science de l'Homme et de la Société
SCALab, UMR CNRS 9193
University of Lille

Composition of the Jury

Pr. Bilge Sayim	University of Lille, France	Director
Pr. Ian M. Thornton	University of Malta, Malta	President
Pr. Valérie Goffaux	Université Catholique de Louvain, Belgium	Reviewer
Dr. Sabrina Hansmann-Roth	University of Iceland, Iceland	Examiner

Public defense on April 19th, 2023

Petit à petit, l'oiseau fait son nid !

Foveal input bias in ensemble emotion perception

A thesis submitted
by
Dandan Yu

for the degree
of
Philosophy Doctor in Psychology

to the
Ecole Doctorale Science de l'Homme et de la Société
SCALab, UMR CNRS 9193
University of Lille

Composition of the Jury

Pr. Bilge Sayim	University of Lille, France	Director
Pr. Ian M. Thornton	University of Malta, Malta	President
Pr. Valérie Goffaux	Université Catholique de Louvain, Belgium	Reviewer
Dr. Sabrina Hansmann-Roth	University of Iceland, Iceland	Examiner

Public defense on April 19th, 2023

Dandan Yu: Foveal input bias in ensemble emotion perception April 2023

SUPERVISOR:
Pr. Bilge Sayim

Location:
SCALab, University of Lille

ABSTRACT

Individuals can extract the summary emotional information from groups of multiple faces, an ability called ensemble emotion perception. Previous demonstrations in ensemble emotion perception have shown that faces at (or close to) fixation weigh more than others, revealing a foveal input bias. Yet the contribution of the foveal face in ensemble emotion perception is under discussion. The general aim of my dissertation was to investigate the role of the foveal input in ensemble emotion perception. In Study 1, I investigated if – and to what extent -- the foveal input biased estimates of the ensemble. Results showed that the ensemble judgment was less accurate when the foveal face was of a different emotion from other face members, revealing a pronounced foveal input bias. In Study 2, I tested if there was a foveal input bias when the foveal face was of the same emotion but different intensity than the average. I found that the reported ensemble emotion was more intense with increasing emotional intensity of the foveal face, suggesting that the foveal input bias occurs within the same emotional category. Study 3 was designed to investigate whether the foveal input bias could be overcome by top-down attentional control. The results showed that a pronounced foveal input bias occurred when asking participants to report the average emotion of the entire face set, and there was no foveal input bias when asked to ignore the foveal face and judge the average emotion of other face members in the face set. The efficiency of the attentional control was unaffected by the variance of the foveal face. Future work will focus on investigating the foveal input bias in more natural settings. Combined, the results demonstrated how the foveal input biased ensemble perception and suggest that ensemble perception fails when salient target information is available in central vision, however, top-down attentional control can save ensemble perception. Together, the findings of this dissertation reveal how the foveal input influences ensemble perception and indicate that ensemble perception can be disrupted by the presence of salient target information in central vision. Importantly, the results also suggest that top-down attentional control can help to overcome the foveal input bias and support accurate ensemble perception.

RÉSUMÉ

Les individus sont capables d'extraire un ensemble d’informations émotionnelles de plusieurs visages, grâce à une capacité appelée "perception d'ensemble des émotions". Des démonstrations antérieures ont montré que les visages au niveau de la fixation (ou à proximité) ont plus d'importance que les autres, révélant un biais d'entrée fovéale. Toutefois, la contribution du visage fovéal dans la perception d'ensemble des émotions est discutée. L'objectif général de cette thèse était d'étudier le rôle de l'entrée fovéale dans la perception d'ensemble des émotions. Dans l'étude 1, Je cherche à savoir si, et dans quelle mesure, l'entrée fovéale biaisait les estimations de l'ensemble. Les résultats ont montré que le jugement de l'ensemble était moins précis lorsque le visage fovéal avait une émotion différente de celle des autres membres du visage, révélant un biais prononcé de l'entrée fovéale. Dans l'étude 2, j’ai testé s'il y avait un biais d'entrée fovéal lorsque le visage fovéal avait la même émotion, mais une intensité différente de la moyenne. J’ai constaté que l'émotion d'ensemble rapportée était plus intense lorsque l'intensité émotionnelle du visage fovéal augmentait, ce qui suggère que le biais d'entrée fovéal se produit au sein de la même catégorie émotionnelle. L'étude 3 visait à déterminer si le biais d'entrée fovéal pouvait être surmonté par un contrôle attentionnel descendant. Les résultats ont montré qu'un biais d'entrée fovéale prononcé se produisait lorsqu'on demandait aux participants de rapporter l'émotion moyenne de l'ensemble des visages et qu'il n'y avait pas de biais d'entrée fovéale lorsqu'on leur demandait d'ignorer le visage fovéal et de juger de l'émotion moyenne des autres membres de l'ensemble des visages. L'efficacité du contrôle attentionnel n'était pas affectée par la variance du visage fovéal. Les travaux futurs se concentreront sur l'étude du biais de l'entrée fovéale dans des environnements plus naturels. Les résultats combinés ont démontré comment l'entrée fovéale biaisait la perception d'ensemble et suggèrent que la perception d'ensemble échoue lorsque des informations saillantes sur la cible sont disponibles dans la vision centrale. Cependant, le contrôle attentionnel descendant peut sauver la perception d'ensemble. Ensemble, les résultats de cette thèse révèlent comment l'entrée

fovéale influence la perception d'ensemble et indiquent que la perception d'ensemble peut être perturbée par la présence d'informations de cibles saillantes en vision centrale. Il est important de noter que les résultats suggèrent également que le contrôle attentionnel descendant peut aider à surmonter le biais de l'entrée fovéale et à soutenir.

CURRICULUM VITAE

Dandan Yu

Laboratoire SCALab UMR CNRS 9193 Université Lille BP 60149
59653 Villeneuve d'Ascq Cedex, France
E-mail: dandan.yu@univ-lille.fr

EDUCATION

10/2019 –	PhD Student Cognitive and Affective Sciences Laboratory, University of Lille, France	Cognitive Psychology
9/2016 – 6/2019	Master Student Institute of Education, Xiamen University, Xiamen, China	Developmental and Educational Psychology
9/2012 – 6/2016	Undergraduate Student Datong University, Datong, China	Psychology

RESEARCH INTERESTS

Ensemble perception; Crowding; Facial expressions; Redundancy masking

RESEARCH EXPERIENCES

2019.10 – Cognitive and Affective Sciences Laboratory, University of Lille
- Project: Foveal input bias in ensemble emotion perception

2021.11–2022.11 Harvard Business School, Harvard University
- Project: Emotions on screen: How we perceive emotions in online meetings

2018.03 – 2019.03
State Key Lab of Brain and Cognitive Science at Institute of Psychology, Chinese Academy of Sciences
- Project: Anxious responses to unpredicted threatening stimuli

MANUSCRIPTS

- **Yu R. Dandan**, Ji, L., Song, Y., & Sayim, B. (2023). Foveal vision determines the perceived emotion of face ensembles. *Attention, Perception, & Psychophysics*, *85*(1), 209-221. https://doi.org/10.3758/s13414-022-02614-z
- Cui, J., **Yu R. Dandan**, & Jiang, G. (2023) Judging emoji by occupation: A case of emoji-based sarcasm interpretation. *Acta Psychologica*, *234*, 103870. https://doi.org/10.1016/j.actpsy.2023.103870
- Cui, J., **Yu, R. Dandan**, & Jiang, G. (in revision) What emojis are ironic markers? A survey on Chinese teenagers in Wechat app. *Journal of Psycholinguistic Research*.

- **Yu R. Dandan**, Song, Y., & Sayim, B. (in preparation) Foveal input bias in ensemble emotion perception.
- **Yu R. Dandan**, Song, Y., & Sayim, B. (in preparation) Foveal input bias can be overcome by attentional control.

ABSTRACTS

- **Yu R. Dandan**, & Sayim, B. (2023) Unbiased by redundant signals: Negativity bias for emotion perception of single but not two identical faces. *Annual meeting of the Vision Sciences Society*.
- **Yu R. Dandan**, & Sayim, B. (2022) Voluntary control can overcome the foveal input bias in ensemble perception. *22nd Conference of the European Society for Cognitive Psychology (ESCoP).*
- Li, M., **Yu R. Dandan**, & Sayim, B. (2022) Redundancy masking prevents the detection of faces. *22nd Conference of the European Society for Cognitive Psychology (ESCoP).*
- **Yu R. Dandan**, & Sayim, B. (2022) We do not have to see what we look at: Foveal input can be ignored in ensemble emotion perception. *Annual meeting of the Vision Sciences Society*.
- L-Miao, L., **Yu R. Dandan**, Reynvoet, B., & Sayim, B. (2022) Redundancy masking of faces: When trios look like duos. *Annual meeting of the Vision Sciences Society*.
- **Yu, D.**, & Sayim, B. (2021) Foveal input can be ignored in ensemble emotion perception. *GDR Vision Forum*.
- **Yu, D.**, Song, Y., & Sayim, B. (2021) Voluntary control eliminates the fovea bias in ensemble emotion perception. *European Conference on Visual Perception*.
- **Yu, D.**, Song, Y., & Sayim, B. (2021) Happy centre, happy whole: Foveal vision determines the perceived emotion of face ensembles. *Virtual Vision Sciences Society*.

TEACHING (130H)

- Cognitive Psychology
- History and Systems of Western Psychology
- A practical introduction to Eprime 3.0

GRANTS

- Mobility grant for study and research at Harvard University (Bourses de mobilité de l'Université de Lille) – November 2021 – November 2022.

Table of content

CHAPTER I

General Introduction

1. General Introduction

Imagine you are in your first Psychology class of this semester, most of the time you can quickly have a general idea about how many students are in the classroom, the percentage of male and female students, whether they are excited or nervous, and so on. The ability to extract the summary statistical information from groups of similar objects is called ensemble perception (for reviews see, Alvarez, 2011; Whitney & Yamanashi Leib, 2018).

1.1 Ensemble perception: from low-level to high-level features

1.1.1 Does ensemble perception of low- and high-level features involve distinct mechanisms?

Ensemble perception occurs not only when perceiving low-level features such as size (Chong & Treisman, 2003, 2005; Haberman & Suresh, 2021), orientation (Dakin & Watt, 1997; Parkes et al., 2001), and motion (Watamaniuk et al., 1989; Sweeney et al., 2012), but also when perceiving high-level features such as facial identity (Neumann et al., 2013; Jung et al., 2017), attractiveness (Luo & Zhou, 2018), and facial expressions (Haberman & Whitney, 2007, 2009; Fischer & Whitney, 2011). For instance, Ariely (2001) presented a set of circles (set size: 4, 8, 12, 16) with different sizes. Participants were required to judge whether a subsequently presented test circle was larger or smaller than the mean of the previous circle set. The results showed that participants were able to accurately discriminate the mean size of the circle set, and that the discriminability was independent of the set size, revealing that ensemble perception occurred with unlimited capacity. In the feature domain of facial expressions, researchers showed that when the emotional intensity of a test face was close to the average emotion of the previously presented face set. In particular, participants were more likely to regard the test face as a set member, suggesting that individuals can extract the summary emotional information of groups of faces (e.g., Haberman et al., 2007, 2009).

Even though numerous studies have demonstrated robust ensemble perception in different feature domains, it remains poorly understood whether the processing of low- and high-level features is based on a common mechanism. The relationship between ensemble perception of low- and high-level features is often investigated by comparing the performance correlation in both tasks (Haberman et al., 2015; Yörük & Boduroglu, 2020; Kacin et al., 2021; Kwon & Chong, 2023). For instance, Haberman et al. (2015) demonstrated an independent ensemble processing mechanism across low- and high-level features. They compared the performance correlation when asked to perceive two different low-level features (e.g., orientation vs. color), two different high-level features (e.g., face identity vs. facial expressions), and one low-level and one high-level feature (e.g., orientation vs. face identity), and found positive correlations between two low-level features and two high-level features, but no correlation between one low-level and one high-level features. Consistent with the findings of Haberman et al. (2015), Kacin, Gauthier, and Cha (2021) compared the mean absolute error when perceiving the ensemble information of two low-level features -- length and orientation, and found a positive performance correlation between length and orientation tasks, suggesting that at least to some extent, there is a common mechanism for ensemble processing of low-level features. In some circumstances, however, researchers showed that the processing of high-level features shares a number of characteristics with the processing of low-level features. For instance, ensemble perception of size, motion direction, and facial expressions showed a pronounced recency effect -- more recent items presented in the stimulus sequence weigh more than others in ensemble perception (Hubert-Wallander & Boynton, 2015; Goldenberg et al., 2022). The set size and stimuli duration had a slight effect on the ensemble perception of both feature domains (Ariely, 2001; Chong & Treisman, 2003; Chong et al., 2008; Haberman & Whitney, 2009; Li et al., 2016). Moreover, some studies have shown that participants could extract multiple ensemble characteristics such as the average speed and size simultaneously (Emmanouil & Treisman, 2008; Albrecht et al., 2012). Meanwhile, they could extract not only the average information but also the variance information and the distribution information of the stimuli set (Solomon 2010; Haberman et al., 2015; Chetverikov et al., 2016, 2017; Hansmann-Roth et al., 2018), suggesting that the ensemble perception can occur in a hierarchical manner, which means that it can happen at different levels of processing in the brain.

1.1.2 The relationship between ensemble perception and individual item perception

One of the significant characteristics of ensemble perception is that it summarizes individual information into an ensemble. Previous studies have demonstrated that ensemble perception of a set of items is as quick and accurate as the processing of one single item (Chong & Treisman, 2003; Haberman & Whitney, 2009; Haberman et al., 2015; Li et al., 2016). For instance, Chong & Treisman (2003) presented two arrays of circles in the left and right visual fields simultaneously. The two arrays of circles were either homogeneous (12 circles of the same size in each array), heterogeneous (12 circles of 4 different sizes in each array), or single (there was only 1 circle on each side). Participants were required to report which circle (array) had the larger size or larger mean size. Results showed that the discrimination was comparable in the three conditions, suggesting that the extraction of mean information was as accurate as the extraction of a single item's information. According to the holistic model (Furtak et al., 2022), the brain extracts the global information before the extraction of individual information, and the global information can influence the subsequent perception of individual information (Navon, 1977; Hochstein & Ahissar, 2002; Campana & Tallon-Baudry, 2013; Furtak et al., 2022). However, the specific individual information can bias ensemble perception. For instance, according to the amplification effect found in ensemble perception, the more salient items were overrepresented in ensemble coding (for low-level features, see Kanaya et al., 2018; Iakovlev & Utochkin, 2021; Choi & Chong, 2020; for high-level features, see Goldenberg et al., 2021, 2022). For instance, Goldenberg and colleagues investigated the amplification effect of simultaneously (Goldenberg et al., 2021) and sequentially (Goldenberg et al., 2022) presented facial expressions. In their study, 1-12 happy or angry faces with different levels of emotional intensity were presented (happy and angry expressions were not mixed in any trial), followed by a single probe face with a neutral emotion. Participants were required to adjust the emotional intensity of the probe face to the average emotionality of the face set. Results showed that participants tended to overestimate the emotionality of the face set, and this amplification effect was more pronounced with larger set sizes (i.e., more faces presented in the face set). Evidence from eye-tracking showed that participants spent more time looking at more emotional faces in the face set, which could cause the amplification effect (Goldenberg et al., 2021).

In redundancy masking (RM), observers tend to underestimate the number of identical items presented in the visual periphery (Sayim & Taylor, 2019; Yildirim et al., 2020, 2021, 2022). For instance, when presented three identical lines in the periphery, participants frequently reported two lines presented instead of three (Yildirim et al., 2020, 2021). Both ensemble perception and RM are a kind of information compression – Ensemble perception occurs by reducing the information from individual objects to the feature dimension that is extracted (and maybe some additional information; for reviews see, Alvarez, 2011; Whitney & Yamanashi Leib, 2018), RM occurs by having a statistical representation of the number of objects (the output is systematically less than the number of presented objects; Yildirim et al., 2020, 2021). Ensemble perception occurs independent of whether all items are detected or not, but usually, it is assumed that all objects are at least in principle detectable, and segmented into individual objects (Ariely, 2001; Chong & Treisman, 2005; Haberman & Whitney, 2009). In RM not all items are detected, and segmentation or individuation in all presented objects fails. In our recent work (L-Miao et al., 2023, in preparation), we investigated the redundancy masking effect with high-level features -- faces. In the study, a face set including 3-6 upright or upside-down neutral faces was presented in the visual periphery (Eccentricity: 10°). Participants were required to judge the number of faces presented in the face set and the orientation of faces (i.e., report whether the face set was upright or upside-down) on each trial. The order of tasks was counterbalanced between subjects. The results showed that participants tended to underestimate the number of faces in the face set, revealing a redundancy masking effect. However, the orientation discriminability in trials with and without redundancy masking was comparable, suggesting that the missing face members had no effect on the orientation judgment of the entire face set. However, when breaking the uniformity of items (e.g., \\\ vs. \\/), the redundancy masking effect disappeared (Rummens & Sayim, 2022) in less regular triplets condition (i.e., \\/), the corresponded ensemble performance is still unclear.

1.1.3 The variance of stimulus distributions modulates ensemble performance

Past studies have shown that individuals can extract not only the average information but also the variance of the stimulus set (Solomon, 2010; Haberman et al., 2015). The two characteristics also interact to help individuals get an overall impression of the scene (Corbett et al., 2012; Im & Halberda, 2013). Then, how do mean and variance information interact during ensemble coding? Chong & Treisman (2003) showed that the variance had a slight effect on mean size perception. In their study, the variance of the circle set was manipulated by varying the distribution of the circle set -- normal, uniform, two-peaks, and homogeneous. Haberman and Whitney (2009) investigated discriminability in discriminating the average emotion of homogeneous and heterogeneous face sets. In the homogeneous condition, the facial expression of each face member was the same, whereas, in the heterogeneous condition, they varied from each other (i.e., there were four unique facial expressions in each face set). The results showed that the discriminability was comparable in the two conditions (the same results were found in a study about ensemble facial attractiveness perception, see Luo & Zhou, 2018). Meanwhile, some studies demonstrated that the variance modulates the effect of set size in ensemble perception. When the variance of the face set was relatively large, increasing set sizes led to a poorer averaging performance. On the contrary, when the variance was relatively small, averaging performance was unaffected by the set size (Ji & Pourtois, 2018; Im et al., 2017; Marchant et al., 2013). Future studies should continue to explore the interactions between the variance and mean in summary statistical processing.

1.2 The internal representation of ensemble information

1.2.1 Different weighting mechanisms involved in ensemble perception

Although individuals can quickly and relatively accurately extract the summary statistical information of groups of similar objects, the underlying mechanism is still under debate. There are several weighting mechanisms demonstrated by previous studies: (1) General averaging account: according to this model, ensemble perception occurs automatically (Chong & Treisman, 2005). Evidence showing that the ensemble performance was unaffected by set size (Ariely, 2001; Chong & Treisman, 2005; Haberman & Whitney, 2007, 2009), presentation mode (simultaneous vs. successive; Goldenberg et al., 2021, 2022), stimulus duration (from 50 ms to 2000 ms; Chong & Treisman, 2003; Li et al., 2016), and set distribution (homogeneous vs. heterogeneous; Chong & Treisman, 2003; Haberman & Whitney, 2009; Luo & Zhou, 2018) supports this statement; (2) Weighted averaging account: the account supports that items in the stimulus set do not weigh equally in ensemble coding. Some items are given more weight than others in ensemble judgments. For instance, researchers showed that attended items (De Fockert & Marchant, 2008; Im et al., 2015), salient items (Kanaya et al., 2018; Cant & Xu, 2020; but see Epstein et al., 2020; Rosenbaum et al., 2021), earlier or recent items in the sequence (Hubert-Wallander & Boynton, 2015; Tong et al., 2019; Goldenberg et al., 2022), items close to the mean of the distribution (Vandormael et al., 2017; Ni & Stocker, 2022; Iakovlev & Utochkin, 2023), and items at fixation (Ji et al., 2014; Jung et al., 2017; Dandan et al., 2023a) weigh more in ensemble coding.

1.2.2 Ensemble perception vs. crowding

When comparing peripheral with foveal vision, one of the important factors that distinguish the two is the extent of crowding. Crowding is the phenomenon that stimuli that are easily identified in isolation are not discernible when surrounded by similar objects (e.g., Bouma, 1970; Pelli & Tillman, 2008; Whitney & Levi, 2011; Herzog et al., 2015; Sayim et al., 2010, 2013, 2017). The distance over which flankers (i.e., the neighboring stimuli that are presented alongside a central target in a crowded visual scene) interfere with target perception (i.e., the ‘crowding zone’) increases with the eccentricity of the target (e.g., Bouma, 1970, 1973; Toet & Levi, 1992; Pelli et al., 2004; Tripathy et al., 2014). Crowding renders the target difficult or impossible to discern, however, the target signal is often not entirely lost. For example, it may become part of an average representation (Parkes et al., 2001), function as a semantic or emotional valence prime (Yeh et al., 2012; Kouider et al., 2011) and be retrieved by long-range grouping mechanisms (Sayim et al., 2014).

The relationship between ensemble perception and crowding is still under debate. Some studies demonstrated that crowding might facilitate ensemble representation (Parkes et al., 2001; Wolfe & Robertson, 2011; Fischer & Whitney, 2011). Despite poor identification, the crowded central target can still be integrated into the ensemble representation (see Parkes et al., 2001 for the orientation discrimination task; see Fischer & Whitney, 2011 for the emotion discrimination task). For example, Fischer & Whitney (2011) presented two groups of seven faces in the left and right visual fields (Eccentricity: 16.5°) and asked participants to judge either which central face (crowding task) or which face group (ensemble task) was more disgusted in separate blocks. There was also a control condition where one single central face was presented. Results showed that the discrimination of the central face in a face group was significantly impaired compared to the control condition where only the central face was presented, revealing a classical crowding effect. In the ensemble task condition, participants were more likely to perceive the face group in the right visual field as more disgusted when the corresponded central face was more disgusted and vice versa. Meanwhile, some studies showed that ensemble perception and crowding involve distinct mechanisms. First, it seems that crowding and ensemble perception are modulated by eccentricity in different ways. In crowding, numerous studies have demonstrated that the crowding effect is stronger in the periphery than the fovea (for reviews see Herzog et al., 2015; Levi, 2008; Whitney & Levi, 2011). However, To et al. (2019) first investigated how eccentricity modulates the ensemble performance and found a parafovea averaging advantage -- the ensemble performance was better in the parafovea than the fovea. In the experiment, they presented a set of 9 faces either in the center of the screen or at 3° eccentricity (parafovea) in separate blocks. Participants were required to judge either the emotion of the central face or the average emotion of the face set. Unsurprisingly, identification of the central face was better in the fovea than in the parafovea, revealing the classical crowding effect. However, participants' ensemble judgments were more accurate in the parafovea than in the fovea, showing a parafovea averaging advantage. Importantly, participants' responses in the fovea condition were biased by the central face, indicating that observers were not able to equally weight foveal and parafoveal faces in ensemble coding (see Section 1.3.1 for the overrepresentation of foveal input in ensemble perception -- the foveal input bias). Second, Bulakowski and colleagues (2011) demonstrated that ensemble perception diverges from crowding in regard to visual field asymmetries. In their study, arrays of bars with different orientations were presented in two different visual fields (upper vs. lower visual field), and participants were required to judge either the orientation of a crowded bar (crowding task; Experiment 1) or the average orientation of all bars (ensemble task; Experiment 2). The result showed that the crowding effect was stronger in the upper than the lower visual field (i.e., vertical meridian asymmetry), whereas the ensemble performance was comparable in both visual fields. Third, Lin et al. (2022) showed that attention strongly modulated the crowding effect, but had only a slight effect on the ensemble performance. In the divert attention condition, a post-cue appeared after the stimuli (an array of bars) and indicated the task participants needed to perform (i.e., either crowding or ensemble task). In the direct attention condition, however, a pre-cue was presented before the stimuli and indicated the location of the target (i.e., either the location of the central target in the crowding task or the location of all items in the ensemble task). The results showed that the performance in crowding and ensemble tasks was worse with fewer attentional resources (divert attention), and the crowding performance was more severely harmed than the ensemble performance. However, when given more attentional resources (direct attention), the pre-cue improved only the performance in the crowding task but not in the ensemble task. Taken together, the relationship between crowding and ensemble perception is complex and context-dependent. While crowding can to some extent facilitate the ensemble representation, the processes involved in crowding and ensemble perception are generally considered to be dissociable.

1.3 The role of attention in ensemble perception

1.3.1 Does ensemble perception require attention?

The visual system can compute statistical information with limited attentional resources (Alvarez & Oliva, 2008; Sekimoto & Motoyoshi, 2022). Previous studies showed that the summary statistical information could be computed from crowded objects in the visual periphery (Parkes et al., 2001; Fischer & Whitney, 2011), with limited presentation time and large set size (Ariely, 2001; Chong & Treisman, 2003; Li et al., 2016). These results demonstrated that ensemble perception could occur with little requirement of focal attention. Ensemble perception and focal attention also have been suggested to be two distinct mechanisms dealing with the limited capacity of the visual system (Baek & Chong, 2020) -- ensemble perception provides the gist about a scene, whereas focal attention selects important information from the scene to recognize a few objects.

Even though ensemble perception occurs with limited attentional resources, attended items are given more weight than unattended items during ensemble coding (De Fockert & Marchant, 2008; Im et al., 2015). Participants tended to overestimate the average size of circle sets when required to fixate on the largest circle member and vice versa (De Fockert & Marchant, 2008). The demonstrated foveal input bias in ensemble perception (see below) could also be explained by the attentional distribution to the foveal input. Specifically, previous studies showed that the face(s) located at the fovea vision are given more weight than faces in the periphery in the perceived ensemble (Ji et al., 2014; Jung et al., 2017), revealing a foveal input bias. For instance, Ji and colleagues (2014) investigated the relative contribution of foveal and extrafoveal faces in the perceived ensemble. In their study, participants were presented with a group of faces, consisting of 4 foveal faces located in the central vision and 12 extrafoveal faces presented in the extrafoveal vision. The foveal faces and extrafoveal faces were manipulated to be either congruent or incongruent in terms of their emotional expression. In the congruent condition, the foveal faces had the same emotion as the extrafoveal faces, while in the incongruent condition, the foveal faces had different emotion than the extrafoveal faces. The results showed that the perceived ensemble was less accurate in the incongruent than in the congruent condition. Participants frequently reported the foveal faces’ emotion as the average emotion of the ensemble when the foveal faces conveyed a different emotion than the extrafoveal faces. Taken together, the results suggest that focused attention can facilitate ensemble perception.

1.3.2 Top-down attentional control used in ensemble tasks

The aforementioned studies of how outliers biased ensemble perception (Kanaya et al., 2018; Cant & Xu, 2020) examined the possibility that ensemble perception occurs in an involuntary, stimulus-driven manner. However, it remains unclear how top-down attentional control biases ensemble performance. In visual search tasks, for instance, when participants were asked to voluntarily ignore one salient item, it can be ignored only when the feature of both the to-be-ignored item and target was constant during the experiment (Theeuwes & Burger, 1998; Wang & Theeuwes, 2018). For example, in Theeuwes and Burger's study (1998), the to-be-ignored salient singleton (letter E or R) was set in a different color (i.e., the singleton was red while other items including the target were green or vice versa). The color of the singleton and other items was interleaved between trials, meaning that the singleton's color was either red or green between trials, and the other items always had a different color from that of the singleton on each trial. The singleton was either congruent or incongruent with the target. In the congruent condition, the singleton and target were the same (i.e., both letters were Es or Rs). In the incongruent condition, they were different (i.e., the singleton was E and the target was R or vice versa). Participants were required to ignore the singleton based on the color information and judge whether an E or R was present among the non-singleton letters. The results showed that participants spent more time identifying the target in the incongruent than the congruent condition, suggesting that the to-be-ignored singletons could not be successfully ignored but biased the search performance. However, when the color of the singleton and the target was kept constant (i.e., the singleton was constantly red while the other items were constantly green in the experiment or vice versa), the RTs were comparable in congruent and incongruent conditions, demonstrating that when the feature of the to-be-ignored singleton and target was constant, individuals could voluntarily ignore the highly salient singleton. In ensemble tasks, researchers showed that the to-be-ignored item(s) contributed to the averaging process (Oriet & Brand, 2013; Chen et al., 2021), suggesting that top-down attentional control may fail in ensemble perception. For example, Alvarez & Oliva (2008) used a divided attention task in which participants tracked the movement of targets (clouds of dots) while ignoring distractors, and then localize either the location of one missing item or the centroid of four missing items. The missing items were either targets or distractors. The results showed that while participants' discrimination of the single missing distractor was close to the chance level, they were able to accurately discriminate the centroid of four missing distractors, indicating that the to-be-ignored distractors were still processed (see also Oriet & Brand, 2013; Chen et al., 2021).

1.4 The framework of the current dissertation

The current dissertation consisted of three studies. Study 1 aimed to investigate whether and to what extent the foveal face determines the perceived emotion of face ensembles (i.e., the foveal input bias). The stimuli consisted of one foveal face and eight surrounding faces ('flankers'). The foveal input bias was tested by manipulating the congruency of the foveal face and flankers. In the congruent condition, the foveal face had the same emotion as flankers. In the incongruent condition, however, the foveal face had different emotion from that of flankers. The significantly impaired discriminability in the incongruent condition would demonstrate a foveal input bias. To preview the result, I found a pronounced foveal input bias -- the ensemble performance was significantly impaired in incongruent compared to congruent conditions, and participants tended to regard the foveal face's emotion as the average emotion of the face set when the foveal face had different emotion than the other face members.

Study 2 investigated whether the foveal input bias is pronounced when the foveal face is of the same emotional category as the flankers. I used 11 intensity levels of foveal face and flankers, and asked participants to report the average emotion of faces sets on a scale with 11 levels (from 0 to 10; with '0' representing 'disgusted', '5' representing 'neutral', and '10' representing 'happy'). I found a pronounced foveal input bias -- participants perceived the average emotion as more intense with the increased emotional intensity of the foveal face. The results suggest that the foveal input bias is pronounced within the same emotional category.

Study 3 tested whether the pronounced foveal input bias found in previous studies can be overcome by top-down attentional control. In the study, I manipulated the attentional control by either asking participants to report the entire face set’s average emotion or asking them to voluntarily ignore the foveal face and report only the average emotion of flankers. I found a pronounced foveal input bias when asked to judge the average emotion of the entire face set, which was consistent with the previous demonstrations. However, the foveal input bias disappeared when asked to voluntarily ignore the foveal face, the performance was comparable in congruent and incongruent conditions, suggesting that the foveal input bias is not ubiquitous, but can be overcome by voluntary control.

Taken together, in the current dissertation, I systematically investigated the foveal input bias in ensemble perception. I first clarified the role of the foveal input by demonstrating that the foveal input can strongly bias ensemble performance. Second, I showed that the foveal input bias is strong regardless of the (un)grouping of the foveal face and flankers, the variance of the foveal face, the modality of face groups, and response formats. Third, I demonstrated that although the foveal input bias is pronounced in many circumstances, it can be overcome by voluntary control. My work contributed to studies that aim to investigate the weighting mechanism of ensemble perception.

CHAPTER II

Study 1: Foveal vision determines the perceived emotion of face ensembles

ABSTRACT: Study 1

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

Published as **Yu R. Dandan**, Ji, L., Song, Y., & Sayim, B. (2023). Foveal vision determines the perceived emotion of face ensembles. *Attention, Perception, & Psychophysics*, *85*(1), 209-221. https://doi.org/10.3758/s13414-022-02614-z

2. Study 1: Foveal vision determines the perceived emotion of face ensembles

2.1 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.

2.2 Experiment 1: Uniform flankers

2.2.1 Method

2.2.1.1 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 ($1-\beta$). 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.

2.2.1.2 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° × 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.

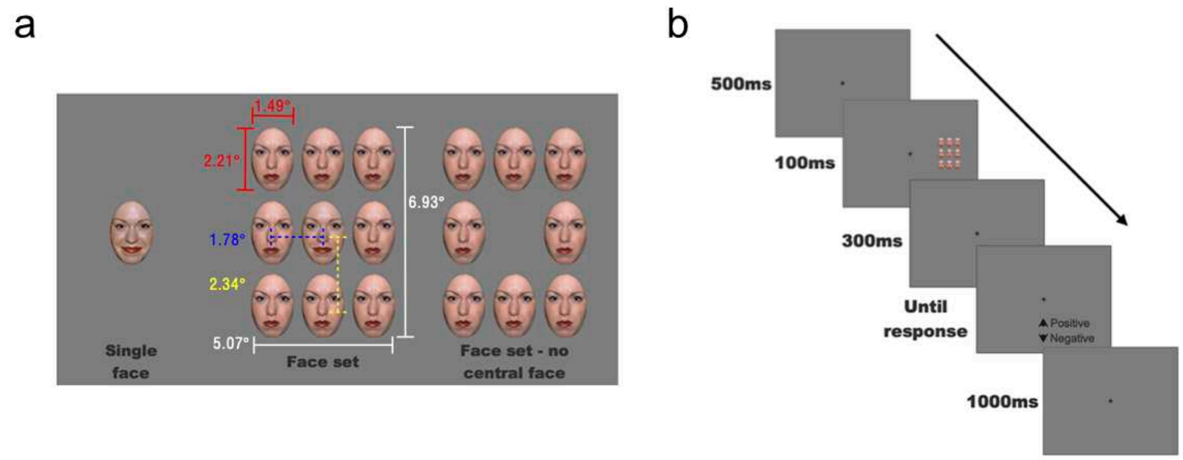
2.2.1.3 Design and Procedure

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'.

2.2.2 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:

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

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

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.

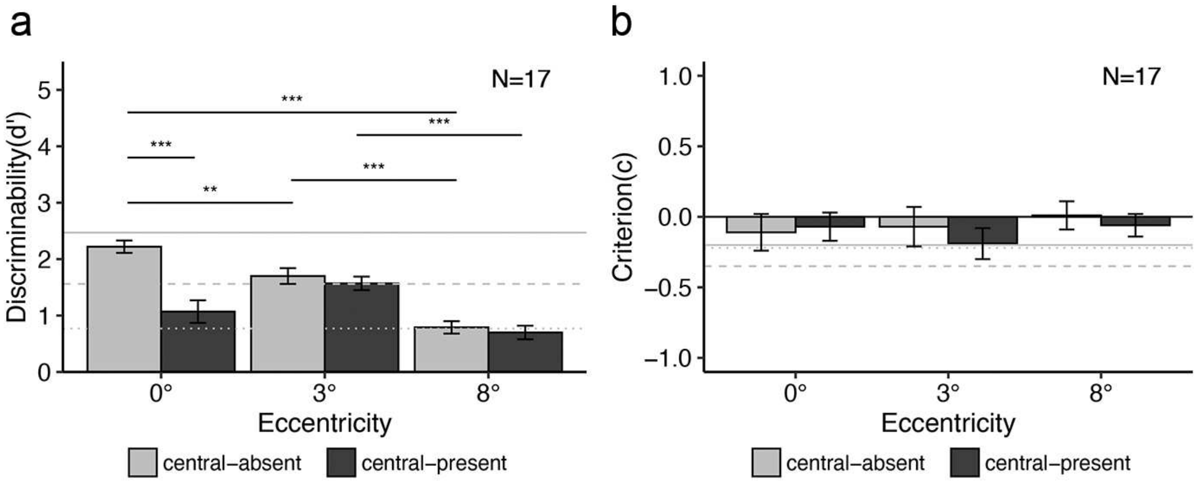
2.2.3 Results

2.2.3.1 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*.

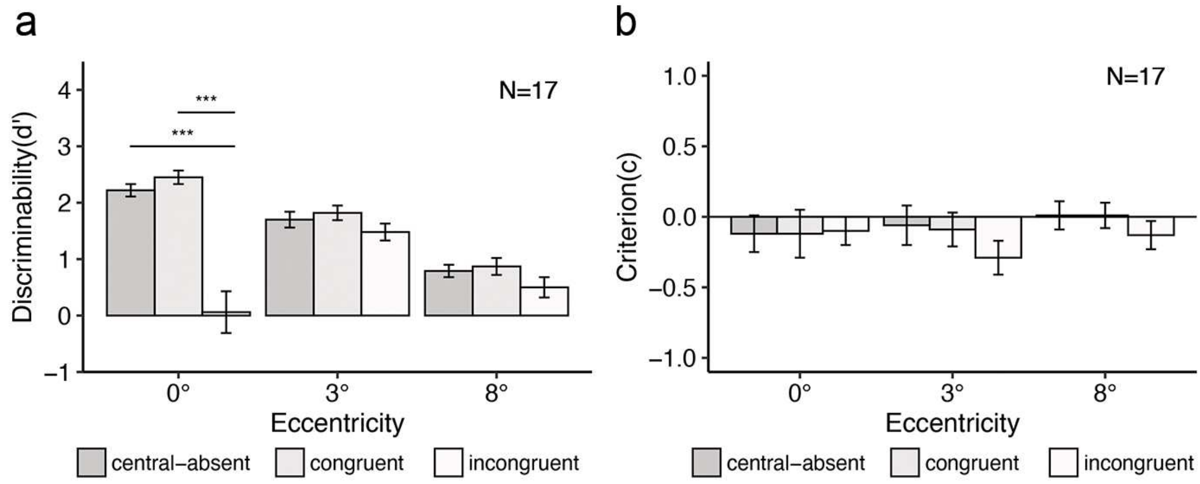
2.2.3.2 Congruency

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



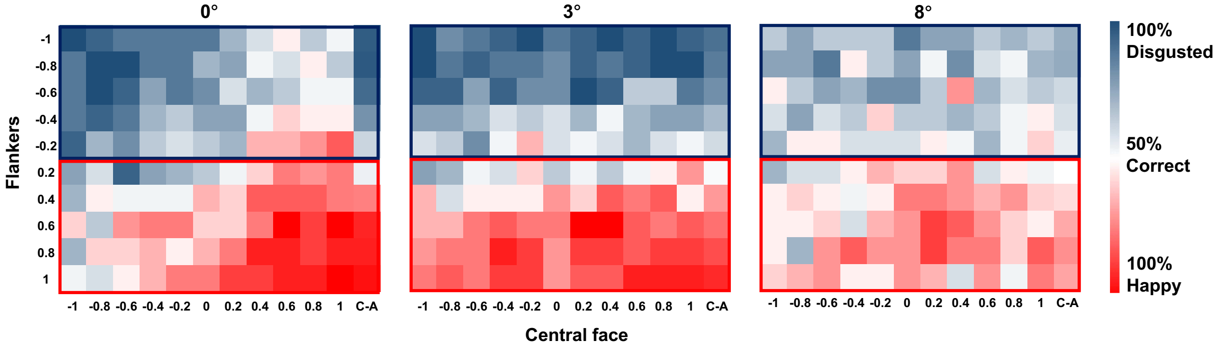
Congruency results of Experiment 1. Discriminability **(a)** and criterion **(b)** separated for face sets with congruent and incongruent central faces and flankers, and without a central face (central-absent condition). Asterisks indicate significance with alpha levels of 0.001 (***). Significance is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

2.2.3.3 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.

2.3 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.

2.3.1 Method

2.3.1.1 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.

2.3.1.2 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.

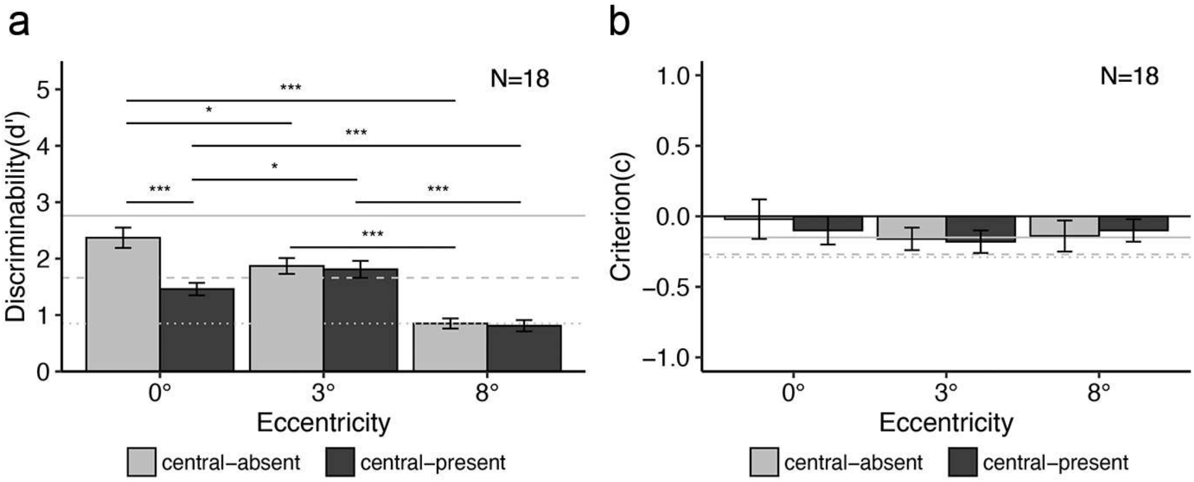
2.3.2 Results

2.3.2.1 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



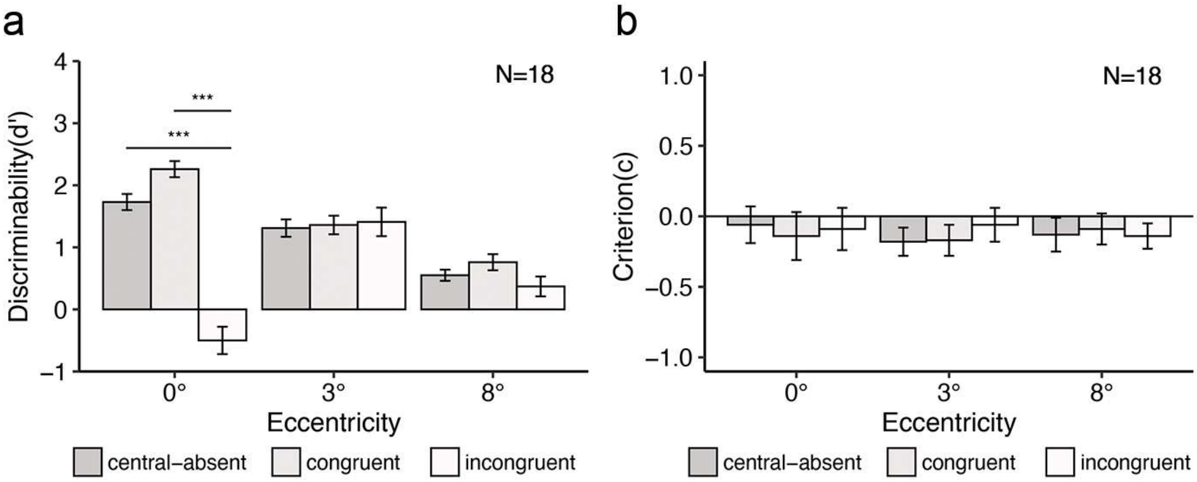
Results of Experiment 2. Discriminability **(a)** and criterion **(b)** separated for central-absent and central-present conditions. The gray horizontal lines represent discriminability and criterion in the single face condition at 0° (solid line), 3° (dashed line) and 8° (dotted line). Asterisks indicate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

2.3.2.2 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*.

2.4 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.

2.5 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.

CHAPTER III

Study 2: Foveal input bias in ensemble emotion perception

ABSTRACT: Study 2

When perceiving the emotionality of a group of faces, fixated faces weigh more than faces that are not fixated. This effect -- the foveal input bias -- has been shown for categorical responses where observers judged the emotions of face groups by indicating whether the average emotion was positive or negative. For example, on average positive groups were judged as negative when the fixated face was negative. Here, to test if there was a foveal input bias when a fixated face was of the same emotion but different intensity than the average, we asked participants to report the average emotion of face sets on a scale with 11 levels (from very disgusted to very happy). Participants were presented with face sets arranged in a 3 × 3 faces matrix, and were asked to judge the average emotion. There were three conditions: the congruent condition where the central face had the same emotion as the surrounding faces ("flankers"); the incongruent condition where the central face had a different emotion than the flankers, and the central-absent condition where no central face was presented. The ensemble performance was impaired in the incongruent compared to the congruent and central-absent conditions, showing a pronounced foveal input bias. When the foveal face and the flankers had the same emotion but different emotional intensity, participants' average emotion judgments were biased by the foveal face, showing a foveal input bias within the same emotional category. There was a significant central-tendency response bias -- participants were more likely to report the midpoint of the rating scale. Our results show that the foveal input bias is pronounced within the same emotional category.

Keywords: ensemble emotion perception, foveal input bias, central-tendency response bias

3. Study 2: Foveal input bias in ensemble emotion perception

3.1 Introduction

The processing of facial expressions has an important role in our social life. Facial expressions are a universal system of signals which reflect the moment-to-moment fluctuations in a person's emotional state (Ekman & Friesen, 1971; Russell et al., 1994; Elfenbein & Ambady, 2002). In our daily life, we often see groups of faces (Haberman et al., 2015; Goldenberg et al., 2021; Mihalache et al., 2021). The visual system's ability to extract the summary emotional information from groups of faces is called ensemble emotion perception (Haberman & Whitney, 2007, 2009, Whitney & Yamanashi Leib, 2018). For example, previous demonstrations showed that the closer the test face's emotion was to the average emotion of the previously presented face set, the more participants tended to regard it as a member of the group, showing that ensemble emotional information was extracted (Haberman & Whitney, 2007, 2009; Li et al., 2016; Ji & Hayward, 2021). Ensemble emotion perception is remarkably efficient. For example, observers could extract the average emotion of a group of faces successfully with short stimulus durations (e.g., as short as 50 ms; Li et al., 2016), large set sizes (e.g., up to 24; Wolfe et al., 2015), and dynamic (Elias et al., 2017), ambiguous (e.g., Mooney face, Han et al., 2021), and sequentially presented facial expressions (Haberman et al., 2009; Hubert-Wallander & Boynton, 2015; Ji et al., 2018; Goldenberg et al., 2022).

Even though individuals can quickly and quite efficiently extract the ensemble emotional information from a group of faces, previous studies demonstrated that faces at (or close to) fixation weigh more than other face members, revealing a foveal input bias (e.g., Ji et al., 2014; Ueda, 2022; Dandan et al., 2023). For instance, we presented a single face (visual angle: 1.49° × 2.21°) in the fovea with 8 faces surrounded ('flankers') (Dandan et al., 2023). To directly test the foveal input bias, we compared the discriminability of the average emotion of the face set with and without a foveal face. When the foveal face was present, its emotion was either congruent or incongruent with that of the flankers: In the congruent condition, the foveal face had the same emotion as the flankers (i.e., both the foveal face and flankers were happy or disgusted). In the incongruent condition, the foveal face had a different emotion from the flankers (i.e., the foveal face was happy, whereas the flankers were disgusted or vice versa). Participants judged whether the average emotion of the face array was positive or negative. The ensemble performance was better when no foveal face was presented. The deteriorated performance with a foveal face was driven by the incongruent condition where the foveal face strongly biased responses. The foveal input bias is pronounced not only when extracting ensemble emotional information, but also when extracting ensemble facial race information (Jung et al., 2017), suggesting that the foveal input bias might be ubiquitous in ensemble face perception.

Yet, it is unclear how the foveal input bias manifests itself when the emotional category of the foveal face is the same as the other face members in the face set (e.g., all faces happy to varying degrees) but different emotional intensity. Several studies investigated ensemble performance when all face members in the face set belonged to the same category (Li et al., 2016; Goldenberg et al., 2021, 2022; Ueda, 2022). For instance, Goldenberg et al. (2021, 2022) demonstrated a crowd emotion amplification effect -- the perceived average emotionality was more intense than it actually was when perceiving groups of faces within the same emotional category (i.e., all face members within the face set were either happy or angry (to different degrees); happy and angry faces were not intermixed within a face set). The amplification effect was driven by attention to more intense facial expressions. Hence, the results suggest that attended faces could bias ensemble performance even within the same emotional category. To test whether there was a “within-category” foveal input bias, in the current study, we used 11 intensity levels of foveal face and flankers (from disgusted to happy) and asked participants to make ensemble judgments on a rating scale with 11 levels (from 0 to 10). When using such rating scales, a central-tendency response bias -- the tendency for observers to prefer responses in the middle of a rating scale – is expected (Stevens, 1971; Boari & Ruscone, 2015; Douven, 2018; Aston et al., 2022). Here, we also investigated the central-tendency response bias induced by the rating scale.

In the current study, stimuli consisted of a 3 × 3 matrix of faces with the foveal face (a single face presented at fixation for 100 ms) either present (‘central-present condition’) or absent ('central-absent condition'), and 8 surrounding faces ('flankers'). In the central-present condition, the emotion of the central face was either congruent (i.e., both the central face and flankers were happy or disgusted) or incongruent (i.e., the central face was happy while flankers were disgusted or vice versa) with that of the flankers. We expected a foveal input bias: the ensemble performance was expected to be impaired in the incongruent compared to the congruent and central-absent conditions. To investigate how the foveal input bias manifested itself when the foveal face was of the same emotion but different intensity than the average, we used 11 intensity levels of happy and disgusted facial expressions as foveal face and flankers (11 foveal faces × 11 flankers) and asked participants to make ensemble judgments on a rating scale with 11 levels (from 0 to 10; with '0' representing 'disgusted', '5' representing 'neutral', and '10' representing 'happy'). Using this rating scale, we tested whether there was a foveal input bias within the same emotional category: A foveal input bias would be observed when a foveal face of the same emotion as the flankers but different emotional intensity resulted in responses systematically biased towards the intensity of the foveal face. If there was no foveal input bias within the same emotional category, the performance was expected to be similar regardless of the intensity difference between the foveal face and flankers. Furthermore, we tested whether the central-tendency response bias occurred and if so, how it biased the ensemble performance. We expect that a central-tendency response bias would occur when using the rating scale as the response format. Taken together, the current study aimed to shed light on the role of the foveal face in ensemble emotion perception, especially when it has the same emotional category as the flankers.

3.2 Method

3.2.1 Participants

17 undergraduate students (18-24 years, 12 females) from Soochow University participated in the current study. All participants provided written consent before their participation. All participants were right-handed and had normal or corrected-to-normal visual acuity. The study was approved by the institutional review board at Soochow University.

3.2.2 Apparatus

The experiment was conducted on a standard 19” LCD display (E196FP, DELL) at a refresh rate of 60Hz and a resolution of 1280 × 1024. Participants were seated at a distance of 57 cm from the screen and gave their responses through an 11-point rating scale. E-prime 3.0 was used for stimulus presentation and data collection. A chin rest was used during the experiment.

3.2.3 Stimuli and Procedure

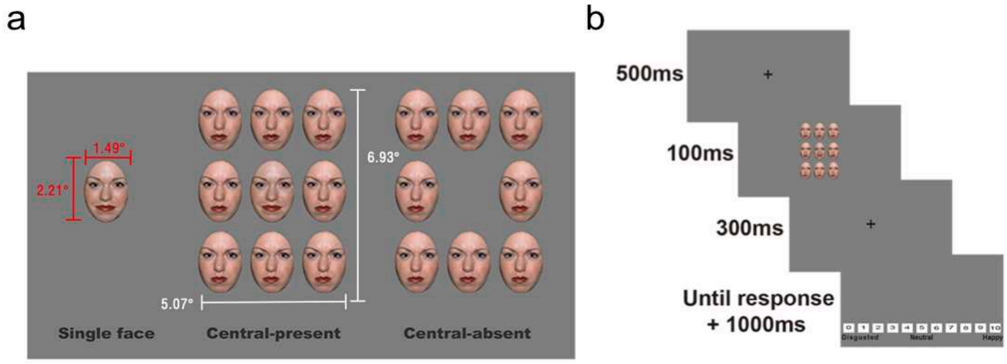
We used three facial expressions (happy, disgusted, neutral) of one female from the NimStim face set (Tottenham et al., 2009) to morph different intensity levels of facial expressions. Specifically, we morphed happy/disgusted faces with the neutral facial expression to generate four intensity levels of happy/disgusted facial expressions: 20%, 40%, 60%, 80%. There were 11 levels of facial expressions in total: 100% disgusted, 80% disgusted, 60% disgusted, 40% disgusted, 20% disgusted, neutral, 20% happy, 40% happy, 60% happy, 80% happy, 100% happy. All the external face contours and hair were removed. And the background of each image was replaced with a uniform gray (85 cd/m^2). Each face subtended a visual angle of 1.49° × 2.21° and was separated by 0.30° horizontally and 0.15° vertically from neighboring faces (the edge-to-edge distance). The whole face set (set size: 8 or 9) subtended a visual angle of 5.07° × 6.93° (Figure 1a).

We used a within-subjects design in which each participant viewed and rated facial expressions of a single face or a face set containing 8 or 9 faces. The task was to report the perceived average emotion of the face set (or the emotion of the single face) by using the mouse to click the corresponding option labels. An 11-point rating scale was used in the experiment (from 0 to 10; ‘0’ representing ‘disgusted’, ‘5’ representing ‘neutral’, ‘10’ representing ‘happy’). Faces were presented at the center of the screen. In the central-present condition, 11 levels of facial expressions were presented either as the central face or flankers (121 trials: 11 central faces × 11 flankers). In the central-present condition, the foveal face was either congruent or incongruent with that of flankers. In the congruent condition, both the central face and flankers were happy or disgusted. In the incongruent condition, the central face was happy while the flankers were disgusted or vice versa. In the central-absent condition, we repeated 11 flankers 11 times (121 trials: 11 flankers × 11 times) to match the trial number of the central-present condition. In the single face condition, 11 levels of facial expressions were repeated 11 times (121 trials: 11 faces × 11 times). The single face was presented at the center of the screen during the experiment. In total, there were 3 blocks (foveal-present condition, foveal-absent condition, single-face condition) and each block consisted of 121 trials. (In addition, the participants performed the same task with the same stimuli when the central face was centered at 3° or 8° eccentricity; results are reported in the supplementary material).

On each trial, a black fixation cross was presented for 500 ms, followed by a face set containing eight or nine faces, or a single face. Stimuli were presented for 100 ms in the center of the screen. After stimulus offset, a blank screen was presented for 300 ms, followed by the response screen. In separate blocks, participants were asked to judge either the face set's average emotion or a single face’s emotion on a rating scale from 0 (disgusted) to 10 (happy; 5 represented neutral). 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) Stimulus examples: a single face, a face set containing 9 faces (central-present condition), and a face set without the central face (central-absent condition). Stimuli consisted of either 1, 8, or 9 faces (9 faces were presented in the sample trial). **(b)** Trial sequence of one example trial: A face set including 9 faces was presented in the center of the screen for 100 ms. Participants were asked to report either the average emotion of the face set (or the single face's emotion) by clicking the corresponding number keys (0-10).

3.3 Analysis

We first compared participants' performance in the congruent, incongruent, and central-absent conditions (Figure 2). The performance was calculated based on the deviation between participants' reported emotional intensity and the actual intensity (i.e., the correct answer). Specifically, we calculated the mean squared error (MSE) in each condition, $MSE = \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{(Perceived\ emotional\ intensity_i - Actual\ intensity_i)^2}{n}$, where parameter i is the trial number and parameter n is the total number of trials in the corresponding condition. Bigger MSE values represent worse performance. The chance level was determined by calculating the MSE based on the deviation between random responses and the actual intensity. To assess the strength of the foveal input bias, we calculated the MSE for each valence and intensity of the foveal face (Figure 2b). To visualize the foveal input bias when the foveal face and flankers belonged to the same category, we plotted MSE against the difference between the foveal face and flankers (Figure 3a). The average responses for each combination of foveal face and flankers are shown in Figure 3b. To test the central-tendency response bias induced by the rating scale and its influence on the ensemble performance, we calculated the probability of each response option and compared them with random (unbiased) probabilities (Figure 4).

3.4 Results

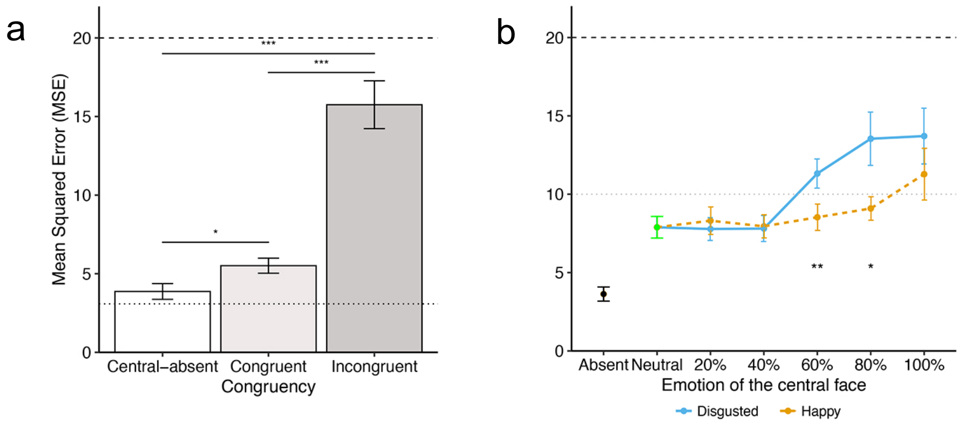
3.4.1 The foveal input bias

We compared participants' performance to identify the average emotion of the face set in congruent, incongruent, and central-absent conditions (Figure 2a). A repeated-measures ANOVA with MSE as the dependent variable showed a main effect of Congruency ($F(2, 32) = 47.09$, $p < 0.001$, partial $\eta^2 = 0.75$). The ensemble performance was strongly impaired in the incongruent compared to the congruent and central-absent condition (congruent (5.51 +/- 0.48) < incongruent (15.75 +/- 1.52): $p < 0.001$; central-absent (3.87 +/- 0.50) < incongruent: $p < 0.001$; central-absent < congruent: $p < 0.05$).

To assess whether the ensemble performance was also modulated by the emotional valence and the intensity of the foveal face, we compared MSEs as a function of the emotional valence and intensity of the foveal face (Figure 2b). MSEs in the central-absent condition show the ensemble performance without any influence from the foveal face. If the foveal face did not bias the ensemble performance, MSEs in the different emotional intensity conditions would be close to the MSEs in neutral and central-absent conditions. If MSEs in different emotional intensity conditions were bigger than MSEs in neutral and central-absent conditions, and increased with increasing emotional intensity of the foveal face, the result would show a bias induced by the foveal face. A 2 × 5 repeated-measures ANOVA with two factors (Valence (happy vs. disgusted) × Intensity of the central face (20%, 40%, 60%, 80%, 100%)) showed a significant main effect of Valence ($F (1,16) = 4.66$, $p < 0.05$, partial $\eta^2 = 0.23$) and Intensity ($F (4,64) = 10.21$, $p < 0.001$, partial $\eta^2 = 0.39$). There was also a significant Valence × Intensity interaction effect ($F (4,64) = 3.45$, $p < 0.05$, partial $\eta^2 = 0.18$): Performance was worse with increasing intensity of the foveal face (20%: 8.05 +/- 0.71; 40%: 7.88 +/- 0.73; 60%: 9.93 +/- 0.76; 80%: 11.31 +/- 0.88; 100%: 12.50 +/- 1.48). There was a difference between the happy and disgusted foveal face conditions with 60% (happy (8.53 +/- 0.84) < disgusted (11.32 +/- 0.93): *p* < 0.01) and 80% (happy (9.09 +/- 0.75) < disgusted (13.54 +/- 1.70): *p* < 0.05) intensity, but not between the lower intensities (20% happy (8.31 +/- 0.88) vs. 20% disgusted (7.78 +/- 0.73): *p* = 0.50; 40% happy (7.95 +/- 0.74) vs. 40% disgusted (7.81 +/- 0.83): *p* = 0.81; 100% happy (11.28 +/- 1.65) vs. 100% disgusted (13.71 +/- 1.78): *p* = 0.18).

Figure 2

MSEs for emotion recognition of face ensembles

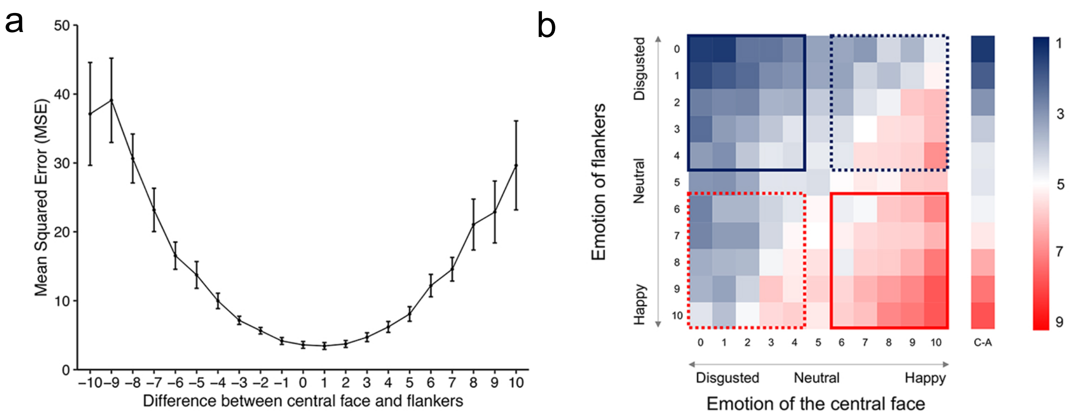


(a) MSE separated for the central-absent condition, and the congruent and incongruent central-present conditions. The dashed line represents the chance level. The dotted line represents the MSE in the single-face condition. Error bars represent ± 1 *SEM*. Asterisks indicate significance with alpha levels of 0.05 (*), 0.001 (***). Participants' ensemble performance in the central-absent condition was close to performance with a single face. The performance was most accurate in the central-absent condition, middle in the congruent condition, and worst in the incongruent condition. **(b)** MSE in the disgusted and happy central face conditions across the 5 emotional intensities of the foveal face (from 20% to 100%), the central-absent and neutral central face condition. The dashed line represents the chance level. The gray dotted line represents the MSE for always responding with the midpoint (i.e., '5' on the rating scale from 0-10). Error bars represent ± 1 *SEM*. Asterisks indicate significance with alpha levels of 0.05 (*), 0.01 (**). The performance was worse with increasing emotional intensity of the foveal face, and with the disgusted face presented as the foveal face.

To further illustrate how the foveal input bias manifested itself within the same emotional category, we calculated MSEs for each difference of the emotional intensity between the foveal face and the flankers (Figure 3a). The results showed that the MSE was minimal when the foveal face approached the emotional intensity of the flankers. The MSE was bigger when the foveal face’s emotional intensity was far from the flankers. Figure 3b depicts the average response for each combination of foveal face and flankers, showing how responses varied when the emotion of the foveal face and flankers belong to different and to the same category. The emotion of the foveal face is shown on the x-axis; the average emotion of the flankers on the y-axis, both varied from 100% disgusted to 100% happy. The results showed that the ensemble performance was strongly biased by the emotion of the foveal face: When the foveal face was disgusted, participants judged the average motion as disgusted even though the flankers were happy (and vice versa). In conditions where the foveal face had the same emotion as flankers but different emotional intensity, participants were biased by the foveal face’s emotion -- they reported the face set as more intense with the increased emotional intensity of the foveal face when the corresponding flankers were the same, revealing a pronounced foveal input bias within the same emotional category.

Figure 3

Ensemble performance in different combinations of foveal face and flankers



(a) MSE plotted against the difference between the foveal face and the flankers. The MSE was minimal when the foveal face had the same emotional intensity as the flankers. MSE increased with increasing (positive and negative) difference between the foveal face and the flankers. Error bars represent ± 1 *SEM*. **(b)** Each cell in the matrix represents participants' average rating with different combinations of foveal face and flankers. A value '0' represents reporting the face set as 100% disgusted, '5' represents reporting the face set as neutral, and '10' represents reporting the face set as 100% happy (the legend was set ranging from 1-9 based on the fact that the smallest average response was 1.45 and the highest one was 7.72). Blue (red) rectangles surround all cells where "disgusted" ("happy") was the correct response. Dashed rectangles surround cells of the incongruent condition where the foveal face had a different emotion from the flankers; solid rectangles surround cells of the congruent condition where the foveal face had the same emotion as the flankers but with different emotional intensity. "C-A" represents the central-absent condition, the values showed the average response modulated by flankers' emotion only. When the foveal face was of the same emotion but different intensity than the flankers, participants perceived the average emotion as more intense with increasing intensity of the foveal face.

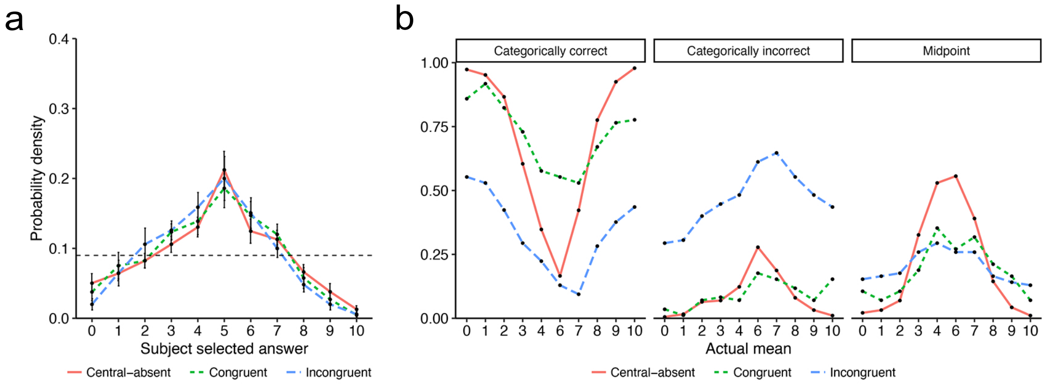
3.4.2 The central-tendency response bias

To test the central-tendency response bias expected by using the rating scale, we calculated the probability density distribution over the 11 possible response options separated by congruent, incongruent, and central-absent conditions (Figure 4). According to the correct answer, each response option's probability density was the same (i.e., 0.09). Since the average probability density of answering “10” (0.01 +/- 0.003) was close to 0, we excluded it in the subsequent statistical analysis. The data were analyzed with a repeated-measures ANOVA with Response (10 levels: response options from 0 to 9) and Congruency as independent variables. We found a significant main effect of Response (F (9,144) = 12.77, $p < 0.001$, partial $\eta^2 = 0.44$), but not Congruency (F (2,32) = 1.90, $p = 0.18$, partial $\eta^2 = 0.11$). There was no significant interaction effect (F (18,288) = 1.28, $p = 0.26$, partial $\eta^2 = 0.07$). The probability density distribution peaked when reporting ‘5’ as the response in the two congruency conditions (congruent and incongruent) and the central-absent condition.

Furthermore, we investigated the influence of the central-tendency response bias on ensemble performance. We separated the probability density of response in each actual mean condition into three subsets: (1) Categorically correct: participants responded happy/disgusted when the actual mean was happy/disgusted. For instance, when the actual mean was 7, all the possible responses from 6-10 were regarded as categorically correct responses. (2) Categorically incorrect: participants responded disgusted when the actual mean was happy (or responded happy when the actual mean was disgusted). For instance, when the actual mean was 7, all the possible responses from 0-4 were regarded as categorically incorrect responses. (3) Midpoints: the probability of choosing ‘5’ in each actual mean condition. In each actual mean condition, the sum of the probability in categorically correct, categorically incorrect, and midpoints subsets equaled 1. If the central-tendency response bias played an important role to explain participants’ responses, the probability density of answering midpoint would be generally high regardless of the actual emotional intensity of the face set (i.e., actual mean). In the Categorically correct subset, the probability density was in general higher in the congruent condition whereas, in the Categorically incorrect subset, the probability density was higher in the incongruent condition, reflecting the higher percentage of incorrect responses in the incongruent condition, which is consistent with the result of the foveal input bias. In the Midpoint subset, the curve in the central-absent condition represents the general central-tendency response bias. The curves in congruent and incongruent conditions were flatter compared to the central-absent condition, suggesting that the central-tendency response bias was less pronounced when the foveal face was present.

Figure 4

Probability density distribution for congruent, incongruent, central-absent conditions



(a) Probability density distribution of participants' responses separated by congruent, incongruent, and central-absent conditions. The dashed line represents the actual probability density of each response option based on the correct answer. Error bars represent ± 1 *SEM*. The probability density distribution peaked when reporting '5' as the response in the congruent, incongruent and central-absent condition. **(b)** Probability density distribution of participants' responses separated by the actual mean, two congruency conditions, and the central-absent condition. In the categorically correct condition, the curves represent the percentage of categorically correct responses. In the categorically incorrect condition, the curves represent the percentage of categorically incorrect responses. In the midpoint condition, the curves represent the percentage of answering '5' in each actual mean condition. The central-tendency response bias was less pronounced in the central-present condition (including congruent and incongruent conditions) when the actual mean was close to ‘5’.

3.5 Discussion

The current study investigated the foveal input bias when judging the average emotion of a group of faces. Our result showed a pronounced foveal input bias: the ensemble performance was significantly impaired when the foveal face was incongruent with that of flankers, and participants frequently reported the foveal face’s emotion as the average emotion of the face set. This was the case for different emotions of the foveal face and the flanker, and when the foveal face had the same emotion as the flankers but different emotional intensities where participants reported the average emotion as more intense with increasing emotional intensity of the foveal face. Using a rating scale for the responses, we found a significant central-tendency response bias: participants were more likely to choose the midpoints of the scale.

One potential mechanism underlying the foveal input bias is an attentional bias toward the foveal face, which leads to more weight than other unattended face members. Previous studies demonstrated that attended stimuli are preferentially weighted over other unattended stimuli in ensemble coding (de Fockert & Marchant, 2008; Im et al., 2015; Ying, 2022). 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. Meanwhile, increased attention may contribute to stronger or more accurate visual working memory (VWM) of the foveal face (Lepsien & Nobre, 2007; Thomas et al.,2014; Goldenberg et al., 2022), which in turn shapes the estimation of the mean. For instance, Goldenberg et al. (2022) investigated whether the stronger VWM for emotional faces predicted the ensemble performance. In their experiment (Goldenberg et al. (2022); see Study 5: testing memory based on facial expression intensity), 8 faces with different levels of happiness or anger were sequentially presented on each trial. Participants did ensemble and memory tasks in separate blocks. In the ensemble task block, participants were required to adjust the emotion of a probe face to the average emotion of the sequentially presented faces. In the memory task, participants reported which of the two probe faces was presented in the previous face sequence. Results showed that face members with stronger emotional intensities were better memorized and affected the ensemble evaluation of the face sequence. In the current study, we asked participants to fixate a single face, which could induce a stronger visual working memory trace of the foveal face, thus causing a bias toward the foveal face when reporting the ensemble emotional information. However, also the ensemble information can bias the VWM of a specific item (Brady & Alvarez, 2011). In Brady & Alvarez (2011)'s study, a display including three red, three blue, and three green circles was presented for 1.5 s. Participants were required to ignore the green circles and recall the size of one randomly chosen circle which was either red or blue. The result showed that observers' adjustment of the size of a given circle were biased toward the mean size of the same color circles. When the task was color-irrelevant (i.e., only presenting red and blue circles, and asked participants to simply remember the size of all circles), however, the bias toward the mean of the same color circles disappeared and observers' adjustments were generally biased toward the mean of the whole circle set, suggesting that selective attention modulated the VWM of the ensemble information, which in turn shaped the representation of single items.

Jung and colleagues (2015) suggested that participants were unconsciously biased by the faces they looked at directly. In the current study, we presented only a single face in foveal vision. Participants were required to fixate on the very same location in which the foveal face was presented. Presentation time was 100 ms, and thereby sufficiently short to prevent eye movements from the foveal face to other faces. The result showed a pronounced foveal input bias: participants frequently made ensemble judgments biased by the foveal face's emotion. In the condition where the foveal face had the same emotion as the flankers but different emotional intensity, participants reported the average emotion as less intense with less intense emotion of the foveal face, and more intense with increasing emotional intensity of the foveal face. Meanwhile, the reaction time in the congruent (1207.39 +/- 370.13 ms), incongruent

(1202.90 +/- 290.22 ms), and central-absent conditions (1153.80 +/- 246.21 ms) were comparable. The results seem to support that the foveal input bias may occur automatically. Recently we investigated whether the foveal input bias can be overcome by attentional control (Yu et al., 2021). In the study, we asked participants to either judge the average emotion of the face set (Experiment 1) or to ignore the foveal face and judge the average emotion of the other face members (Experiment 2). We found a pronounced foveal input bias when asked to integrate the foveal face and flankers' emotions. However, the foveal input bias disappeared when asked to voluntarily ignore the foveal face, suggesting that even though the foveal input bias might occur automatically, it can be modulated by selective attention.

The contribution of individual items to ensemble perception has been widely investigated in previous studies (Haberman & Whitney, 2010; Kanaya et al., 2018; Li & Yeh, 2017; Choi & Chong, 2020; Ueda, 2022). It was shown that observers tend to integrate only the square-root of the number of items during ensemble coding (see a review, Whitney & Yamanashi Leib, 2018). Meanwhile, what kind of individual items can be integrated into the ensemble is hotly debated. For instance, in saliency-based weighting, salient items are more heavily weighted than items that are less salient (Im et al., 2015; Kanaya et al., 2018; Cant & Xu, 2020), however, other studies showed that statistical outliers contribute less to the ensemble (see, for example, Haberman & Whitney, 2010; Epstein et al., 2020). In the current study, the foveal face became an outlier when surrounded by eight identical facial expressions which had different emotions from that of the foveal face (i.e., in the incongruent condition). The result showed a significant bias induced by the foveal face (i.e., the foveal input bias), meaning that participants did not ignore the outlier because it was the minority, rather, they gave more weight to it when doing ensemble judgment. The result seemed to support the saliency-based weighting mechanism in ensemble coding. However, in our recent work (Dandan et al., 2023), we attempted to make the foveal face less salient by manipulating the similarity of the foveal face and the flankers. Still, we found a pronounced foveal input bias, suggesting that the saliency of the foveal face is not the only reason for the observed foveal input bias.

The current study used different emotional intensities of happy and disgusted faces as the foveal face, thus providing us an insight into how the emotional valence and intensity of the foveal face modulate the ensemble performance. The result showed that disgusted facial expressions induced stronger foveal input bias than happy ones, especially with higher intensities (Figure 3a). Even in the single face condition, the discrimination was better with disgusted faces presented (MSE: 2.33 +/- 0.37) than the happy ones (MSE: 4.28 +/- 0.52). The results are consistent with previous demonstrations about the negativity superiority effect (Fox et al., 2000; Gong & Smart, 2021; Gong & Li, 2022; Goldenberg et al., 2021, 2022) – meaning that the detection of negative facial expressions have priority over the detection of positive facial expressions. In the current rating scale, the small values corresponded to disgusted facial expressions while big values corresponded to happy facial expressions, which could cause a possible response bias (e.g., the left-side bias, see for example, Chan, 1991). However, the probability density distribution across the 11-point responses (Figure 4a) showed that the percentage of answering small values were comparable with the percentage of answering bigger values, suggesting that the response bias for preferring small cannot explain the demonstrated negativity superiority effect. In our recent work, we found that when the foveal face contained no emotional information (i.e., scrambled face), the ensemble judgment was as accurate as the condition when no foveal face was presented, suggesting that the foveal input bias occurs only when emotional faces presented in the foveal vision. Similarly, Ueda (2022) investigated how emotional valence interacted with the spatial arrangement of face sets in ensemble coding. In their study, 12 faces with a mix of emotional and neutral facial expressions (i.e., either a mix of 100% happy and neutral faces or 100% angry and neutral faces) were presented. The ratio of emotional faces to neutral faces varied between trials. Participants were required to report which of two facial expressions (i.e., either neutral or emotional) was presented more frequently. In the 'distributed' condition, emotional faces were presented randomly in the face matrix, whereas in the 'dense' condition, emotional faces were presented in the center of the face matrix. The probabilities of positive responses (i.e., reporting emotional faces were presented more frequently) were significantly higher in the dense condition than in the distributed condition, suggesting that the emotional faces presented in the central visual field were more heavily weighted than other face members. In a follow-up experiment where emotional faces were presented at the corner of the face matrix (i.e., far from the central vision field), the probabilities of positive responses were comparable in the dense and distributed conditions. In the current study, we compared the ensemble performance in conditions with varying intensity levels of foveal face, the result showed that the performance was less accurate with more intense foveal faces presented compared to less intense ones (Figure 2b), suggesting that the emotional intensity of the foveal face modulated the ensemble performance. Furthermore, the visualization of the responses for each combination of foveal face and flankers (Figure 3b) showed that when the foveal face had the same emotion as the flankers but different emotional intensity, participants reported the ensemble emotion as more intense with the increased intensity of the foveal face, revealing a within-category foveal input bias in ensemble perception. Taken together, the results demonstrate that both the valence and intensity of the foveal face modulates the ensemble performance.

The central-tendency response bias is common when using continuous scales as the response format (Pimentel, 2000; Chyung et al., 2017; Douven 2018; Aston et al., 2022). As expected, we found that participants tend to choose midpoints of the scale. This was the case in the congruent, incongruent, and central-absent conditions. Even though we found a significant central-tendency response bias induced by the rating scale, the results still showed a pronounced foveal input bias, suggesting that the foveal input bias is pronounced despite the strong central-tendency response bias. The MSE calculated based on midpoints (i.e., 4,5,6; MSE: 4.53 +/- 0.64) was significantly lower than the real MSE (9.75 +/- 0.75) calculated based on participants' responses. However, the MSE calculated based on the foveal face only (8.03 +/- 0.92) was very close to the real MSE, showing that participants tended to report the foveal face's emotion as the average emotion of the face set. In another study (Dandan et al., 2022), we used the binary choice as the response format and still found a significant foveal input bias, suggesting that the foveal input bias is robust regardless of response formats.

Given that the rating scale used in the current study is susceptible to the central-tendency response bias, and thus interrupting the measurement reliability, we suggest using binary choice as an alternative to the Likert scale in future studies. First, our results showed that participants' response accuracy was similar when using the rating scale and binary choices as response formats. Second, continuous scales can induce acquiescence bias – a tendency to select a positive response option, central-tendency bias, and flatline issues – a tendency to consistently choose the same response (Ray, 1990; Qasem & Gul, 2014; Kuru & Pasek, 2016; Simms et al., 2019) and the number of response options also modulates the measurement reliability (Finn et al., 2015; Simms et al., 2019). Participants need more time to react based on the scale, causing higher reasoning costs (Givon & Shapira, 1984; Kunz, 2015; Rivera-Garrido et al., 2022). Meanwhile, when using binary choice as the response format, researchers can avoid the interruption from the central-tendency response bias. The binary choice is also more in accord with the general decision-making process (Harvey, 2016), and economically more efficient (Hilbert et al., 2016; Rivera-Garrido et al., 2022). However, in certain conditions when the continuous scale is necessary, researchers could eliminate the possible response bias either in the design (e.g., counterbalancing the location of the midpoint, see Goldenberg et al. (2022), or in the data analysis (e.g., testing how much variance the effect can be explained by the response bias induced by the scale).

3.6 Conclusion

To conclude, we found a pronounced foveal input bias – participants' ensemble performance was strongly impaired when the foveal face had a different emotion from that of flankers. Even in conditions where the foveal face had the same emotion as flankers but different emotional intensity, participants reported the average emotion as more intense with increasing emotional intensity of the foveal face. Our results suggest that foveal information is strongly overweighted when judging ensembles.

CHAPTER IV

Study 3: Foveal input bias in ensemble emotion perception can be overcome by voluntary control

ABSTRACT: Study 3

Individuals can extract the summary emotional information from groups of multiple faces, an ability called ensemble emotion perception. However, not every member in the face group is weighted equally. For example, it was shown that members presented around fixation weighed more in ensemble emotion perception, revealing a ‘foveal input bias’. To test whether the foveal input bias is ubiquitous, we investigated whether it can be overcome by voluntary control. In particular, we compared discriminability when participants were asked to report the average emotion of a face set with the discriminability when asked to voluntarily ignore the fixated face (the foveal input). The results showed a pronounced foveal input bias when asked to report the average emotion of a face set. However, the foveal input bias disappeared when participants were asked to voluntarily ignore the foveal face. The efficiency of voluntary control was unaffected by the variance of the emotion of the foveal face throughout a block of trials. The results suggest that the foveal input bias in ensemble perception can be overcome by attentional control.

Keywords: ensemble emotion perception, foveal input bias, attentional control

4. Study 3: Foveal input bias in ensemble emotion perception can be overcome by voluntary control

4.1 Introduction

Ensemble perception represents the visual system's ability to extract summary statistical information from groups of similar objects (e.g., Whitney & Yamanashi Leib, 2018). Multiple statistical features such as mean, variance, and distribution properties can be extracted during ensemble coding (Solomon 2010; Haberman et al., 2015; Chetverikov et al., 2016, 2017; Hansmann-Roth et al., 2019). Even though ensemble perception is flexible and reasonably efficient, how different objects are weighted is still under debate. According to the whole-set averaging account (Ariely, 2001; Chong et al., 2008; Joo et al., 2009), the ensemble information is extracted automatically and calculated based on all or most of the items in the stimulus set. By contrast, some studies demonstrated that only the square root of the number of items is integrated into the ensemble (see a review, Whitney & Yamanashi Leib, 2018), suggesting a subsampling 'strategy' in ensemble perception (Allik et al., 2013; Solomon et al., 2011; Maule & Franklin, 2016; Ji et al., 2020). Meanwhile, researchers showed that items close to the mean weigh more than items far from the mean in ensemble perception (Vandormael et al., 2017; Ni & Stocker, 2022; Iakovlev & Utochkin, 2023). Besides, it was shown that observers tended to give more weight to attended items (De Fockert & Marchant, 2008; Im et al., 2015), salient items (Kanaya et al., 2018; Cant & Xu, 2020; but see Epstein et al., 2020; Rosenbaum et al., 2021), most recent or earliest items in a sequence (Hubert-Wallander & Boynton, 2015; Tong et al., 2019; Goldenberg et al., 2022), and items in foveal vision (Ji et al., 2014; Jung et al., 2017; Dandan et al., 2023a). The availability of high-acuity foveal vision is important to visual perception (Rayner & Bertera, 1979; Geringswald et al., 2012; Geringswald & Pollmann, 2015; Kroell & Rolfs, 2022). For instance, in visual search tasks, the search accuracy was lower, and the search time was longer when the foveal vision was masked (Bertera & Rayner, 2000; Geringswald et al., 2012). Interestingly, foveal processing can bias peripheral object discrimination (Williams et al., 2008; Chambers et al., 2013; Fan et al., 2016; Weldon et al., 2020; Contemori et al., 2022).

The role of foveal input in ensemble perception is still debated. Some studies demonstrated that the foveal item weighs more than the extrafoveal items in ensemble perception (Ji et al., 2014; Jung et al., 2017). For example, by manipulating the congruency of foveal and extrafoveal faces, Ji and colleagues (2014) found that participants' accuracy in discriminating the average emotion of a face set was lower in the incongruent condition where foveal faces had a different emotion from extrafoveal ones than the congruent condition where both foveal and extrafoveal faces had the same emotion. Meanwhile, some studies showed that the foveal input is not necessary for ensemble perception (e.g., Wolfe et al., 2015). Recently, worse ensemble performance was found for stimuli with foveal items compared to stimuli viewed entirely in the periphery (To et al., 2019; Dandan et al., 2023a). Our recent work showed that the foveal input can strongly bias ensemble performance even when the foveal face is of the same emotional category as the other face members but with different emotional intensities (Dandan et al., 2023b, in preparation). Meanwhile, when presenting a face set with a foveal face, we found a pronounced foveal input bias regardless of the similarity between the foveal face and the surrounding face members ('flankers') and response methods (Dandan et al., 2023a, 2023b). For example, there was a strong foveal input bias when we used a binary choice paradigm and a rating-scale as the response format, suggesting that the foveal input bias might occur at the encoding stage, and that is not due to response bias.

One possible explanation for the foveal input bias is that the attentional allocation to the foveal item contributes to an overrepresentation of the item in the ensemble. For instance, when asking participants to judge the ensemble size of a circle set while locating a particular item (i.e., either the largest or smallest item), their ensemble judgment was shifted towards the size of the attended item (de Fockert & Marchant, 2008). If the foveal input bias is a result of focused attention to the foveal item, an important question is whether it can be overcome by top-down attentional control. There are only a few studies that have investigated the role of top-down attentional control in ensemble perception. Chong & Treisman (2005) demonstrated that distributed attention to a circle set facilitated ensemble size perception compared to focused attention to a foveal item. In their study, the distribution of attention was manipulated by requiring participants to judge the orientation of a large rectangular frame that framed the entire circle set. In the focused attention condition, participants were required to judge the orientation of a small rectangular frame located in foveal vision. Participants first judged which of the two probe circles that appeared in the test phrase matched the mean size of the previously presented circle set (ensemble task) and then judged whether the orientation of the rectangular frame was horizontal or vertical (orientation task). The results showed that the accuracy of the orientation task was better in the focused than in the distributed attention condition. However, mean size estimates were more accurate in the distributed than in the focused attention condition, suggesting that top-down attentional control can bias ensemble performance. Recently, Chen and colleagues (2021) found that a to-be-ignored ensemble still affected the judgment of the attended ensemble, suggesting that voluntary control may fail during ensemble coding (see also Oriet & Brand, 2013). Here, we aimed to investigate whether participants could successfully ignore a foveal face that caused a strong foveal input bias when not ignored.

In the current study, stimuli consisted of a 3 × 3 matrix of faces with the foveal face either present or absent and eight surrounding faces (‘flankers’). The aim was to investigate whether the foveal input bias is ubiquitous. In particular, we investigated whether the to-be-ignored foveal face still biased the ensemble performance by comparing discriminability (d′) when asking participants to report the average emotion of an entire face set (Experiments 1 & 3) with the discriminability when asking participants to voluntarily ignore the foveal face and judge the average emotion of the surrounding faces only (‘flankers’, Experiments 2 & 4). Like previous studies (Ji et al., 2014; Dandan et al., 2023), the foveal input bias was calculated by comparing the discriminability between congruent and incongruent conditions. In the congruent condition, the foveal face had the same emotion as the flankers. In the incongruent condition, the foveal face had a different emotion than the flankers. A significant lower discriminability in the incongruent condition would indicate a foveal input bias. If participants were able to ignore the foveal face successfully, there should be a diminished (or no) foveal input bias. The discriminability would be expected to be comparable in congruent and incongruent conditions.

Alternatively, if participants were not capable of completely ignoring the foveal face, we expected a pronounced foveal input bias. The discriminability would be expected to be lower in the incongruent than in the congruent condition. Taken together, the results can help better understand the effect of top-down attentional control in eliminating the foveal input bias in ensemble perception.

4.2 Experiment 1

The purpose of Experiment 1 was to investigate whether the foveal face determined the perceived emotion of face ensembles (i.e., the foveal input bias). We expected lower discriminability in the incongruent condition compared to the congruent condition (see also Ji et al., 2014; Jung et al., 2017; Dandan et al., 2023a).

4.2.1 Method

4.2.1.1 Participants

We conducted power analyses for sample size using the G*Power software (Version: 3.1.9.6; Faul et al., 2007, 2009). The power analysis showed that a repeated-measures ANOVA with three factors required 13 participants to identify a main effect ($f = 0.25$, power ($1-\beta$) = 0.80, $\alpha = 0.05$, and a correlation of 0.5; see Kang, 2021). A total of 14 participants (13 females, mean ± *SD* age = 20.64 ± 2.68 years) participated in the current experiment. All participants had normal or corrected-to-normal visual acuity, and gave written informed consent after being provided with an explanation of the experimental procedure.

4.2.1.2 Stimuli and design

We selected three facial expressions (happy, angry, neutral) of one female identity from the NimStim face database (Tottenham et al., 2009). The face images were trimmed to an oval shape and all external features, such as hair, ears, and neck, were cropped. To create different intensity levels of happy and angry expressions, we morphed happy/angry expressions with a neutral expression. There were 6 intensity levels: neutral, 20% happy/angry, 40% happy/angry, 60% happy/angry, 80% happy/angry, and 100% happy/angry. Each face image subtended a visual angle of 1.49° × 2.21°, and was separated by 0.30° horizontally and 0.15° vertically from neighboring faces (edge-to-edge distance).

In the experiment, 8 or 9 faces were presented on each trial and occupied a matrix of 5.07° × 6.93°. The central face located at fixation was the 'foveal face' and the 8 surrounding faces as flankers. The presentation time was 100 ms, short enough to prevent eye movements during stimulus presentation. The foveal face was either 'constant' or 'mixed'. In the constant condition, the foveal face was either 100% happy or 100% angry in separate blocks. For instance, in the block where the foveal face was 100% happy, all the trials within this block had the same foveal face (i.e., 100% happy). In the mixed condition, the foveal faces of 100% happiness and 100% anger were randomly interleaved within a block. As controls, there were three conditions where no foveal face was presented or a neutral or scrambled foveal face was presented in separate blocks. No foveal input bias was expected in the control conditions. The average intensity value of the face set was either 20%, 40%, 60%, 80%, or 100% happy/angry and was set based on flankers' expressions. Within each face set, the 8 flankers were either identical or varied. We did this to test whether the (un)grouping of the foveal face and the flankers modulated the expected foveal input bias (see also Dandan et al., 2023a). In the identical-flankers condition, all flankers had the same emotional intensity (e.g., all 8 faces were 60% happy). Each of the 10 emotional faces was presented in 8 trials, resulting in 80 trials per block. In the varying-flankers condition, we iteratively selected faces that maximized the number of different emotions within the set based on the constraint that the same face did not appear more than four times within a face set while keeping the variance of the face set in each average intensity condition relatively constant (average variance: 0.47 ± 0.14). The reason why we used relatively constant variance was to rule out the possible modulation of ensemble perception by stimulus set variance (Ji & Pourtois, 2018). As in the identical-flankers condition, each block consisted of 80 trials in the varying-flankers condition (i.e., there were 8 face combinations per average emotional intensity). In total, there were 960 trials (12 blocks ×

80 trials). There were two congruency conditions: (1) In the ‘congruent’ condition the emotion of the foveal face and the flankers was the same (either both happy or both angry); (2) In the ‘incongruent’ condition the emotions of the foveal face and the flankers were different (the foveal face was happy and the flankers were angry or vice versa). A foveal input bias would show as lower discriminability (d′) in the incongruent condition compared to the congruent condition.

4.2.1.3 Apparatus and procedure

The experiment was conducted in a dimly illuminated room. Participants sat 57 cm in front of a 19-inch LCD screen (E196FP, DELL). The screen had a refresh rate of 60 Hz and a resolution of 1280 × 1024 pixels. A chin rest was used. Participants were required to focus on the fixation during the experiment. On each trial, a fixation cross was presented at the center of the screen for 500 ms, followed by a face set containing 8 or 9 faces with a 100 ms duration. After a 300 ms blank, participants reported whether the presented face set was positive or negative by using ‘upper’ or ‘lower’ keys (the ‘upper’ key representing positive, and the ‘lower’ key representing negative). The subsequent trial started automatically 1000 ms after the participants’ response.

4.2.2 Analysis

Trials with response times (RTs) exceeding 3 SDs above or below the average RT for each participant were excluded (59 out of 13440 trials). We measured participants’ discriminability (d') and response bias (criterion) in each condition and conducted a repeated-measures ANOVA. We regarded angry faces reported as negative as ‘hits’, angry faces reported as positive as ‘misses’, happy faces reported as positive as ‘correct rejections’, and happy faces reported as negative as ‘false alarms’. 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.

4.2.3 Results

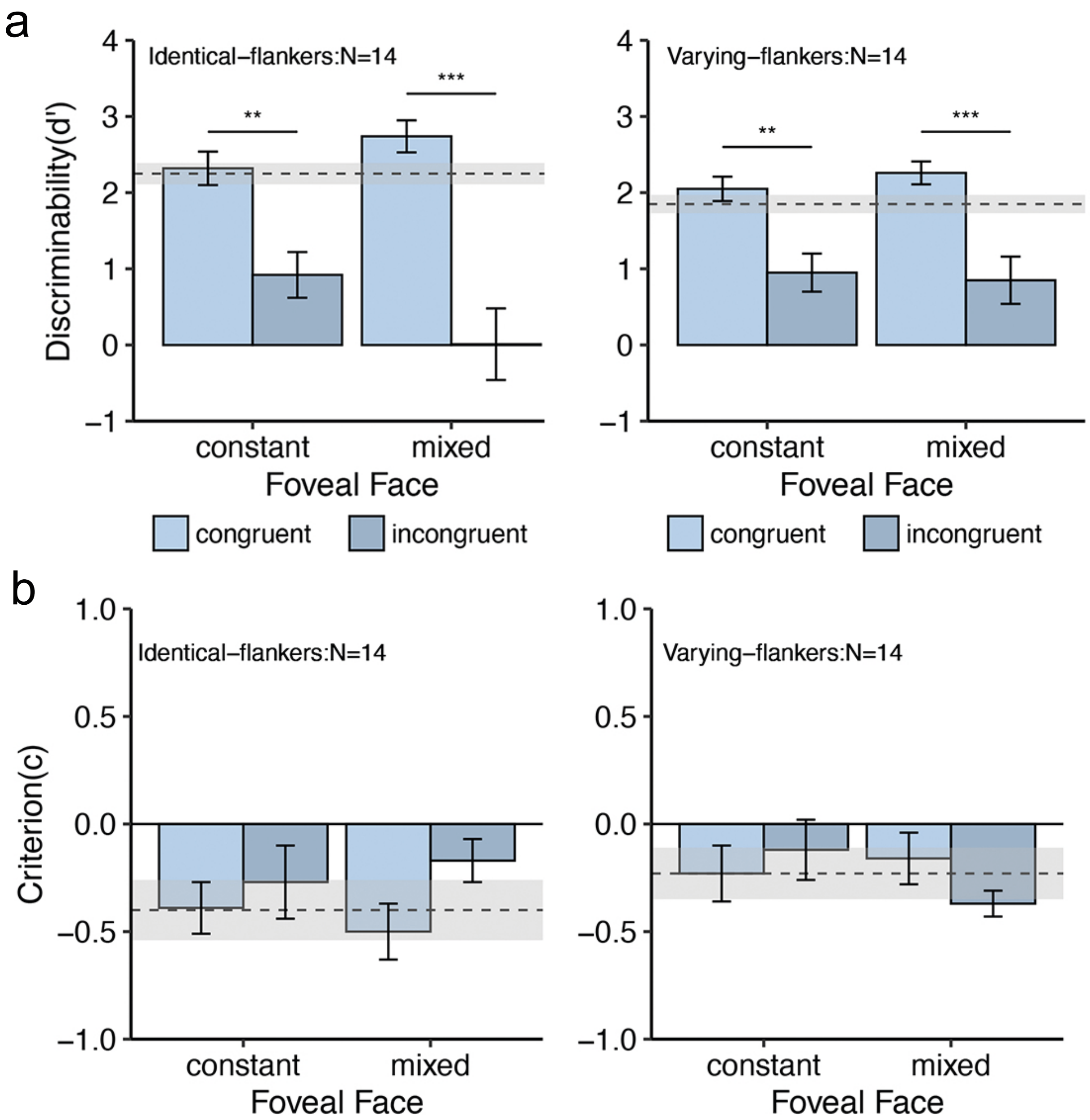
A repeated-measures ANOVA with three factors -- Congruency (congruent vs. incongruent), Foveal Face (constant vs. mixed), and Flankers (identical vs. varying) revealed a significant main effect of Congruency (F (1,13) = 27.93, $p < 0.001$, partial $\eta^2 = 0.68$) as well as interaction effects between Foveal Face and Congruency (F (1,13) = 4.64, $p < 0.05$, partial $\eta^2 = 0.26$) and between Flankers and Congruency (F (1,13) = 7.82, $p < 0.05$, partial $\eta^2 = 0.38$). The discriminability was lower in the incongruent than in the congruent condition. The discriminability difference between the two conditions was bigger in mixed ($p < 0.001$) than in constant foveal face condition ($p < 0.01$). In the congruent condition, the discriminability was lower in constant compared to the mixed foveal face conditions ($p < 0.05$). In the incongruent condition, the discriminability was comparable ($p = 0.13$). There was a trend for lower discriminability in the mixed (0.43 ± 0.35) compared to the constant (0.94 ± 0.26) condition. Meanwhile, the discriminability difference between congruent and incongruent conditions was more pronounced in the identical-flankers condition ($p < 0.001$) compared to the varying-flankers condition ($p < 0.01$). In the congruent condition, the discriminability was lower in the varying-flankers condition than in the identical-flankers condition ($p < 0.05$). In the incongruent condition, however, the discriminability was comparable ($p = 0.09$). There was no significant main effect of Foveal Face (F (1,13) = 0.41, $p = 0.53$, partial $\eta^2 = 0.03$), Flankers (F (1,13) = 0.05, $p = 0.83$, partial $\eta^2 = 0.00$) as well as interaction effect between Foveal Face and Flankers (F (1,13) = 1.70, $p = 0.22$, partial $\eta^2 = 0.12$). There was no three-factors interaction effect (F (1,13) = 3.75, $p = 0.08$, partial $\eta^2 = 0.22$).

To investigate the possible anger bias, we compared the criterion value first with zero, and found an overall bias to overreport face sets as angry in all conditions. When collapsed across all conditions, the average criterion (-0.28 ± 0.08) was consistently negative (one-sample t test: $t(13) = -3.25$, $p < 0.01$, $d = -0.87$). A repeated-measures ANOVA on the criterion yield main effects of Congruency (F (1,13) = 4.67, $p < 0.05$, partial $\eta^2 = 0.26$), and Flankers (F (1,13) = 11.79, $p < 0.01$, partial $\eta^2 = 0.48$). There was no main effect of Foveal Face (F (1,13)

= 0.20, *p* = 0.66, partial $\eta^2 = 0.02$) and no interaction effects (Foveal Face × Flankers: *F* (1,13) = 0.70, *p* = 0.42, partial $\eta^2 = 0.05$; Foveal Face × Congruency: *F* (1,13) = 0.17, *p* = 0.69, partial $\eta^2 = 0.01$; Flankers × Congruency: *F* (1,13) = 0.31, *p* = 0.59, partial $\eta^2 = 0.02$; Foveal Face × Flankers × Congruency: *F* (1,13) = 0.44, *p* = 0.52, partial $\eta^2 = 0.03$). Participants had a higher possibility to respond 'negative' in the identical-flankers condition (-0.38 ± 0.09) compared to the varying-flankers condition (-0.17 ± 0.09; *p* < 0.01). They also responded 'negative' more frequently in congruent (-0.32 ± 0.08) than incongruent conditions (-0.23 ± 0.09; *p* = 0.05).

Figure 1

Discriminability and criterion *for emotion recognition of face ensembles in Experiment 1*



Results of Experiment 1. Discriminability **(a)** and criterion **(b)** are separated for face sets with constant and mixed foveal faces (angry and happy expressions were interleaved within a block). Black horizontal lines represent discriminability and criterion in the absent foveal face condition and grey ribbons represent the corresponding standard error (*SEM*). Asterisks indicate significance with alpha levels of 0.01 (**) and 0.001 (***). Error bars represent ± 1 *SEM*.

4.2.4 Discussion

In the current experiment, we investigated the foveal input bias by comparing the discriminability in congruent and incongruent conditions. The results showed that participants' discriminability (d′) was lower in incongruent than congruent conditions, revealing a pronounced foveal input bias, consistent with previous demonstrations (Ji et al., 2014; Jung et al., 2017; Dandan et al., 2023a). Furthermore, the discriminability difference between congruent and incongruent conditions was more pronounced in mixed than constant foveal face conditions, suggesting that the variance of the foveal face might play a role in the strength of the foveal input bias. To investigate whether the foveal input bias is ubiquitous, in Experiment 2, we sought to investigate whether the foveal input bias can be overcome by top-down attentional control. To further investigate whether the variance of the foveal face played a role, we systematically manipulated the variance of the foveal face in Experiments 3 & 4.

4.3 Experiment 2

The main goal of the second experiment was to examine whether the pronounced foveal input bias found in Experiment 1 could be overcome by voluntary control. In the current experiment, we kept all the parameters the same as Experiment 1 but changed the task demands -- participants were required to voluntarily ignore the foveal face and judge the average emotion of the 8 flankers.

4.3.1 Method

4.3.1.1 Participants

12 new participants (7 females, mean ± *SD* age = 21.08 ± 1.88 years) participated in the current experiment. All participants had normal or corrected-to-normal visual acuity and were naïve to the purpose of the experiment.

4.3.1.2 Procedure

The stimuli and design were the same as Experiment 1, except that participants were explicitly asked to ignore the face presented at fixation (i.e., the foveal input), and judge the average emotion of the 8 flankers.

4.3.2 Analysis

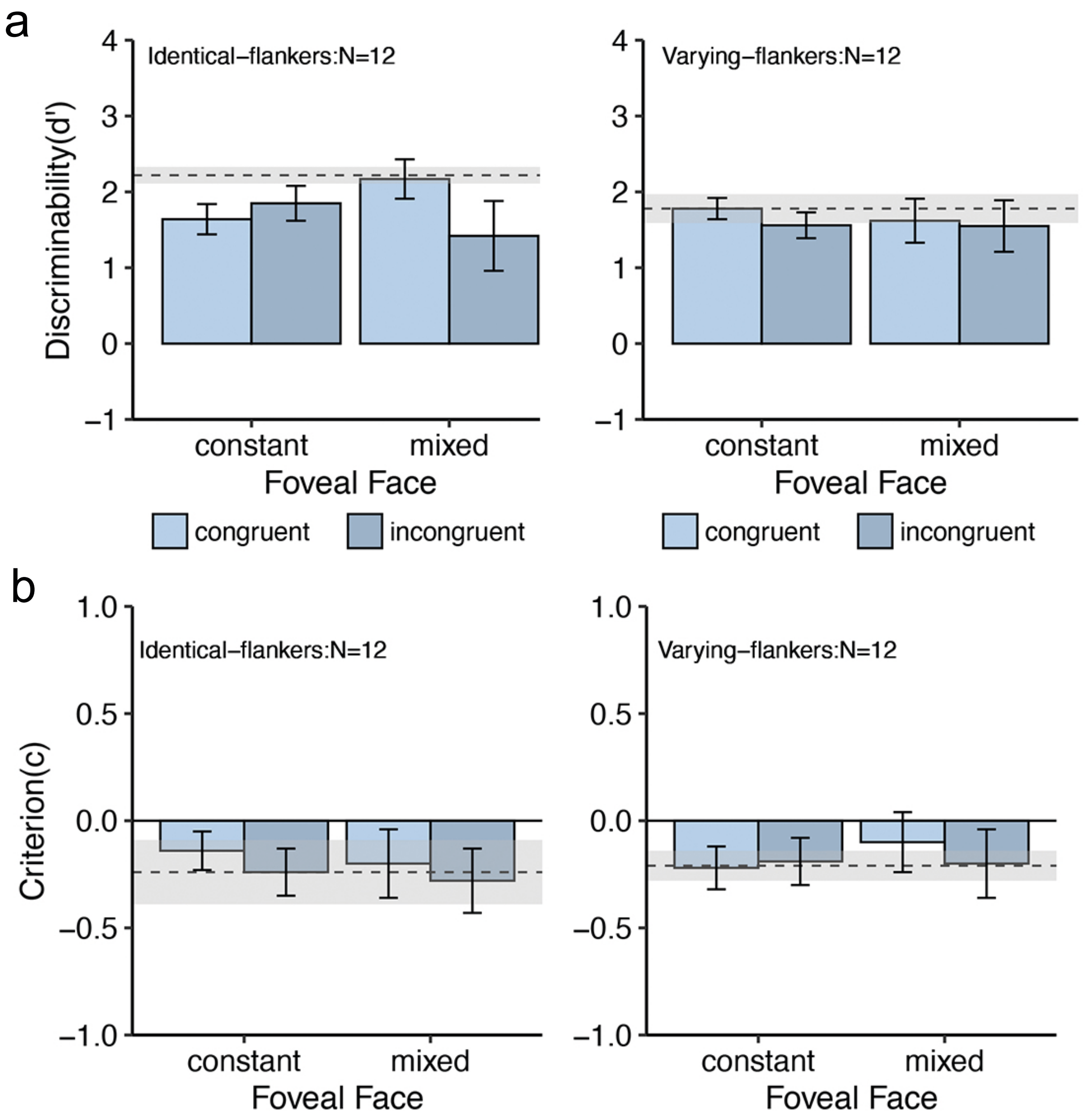
The data analysis in the current experiment was the same as in Experiment 1. Trials with response times (RTs) exceeding 3 SDs above or below the average RT for each participant were excluded (127 out of 11520 trials). We measured participants’ discriminability (d') and response bias (criterion) in each condition and conducted a repeated-measures ANOVA.

4.3.3 Results

We conducted a repeated-measures ANOVA with d′ as the dependent variable and factors of Congruency (congruent vs. incongruent), Foveal Face (constant vs. mixed), and Flankers (identical vs. varying). This analysis yielded no main effect of Congruency (F (1,11) = 1.76, p = 0.21, partial $\eta^2 = 0.14$), no main effect of Foveal Face (F (1,11) = 0.01, p = 0.91, partial $\eta^2 = 0.00$), no main effect of Flankers (F (1,11) = 1.17, p = 0.30, partial $\eta^2 = 0.10$), as well as no interaction effects (Foveal Face × Flankers: F (1,11) = 0.12, p = 0.73, partial $\eta^2 = 0.01$; Foveal Face × Congruency: F (1,11) = 1.79, p = 0.21, partial $\eta^2 = 0.14$; Flankers × Congruency: F (1,11) = 0.34, p = 0.57, partial $\eta^2 = 0.03$; Foveal Face × Flankers × Congruency: F (1,11) = 4.48, p = 0.06, partial $\eta^2 = 0.29$). When collapsed criterion values across all conditions, the average criterion was -0.20 (± 0.11), which was not significantly different from zero (t(11) = -1.85, p = 0.09, d = -0.53).

Figure 2

Discriminability and criterion *for emotion recognition of face ensembles in Experiment 2*



Results of Experiment 2. Discriminability **(a)** and criterion **(b)** are separated for face sets with a constant and mixed foveal face. Black horizontal lines represent discriminability and criterion in the absent foveal face condition and grey ribbons represent the corresponding standard error (*SEM*). Error bars represent ± 1 *SEM*.

4.3.4 Discussion

The critical question we tested in Experiment 2 was whether the to-be-ignored foveal face would still bias the ensemble performance. If the foveal input bias could not be overcome by the top-down attentional control, the discriminability would be lower in incongruent than congruent conditions, consistent with the finding of Experiment 1. On the contrary, if the discriminability was comparable in both conditions, the results would suggest that the foveal input bias can be overcome by voluntary control. Indeed, we found that the discriminability was comparable in the congruent and incongruent condition, as well as in the constant and mixed foveal face condition, suggesting that the foveal input bias is not ubiquitous but can be overcome by voluntary control.

4.4 Experiment 3

In Experiment 1, we found that the foveal input bias was more pronounced in the mixed compared to the constant foveal face conditions, suggesting that the variance of the foveal face might, to some extent, modulate the strength of the foveal input bias. To investigate this, in the current experiment, we systematically manipulated the variance of the foveal face by presenting three different variance levels of the foveal face (high, low, no variance) in separate blocks.

4.4.1 Method

4.4.1.1 Participants

14 participants (10 females, mean ± *SD* age = 24.5 ± 2.58 years) participated in this experiment. All participants reported normal or corrected-to-normal visual acuity.

4.4.1.2 Stimuli and design

The stimuli were the same as in Experiment 1. We manipulated the variance of the foveal face: (a) a high variance condition where the foveal face had different levels of happy and angry (i.e., a mix of 5 levels of happiness and anger (20%, 40%, 60%, 80%, 100% happy and angry); variance: 0.66, the standard deviation of the foveal face's emotional intensity within a block); (b) a low variance condition where the foveal face was different levels of either happy or angry (i.e., happy and angry foveal faces were not mixed within a block; variance: 0.28); (c) a no variance condition where the foveal face was kept constant (i.e., variance: 0), in other words, the foveal face was either happy, angry, 60% happy, 60% angry, neutral, or absent in separate blocks. The aim of adding 60% happy and 60% angry conditions was to test whether the intensity of the foveal face rather than the variance modulated the foveal input bias. The flanker manipulation was the same as in Experiments 1 & 2. There were an identical-flankers condition and a varying-flankers condition. In the identical-flankers condition, each of the 10 emotional faces (i.e., 5 intensity levels of happiness and anger) was presented as flankers and repeated 10 times in each block, resulting in 100 trials per block. In the varying-flankers condition, each of the 10 average values of the face matrix (i.e., 20%, 40%, 60%, 80%, 100% happy, and angry) was distributed in 10 trials (i.e., there were 10 unique stimuli per average value), resulting in 100 trials per block as well. The selection criterion of the face combination was the same as in the above two experiments. In total, we had 18 blocks (9 foveal face conditions × 2 flankers conditions). The procedure was the same as Experiment 1. Participants were required to judge whether the average emotion of the face set was positive or negative by pressing the left or right key. The key response was counterbalanced between participants.

4.4.2 Analysis

The data analysis was the same as in Experiment 1. Trials with response times (RTs) exceeding 3 SDs above or below the average RT for each participant were excluded (27 out of 25200 trials). We measured participants' discriminability (d') and response bias (criterion) in each condition and conducted a repeated-measures ANOVA.

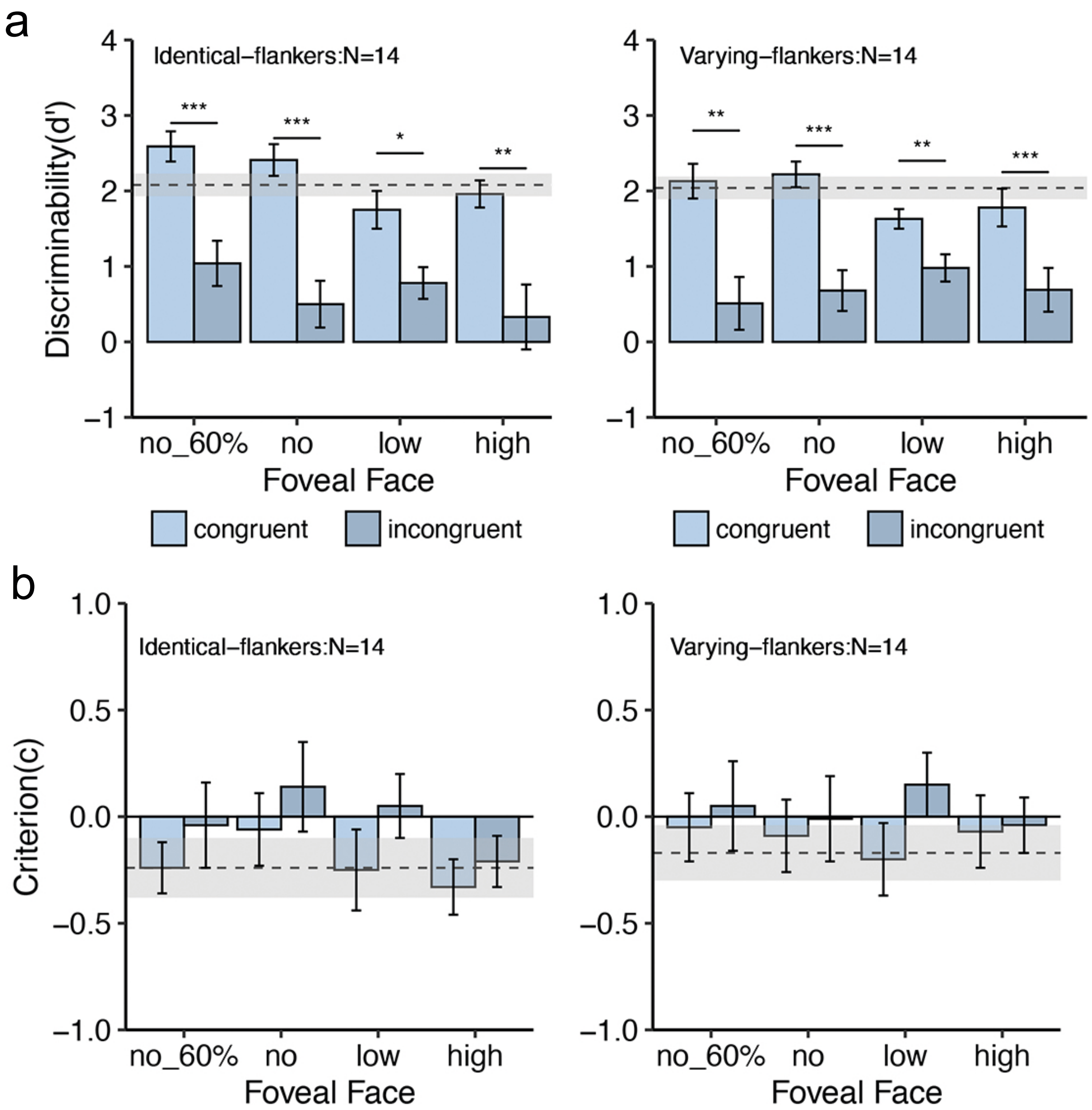
4.4.3 Results

To investigate the potential foveal input bias, a repeated-measures ANOVA with three factors (Congruency (congruent vs. incongruent) × Foveal Face (no, no_60%, low, high variance) × Flankers (identical vs. varying)) was conducted. The results showed a significant main effect of Congruency (F (1,13) = 34.60, $p < 0.001$, partial $\eta^2 = 0.73$), but not the main effects of Foveal Face (F (3,39) = 2.16, $p = 0.15$, partial $\eta^2 = 0.14$), and Flankers (F (1,13) = 1.63, $p = 0.22$, partial $\eta^2 = 0.11$). There were no significant interaction effects of Congruency and Foveal face (F (3,39) = 2.72, $p = 0.09$, partial $\eta^2 = 0.17$), Congruency and Flankers (F (1,13) = 1.98, $p = 0.18$, partial $\eta^2 = 0.13$), Foveal Face and Flankers (F (3,39) = 2.04, $p = 0.12$, partial $\eta^2 = 0.14$), and Congruency, Foveal Face, and Flankers (F (3,39) = 0.40, $p = 0.70$, partial $\eta^2 = 0.03$). The result showed that the d′ was significantly lower in incongruent than congruent conditions regardless of the variance, the intensity of the foveal face, and the homogeneity of flankers.

The criterion analysis (ANOVA) showed that there was a significant main effect of Congruency (F (1,13) = 17.08, $p < 0.001$, partial $\eta^2 = 0.57$). Participants were more likely to misclassify face sets as angry when the foveal face and flankers were congruent than when they were incongruent. There were no main effects of Foveal Face (F (3,39) = 1.79, $p = 0.18$, partial $\eta^2 = 0.12$) and Flankers (F (1,13) = 2.50, $p = 0.14$, partial $\eta^2 = 0.16$), no interaction effects of Congruency and Foveal Face (F (3,39) = 1.61, $p = 0.22$, partial $\eta^2 = 0.11$), Congruency and Flankers (F (1,13) = 0.81, $p = 0.39$, partial $\eta^2 = 0.06$), Foveal Face and Flankers (F (3,39) = 2.78, $p = 0.06$, partial $\eta^2 = 0.18$), and Congruency, Foveal Face, and Flankers (F (3,39) = 0.41, $p = 0.72$, partial $\eta^2 = 0.03$).

Figure 3

Discriminability and criterion *for emotion recognition of face ensembles in Experiment 3*



Results of Experiment 3. Discriminability **(a)** and criterion **(b)** are separated for face sets with a constant 60% intensity (no_60%), constant 100% intensity (no), slightly varied (low), and highly varied (high) foveal face. Black horizontal lines represent discriminability and criterion in the absent foveal face condition and grey ribbons represent the corresponding standard error (*SEM*). Asterisks indicate significance with alpha levels of 0.05 (*), 0.01 (**), and 0.001 (***). Error bars represent ± 1 *SEM*.

4.4.4 Discussion

Experiment 3 investigated whether the variance of the foveal face modulated the strength of the foveal input bias. The results seem inconsistent with the finding of Experiment 1 where the foveal input bias was more pronounced in the condition where the foveal face had a higher variance (i.e., the mixed condition). By systematically manipulating the variance of the foveal face, the current results showed that the foveal input bias was strong regardless of the variance of the foveal face, suggesting that the variance of the foveal face had no effect on the reported ensemble. Moreover, the discriminability was comparable in 60% happy/angry condition and 100% happy/angry condition, suggesting that the intensity of the foveal face did not modulate the effect of foveal variance in ensemble performance.

4.5 Experiment 4

The current experiment aimed to investigate whether the variance of the foveal face modulated the efficiency of voluntary control.

4.5.1 Method

4.5.1.1 Participants

13 participants participated in the experiment (12 females, mean ± SD age = 20.92 ± 1.77 years). All participants reported normal or corrected-to-normal visual acuity.

4.5.1.2 Procedure

The whole design was the same as Experiment 3 except that in the current experiment, participants were required to ignore the foveal face and judge the average emotion of the eight flankers (the task was the same as in Experiment 2).

4.5.2 Analysis

The data analysis in the current experiment was the same as in Experiment 3. We measured participants' discriminability (d') and response bias (criterion) in each condition and conducted a repeated-measures ANOVA.

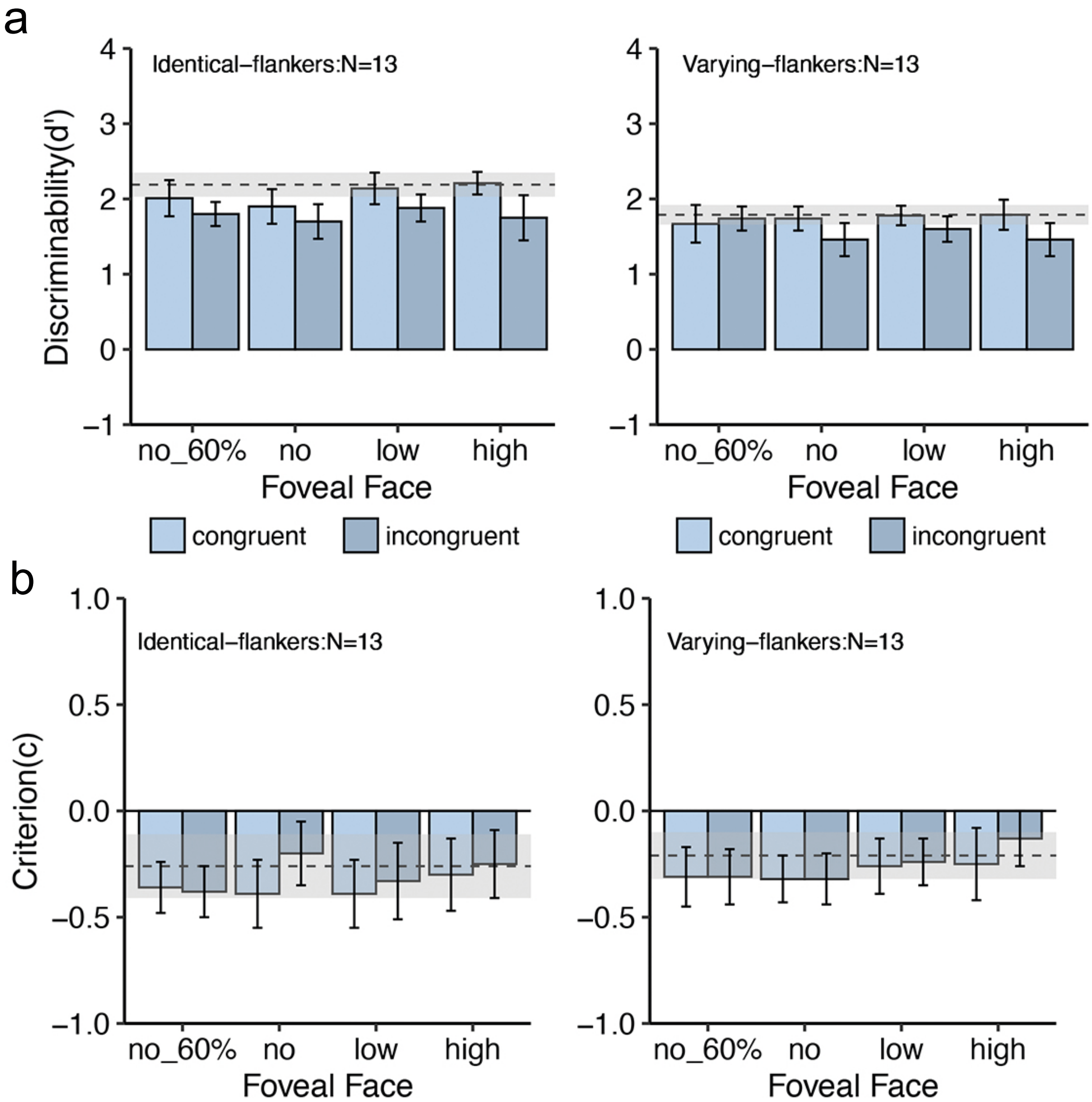
4.5.3 Results

A repeated-measures ANOVA with d′ as the dependent variable and Congruency (congruent vs. incongruent), Foveal Face (no, no_60%, low, high variance), and Flankers (identical vs. varying) as independent variables yield a main effect of Flankers (F (1,12) = 44.38, p < 0.001, partial η^2 = 0.79). The discriminability was higher when the flankers were identical compared to when they were varied. There was no main effect of Congruency (F (1,12) = 3.73, p = 0.08, partial η^2 = 0.24), Foveal Face (F (3,36) = 1.27, p = 0.30, partial η^2 = 0.10), as well as no interaction effects (Foveal Face × Flankers: F (3,36) = 0.52, p = 0.59, partial η^2 = 0.04; Foveal Face × Congruency: F (3,36) = 0.68, p = 0.54, partial η^2 = 0.05; Flankers × Congruency: F (1,12) = 0.86, p = 0.37, partial η^2 = 0.07; Foveal Face × Flankers × Congruency: F (3,36) = 0.30, p = 0.78, partial η^2 = 0.02).

A repeated-measures ANOVA with criterion as the dependent variable yield a significant main effect of Congruency (F (1,12) = 4.74, p < 0.05, partial η^2 = 0.28), showing that participants were more likely to evaluate faces as angry when the foveal face and the flankers were congruent compared to when they were incongruent. There was no main effect of Foveal Face (F (3,36) = 0.93, p = 0.42, partial η^2 = 0.07) and Flankers (F (1,12) = 1.09, p = 0.32, partial η^2 = 0.08) and no interaction effects (Foveal Face × Flankers: F (3,36) = 0.24, p = 0.80, partial η^2 = 0.02; Foveal Face × Congruency: F (3,36) = 0.80, p = 0.46, partial η^2 = 0.06; Flankers × Congruency: F (1,12) = 0.62, p = 0.45, partial η^2 = 0.05; Foveal Face × Flankers × Congruency: F (3,36) = 1.01, p = 0.39, partial η^2 = 0.08).

Figure *4*

Discriminability and criterion *for emotion recognition of face ensembles in Experiment 4*



Results of Experiment 4. Discriminability **(a)** and criterion **(b)** are separated for face sets with a constant 60% intensity (no_60%), constant 100% intensity (no), slightly varied (low), and highly varied (high) foveal face. Black horizontal lines represent discriminability and criterion in the absent foveal face condition and grey ribbons represent the corresponding standard error (*SEM*). Error bars represent ± 1 *SEM*.

4.6 Discussion

The current study investigated whether the foveal input bias is ubiquitous. In particular, we asked whether the foveal input bias can be overcome by voluntary control. To test this, we asked participants to either report the average emotion of the entire face set or ignore the foveal face and judge the average emotion of flankers only. The results showed that when participants were asked to judge the average emotion of the entire face set (including the foveal face), the ensemble performance was worse in the incongruent condition where the foveal face had a different emotion from that of flankers compared to the congruent condition where the foveal face had the same emotion as the flankers. Participants tended to report the foveal face's emotion as the average emotion of the face set (see Experiments 1 & 3), revealing a pronounced foveal input bias. However, when instructed to ignore the foveal face, the results showed that participants were not biased by the foveal face at all -- the performance was comparable in the congruent and incongruent condition (see Experiments 2 & 4). The generally increased discriminability with attentional control (vs. without attentional control) was driven by the incongruent condition where the foveal face barely biased the ensemble performance. In Experiments 3 & 4, we presented the foveal face with different variances within a block (i.e., no, low, and high variance) and found a pronounced foveal input bias regardless of the variance of the foveal face when asked to judge the ensemble of the whole face set (Experiment 3). Meanwhile, the variance of the to-be-ignored foveal face in Experiment 4 did not yield any significant foveal input bias, suggesting that individuals can successfully ignore the foveal face in ensemble emotion perception when instructed to do so.

It has been argued that some items are overrepresented in ensemble perception ('weighted averaging'; De Fockert & Marchant, 2008; Im et al., 2015; Hubert-Wallander & Boynton, 2015; Kanaya et al., 2018; Tong et al., 2019; Goldenberg et al., 2022). In the current study, we showed that participants gave more weight to the foveal face, revealing a foveal input bias in ensemble perception (see also Ji et al., 2014; Jung et al., 2017; Dandan et al., 2023a). Recently, Ueda (2022) showed that the foveal face only biased ensemble emotion perception when emotional rather than neutral faces were presented in foveal vision. The current study (Experiment 1: without voluntary control) compared response accuracy when the foveal face was happy than when it was angry, the result showed that ensemble performance was strongly impaired when angry faces were presented in foveal vision compared to happy faces. The results suggest that the emotional valence modulates the foveal input bias in ensemble perception, which is consistent with our previous findings (Dandan et al., 2023b, in preparation).

Experiment 1 showed how foveal faces biased ensemble perception. Remarkably, we found that the foveal face can be ignored in ensemble perception. When asking participants to voluntarily ignore the foveal face and report the average emotion of the flankers, they were able to ignore the foveal face and correctly report the average. The performance was comparable with the control condition where the foveal face was absent. There are several possible reasons for the efficiency of voluntary control in ensemble perception. First, participants could probably adopt a strategy not to attend to the central face much. In that way, there is no need to 'subtract' the foveal face's emotion. The average reaction time with and without voluntary control was comparable (without voluntary control (Exp.1): 750 ms; with voluntary control (Exp.2): 741 ms), revealing that the attention control did not cost extra attentional resources. Second, there is a possibility that participants could segregate the face set into two subsets on the basis of location and process the two sets concurrently. Chong and Treisman (2005) demonstrated that participants could calculate the average size of two sets of items as easily as one. In their study, a display containing two circle sets was presented, one in red and one in green. Participants were required to judge which of the two probe circles was the average of the relevant set. The relevant set was either informed by a pre-cue (two lines appeared in the relevant color prior to the display) or post-cue (the probe circles had the same color as the relevant set). In the control condition, one single circle set was present. The results showed that participants' discriminability was comparable in pre-cue and post-cue conditions, as well as in the control condition. In the current study, we did not test whether participants still extracted the foveal face's information when asked to ignore it, however, by comparing the accuracy in congruent and incongruent conditions (Experiment 2: with voluntary control), the result showed that the accuracy was slightly lower in the incongruent condition when the foveal face was angry. However, the accuracy in the two conditions was comparable when the foveal face was happy, suggesting that angry foveal faces could not be easily ignored compared to happy ones, which in some way demonstrated that to some extent, the foveal face information was processed even though it was irrelevant to the task.

The current study also tested whether the variance of the foveal face modulated the foveal input bias. In Experiment 3, we used constant, low, and highly varied foveal faces and found that the variance of the foveal face had no effect on the strength of the foveal input bias -- participants were frequently biased by the foveal face in the incongruent condition regardless of its variance. More importantly, when the variance information was not fixed within a block (i.e., in low and high variance conditions, Experiment 4), participants could still ignore the foveal face when instructed to do so. There are studies showed that the to-be-ignored item biased performance when the variance information was not fixed within a block (e.g., Theeuwes & Burger, 1998). Specifically, Theeuwes & Burger (1998) showed that the to-be-ignored salient singleton biased target discrimination when its color information was not fixed within a block. Participants could only successfully ignore the salient singleton when the color information of the singleton and other items in the stimuli set were kept constant. Notably, the salient singleton in their study was presented in the periphery. Beck and Lavie (2005) investigated whether participants could voluntarily ignore the distractor presented in the fovea vision. In their study, the distractor was the letter X or N, and it was either congruent (both the distractor and target were X or N) or incongruent (the distractor was X and the target was N or vice versa) with the target letter. Participants were required to focus on the fixation and search for the target letter among the other five letters presented in the visual periphery. Results showed that participants were slower in incongruent than congruent conditions, demonstrating a fixation distractor effect. One possible explanation for the inconsistent result (i.e., whether participants can successfully ignore the foveal item when instructed to do) is that it might result from differences in the task demands for targets in the periphery. The aforementioned two studies investigated the distractor interference effect in visual search tasks. In our study, we investigated how the distractor (i.e., the to-be-ignored foveal face) biased the ensemble performance. As has been shown, ensemble perception may occur with limited attentional resources (Chong & Treisman, 2003, 2005; Joo et al., 2009; Corbett & Oriet, 2011). Previous demonstrations showed that participants could still extract the average information of to-be-ignored items (Oriet & Brand, 2013; Chen et al., 2021). Further studies could investigate to what extent the saliency of the foveal face modulates voluntary control in ensemble perception. For instance, to investigate whether participants could still successfully ignore the foveal face in the ensemble task when the foveal face was an angry male/female facial expression while flankers were happy female/male expressions.

The present criterion results are consistent with previous demonstrations about the anger bias (Gong & Smart, 2021; Mihalache et al., 2021). It has been shown that angry facial expressions are detected more quickly than neutral and happy ones (Fox et al., 2000; Öhman et al., 2001; Shasteen et al., 2014). When facing a crowd of faces, participants were more likely to perceive them as angry (Mihalache et al., 2021). By comparing the criterion value in congruent and incongruent conditions, we found that the negativity response bias was more pronounced in congruent than incongruent conditions -- participants were more likely to respond ‘negative’ when the foveal face had the same emotion as flankers, which is consistent with Goldenberg et al. (2021, 2022)’s finding. In their study, different intensity levels of happy or angry facial expressions were presented on each trial, and the two emotions were not intermixed within a trial. Participants were required to judge the average emotion of the face set. The results showed that participants tended to overestimate the emotionality of the face set, revealing a crowd-amplification effect. The amplification effect was more pronounced with angry facial expressions than happy ones. Here, we extended Goldenberg et al.’s finding by demonstrating that the anger bias exists when mixing happy and angry facial expressions in the face set.

4.7 Conclusion

To summarize, we found that the ensemble judgments were systematically affected by the emotion of the foveal face, revealing a pronounced foveal input bias. This effect occurred regardless of the variance of the foveal face. When the foveal face had a different emotion from that of flankers, participants were more likely to report the foveal face’s emotion as the average emotion of the face set. However, the foveal input bias is not ubiquitous. When asked to voluntarily ignore the foveal face, the ensemble performance was generally more accurate than the condition without voluntary control, and the foveal input bias disappeared, suggesting that the foveal input bias can be overcome by voluntary control.

CHAPTER V

General Discussion

5. General discussion of the dissertation

In the current dissertation, I investigated the foveal input bias in ensemble perception. I first summarized the main findings of the three studies mentioned in previous chapters. Second, I discussed the contribution of the current dissertation to our understanding of ensemble perception. Last, I presented the limitations of current studies and some perspectives for future work based on the reported findings.

5.1 Overview of the current dissertation

Study 1 investigated to what extent, the foveal input can bias ensemble performance. Stimuli consisted of a 3 × 3 matrix of faces with a foveal face either present or absent, and 8 identical (Experiment 1) or varied (Experiment 2) surrounding faces ('flankers'). In the congruent condition, the foveal face had the same emotion as the flankers (both were happy or disgusted). In the incongruent condition, the foveal face had a different emotion from the flankers (the foveal face was happy while the flankers were disgusted or vice versa). In the central-absent condition, there was no foveal face; only the 8 flankers were presented. The significantly impaired performance in the incongruent compared to the congruent and central-absent conditions indicated a foveal input bias. As a control, a single face was presented in the foveal location. The results showed that performance was worse in the incongruent compared to congruent and central-absent conditions. In the incongruent condition, participants frequently reported the foveal face's emotion as the average emotion of the face set, revealing a pronounced foveal input bias. Besides, the ensemble performance in the central-absent condition was as accurate as the discrimination of the single face, showing the efficiency of ensemble coding (see also Chong & Treisman, 2005; Haberman & Whitney, 2009; Li et al., 2016). With identical flankers, grouping of the flankers due to similarity – and, correspondingly, ungrouping of the flankers from the foveal face – could have made the foveal 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. The result showed a pronounced foveal input bias, suggesting that the homogeneity of flankers was not the reason for the foveal input bias. Taken together, the current study demonstrates that the foveal face can strongly bias the ensemble performance when it has a different emotion from other face members.

In the second study (Study 2), I investigated whether the foveal input bias occurred when the foveal face was of the same emotion but different intensity than the average. To test this, I used 11 intensity levels of foveal face and flankers (from 100% disgusted to 100% happy; 11 levels of foveal face × 11 levels of flankers) and asked participants to report the average emotion of the face set by using a 0-10 rating scale (with ‘0’ representing ‘disgusted’, ‘5’ representing ‘neutral’, and ‘10’ representing ‘happy’). Again, there were a congruent, an incongruent, and a central-absent condition. The study replicated the results of the foveal input bias in Study 1 -- the performance was significantly impaired in the incongruent compared to the congruent and central-absent conditions. When the foveal face was of the same emotion as the flankers but different emotional intensity, participants reported the average emotion as more intense with increasing emotional intensity of the foveal face, revealing a ‘within-category’ foveal input bias. Furthermore, the discrimination was less accurate when the foveal face was disgusted than when it was happy, especially when the foveal face’s emotion was more intense. In the single-face condition, the discrimination of the disgusted face was more accurate than the discrimination of the happy face. The result is consistent with previous findings on the negativity superiority effect – meaning that negative facial expressions are detected with greater priority compared to positive facial expressions (Gong & Smart, 2021; Gong & Li, 2022; Goldenberg et al., 2021, 2022). Last, I found a significant central-tendency response bias -- a tendency for observers to frequently respond the midpoint of the scale (i.e., ‘5’) when using the rating scale as the response format. However, the foveal input bias was pronounced despite the strong central-tendency response bias. Taken together, the results suggest that the foveal input bias occurs when the foveal input is of the same emotional category as other face members.

Study 3 further investigated whether the foveal input bias is ubiquitous. In particular, I investigated whether the foveal input bias can be overcome by attentional control. In the study, I asked participants to either judge the average emotion of the entire face set (Experiment 1) or to ignore the foveal face and judge the average emotion of the flankers (Experiment 2). As in the two aforementioned studies, I manipulated the congruency of the foveal face and flankers to investigate the possible foveal input bias. I found a pronounced foveal input bias when asked to judge the average emotion of the entire face set, participants frequently reported the foveal face's emotion as the average emotion of the face set. However, the foveal input bias disappeared when participants were asked to ignore the foveal face. In Experiment 1, participants' ensemble judgment in the incongruent condition was less accurate with mixed (i.e., happy and angry foveal faces were interleaved within a block) than constant (i.e., the foveal face was either happy or angry; happy and angry foveal face were not mixed within a block) foveal face, suggesting that the variance of the foveal face might modulate the ensemble performance. In the follow-up experiments (Experiment 3 & 4), I systematically varied the variance of the foveal face to investigate whether participants could ignore the foveal face when it was varied. I found that the variance of the foveal face had a slight effect on the efficiency of attentional control -- participants could successfully ignore the foveal face regardless of its variance. Together, the results demonstrated that the foveal input bias is not ubiquitous, but can be overcome by attentional control.

5.2 Original contribution of the current dissertation

The current dissertation systematically investigated the role of the foveal input in ensemble emotion perception. I found a pronounced foveal input bias when the foveal face had different emotions from the flankers. Participants frequently reported the foveal face's emotion as the average emotion of the face set. In conditions where the foveal face was of the same emotional category as the flankers but different emotional intensities, participants' ensemble judgments were biased toward the emotion of the foveal face, revealing a 'within-category' foveal input bias. Previous studies on ensemble emotion perception within the same emotional category demonstrated a crowd emotion amplification effect -- the perceived average emotion is more intense when presenting a face set compared to a single face

(Goldenberg et al., 2021, 2022). The current dissertation revealed that, in addition to the amplification effect, there is also a bias toward the face(s) being fixated on in the perceived ensemble emotion.

Furthermore, the current dissertation investigated whether the foveal input bias is ubiquitous. I found that top-down attentional control can overcome the pronounced foveal input bias. The foveal input bias can be explained by attention to the foveal face, thus causing more weight given to the attended face (de Fockert & Marchant, 2008; Im et al., 2015; Ying, 2022). In Study 3, I found that when asking participants to voluntarily ignore the foveal face and judge the average emotion of flankers only, the foveal input bias disappeared. The phenomenon could be explained by reduced attention to the foveal face during the experiment. The findings further support the idea that attention plays a crucial role in modulating the perceived ensemble.

5.3 Limitations and future directions

The current dissertation has a number of limitations. The first limitation relates to the attentional control we used in Study 3. In the study, I manipulated attentional control through task demands. Specifically, I asked participants to either judge the average emotion of the face set directly or voluntarily ignore the foveal face and judge the average emotion of flankers only. The results showed that a pronounced foveal input bias occurred when asked to judge the average emotion of the face set, but not when asked to voluntarily ignore the foveal face. However, whether the voluntary control occurred in the encoding or decision-making stage is still unclear. Attentional control can occur at both encoding and decision-making stages (Vogel et al., 2005, 2006). In the current study, when asking participants to voluntarily ignore the central face, they could probably adopt a strategy not to attend to the central face much. In that way, there is no need to ‘subtract’ the foveal face’s emotion at the decision-making stage. On the other hand, they could also divide the face set into two subsets -- the foveal face and flankers, and process the two subsets in parallel. Previous demonstrations showed that observers could extract the ensemble information of to-be-ignored items (Oriet & Brand, 2013; Chen et al., 2021). Meanwhile, ensemble perception is quick, efficient and occurs with limited attentional resources (Alvarez & Oliva, 2009; Li et al., 2016; Lin et al., 2022), thus making it possible to process two subsets' emotional information at the same time (Chong & Treisman, 2005). To investigate which strategy participants could have used during the experiment, in future experiments, participants will be required to do dual tasks – reporting the foveal face in one task, and the average in the other task within the same trial, and extract different facial information from the foveal face and the entire face set. For instance, reporting the gender information of the foveal face and the emotional information of the face set. Doing the two tasks would require participants to attend to the entire face set as much as possible. Additionally, selective attention to the face set or the foveal face can be manipulated by varying the probability of performing the ensemble task and the foveal face task. For example, asking participants only in a small number of trials during a block to report the foveal face will most likely reduce attention to the foveal face. Participants are expected to be less biased by the foveal face in this condition, consistent with the result of Study 3.

A second limitation relates to how the foveal input bias manifests itself in natural settings. In the current dissertation, the foveal input was manipulated by presenting a single face at fixation for a limited presentation time (i.e., 100 ms). However, it remains unclear how the foveal input bias would vary when fixating different parts of the stimulus. In my recent work, I investigated ensemble perception in the context of online meetings where facial expressions are dynamic and last longer. The results showed that participants tended to overestimate the emotional intensity of the face set, revealing an amplification effect (see also Goldenberg et al., 2021, 2022). When comparing the weight of each face member on the perceived ensemble, I found that the central face located at the start fixation weighted more heavily than the other face members in ensemble perception. In the next step, I plan to record eye movements and test how the foveal input bias varies when fixating on a varied number of faces in online meeting contexts.

5.4 Conclusion

To summarize, the current dissertation showed that the foveal input strongly biased ensemble performance. Study 1 showed a pronounced foveal input bias when the emotional category of the foveal face was different from the other face members in the face set. In Study 2, where the foveal face was of the same emotion as the other face members but of different emotional intensity, participants reported the average emotion as more intense with increasing emotional intensity of the foveal face, revealing a foveal input bias within the same emotional category. The foveal input bias disappeared when asking participants to voluntarily ignore the foveal face, suggesting that the foveal input bias can be overcome by top-down attentional control. Taken together, the findings of this dissertation reveal how the foveal input influences ensemble perception and indicate that ensemble perception can be disrupted by the presence of salient target information in central vision. Importantly, the results also suggest that top-down attentional control can help to overcome the bias and support accurate ensemble perception.

CHAPTER VI

Le résumé substantial

6. le résumé substantial

6.1 Perception d’ensemble

La perception d'ensemble représente la capacité du système visuel à percevoir les informations statistiques sommaires d'un groupe d'objets similaires (pour des revues, voir Alvarez, 2011 ; Whitney & Yamanashi Leib, 2018). La perception d'ensemble se produit non seulement lors de la perception de caractéristiques de bas niveau telles que la taille (Chong & Treisman, 2003, 2005 ; Haberman & Suresh, 2021), l'orientation (Dakin & Watt, 1997 ; Parkes et al., 2001), et le mouvement (Watamaniuk et al., 1989 ; Sweeney et al., 2012), mais aussi lors de la perception de caractéristiques de haut niveau telles que l'identité faciale (Neumann et al., 2013 ; Jung et al., 2017), l'attractivité (Luo & Zhou, 2018) et les expressions faciales (Haberman & Whitney, 2007, 2009 ; Fischer & Whitney, 2011). Par exemple, Ariely (2001) a présenté un ensemble de cercles (taille de l'ensemble : 4, 8, 12, 16) de différentes tailles. Les participants devaient juger si un cercle test présenté ultérieurement était plus grand ou plus petit que la moyenne de l'ensemble de cercles précédent. Les résultats ont montré que les participants étaient capables de discriminer avec précision la taille moyenne de l'ensemble de cercles, et que la discriminabilité était indépendante de la taille de l'ensemble, révélant que la perception d'ensemble se produisait avec une capacité illimitée. Dans le domaine des caractéristiques des expressions faciales, les chercheurs ont montré que lorsque l'intensité émotionnelle d'un visage testé était proche de l'émotion moyenne de l'ensemble des visages présentés précédemment. En particulier, les participants étaient plus enclins à considérer le visage test comme un membre de l'ensemble, ce qui suggère que les individus peuvent extraire les informations émotionnelles sommaires de groupes de visages (par exemple, Haberman et al., 2007, 2009).

Bien que de nombreuses études aient démontré la robustesse de la perception d'ensemble dans différents domaines de caractéristiques, on ne sait toujours pas si le traitement des caractéristiques de bas et de haut niveau repose sur un mécanisme commun. La relation entre la perception d'ensemble des caractéristiques de bas et de haut niveau est souvent étudiée en comparant la corrélation des performances dans les deux tâches (Haberman et al., 2015 ; Yörük & Boduroglu, 2020 ; Kacin et al., 2021 ; Kwon & Chong, 2023). Par exemple, Haberman et al. (2015) ont démontré l'existence d'un mécanisme de traitement d'ensemble indépendant pour les caractéristiques de bas et de haut niveau. Ils ont comparé la corrélation des performances lorsqu'on leur demandait de percevoir deux caractéristiques de bas niveau différentes (par exemple, l'orientation par rapport à la couleur), deux caractéristiques de haut niveau différentes (par exemple, l'identité du visage par rapport aux expressions faciales), et une caractéristique de bas niveau et une caractéristique de haut niveau (par exemple, l'orientation par rapport à l'identité du visage), et ont trouvé des corrélations positives entre deux caractéristiques de bas niveau et deux caractéristiques de haut niveau, mais aucune corrélation entre une caractéristique de bas niveau et une caractéristique de haut niveau. Conformément aux conclusions de Haberman et al. (2015), Kacin, Gauthier et Cha (2021) ont comparé l'erreur absolue moyenne lors de la perception de l'information d'ensemble de deux caractéristiques de bas niveau, la longueur et l'orientation, et ont trouvé une corrélation positive de performance entre les tâches de longueur et d'orientation, ce qui suggère qu'au moins dans une certaine mesure, il existe un mécanisme commun pour le traitement d'ensemble des caractéristiques de bas niveau. Dans certaines circonstances, cependant, les chercheurs ont montré que le traitement des caractéristiques de haut niveau partageait un certain nombre de caractéristiques avec le traitement des caractéristiques de bas niveau. Par exemple, la perception d'ensemble de la taille, de la direction du mouvement et des expressions faciales a montré un effet de récence prononcé - les éléments les plus récents présentés dans la séquence de stimuli pèsent plus que les autres dans la perception d'ensemble (Hubert-Wallander & Boynton, 2015 ; Goldenberg et al., 2022). La taille de l'ensemble et la durée des stimuli ont eu un léger effet sur la perception d'ensemble des deux domaines de caractéristiques (Ariely, 2001 ; Chong & Treisman, 2003 ; Chong et al., 2008 ; Haberman & Whitney, 2009 ; Li et al., 2016). En outre, certaines études ont montré que les participants pouvaient extraire simultanément plusieurs caractéristiques de l'ensemble, telles que la vitesse moyenne et la taille (Emmanouil & Treisman, 2008 ; Albrecht et al., 2012).

Parallèlement, ils pouvaient extraire non seulement les informations moyennes, mais aussi les informations sur la variance et la distribution de l'ensemble de stimuli (Solomon 2010 ; Haberman et al., 2015 ; Chetverikov et al., 2016, 2017 ; Hansmann-Roth et al., 2018), ce qui suggère que la perception de l'ensemble peut se produire de manière hiérarchique, ce qui signifie qu'elle peut se produire à différents niveaux de traitement dans le cerveau.

6.2 La relation entre la perception de l'ensemble et la perception de l'élément individuel

L'une des principales caractéristiques de la perception d'ensemble est qu'elle résume les informations individuelles en un ensemble. Des études antérieures ont démontré que la perception d'un ensemble d'éléments est aussi rapide et précise que le traitement d'un seul élément (Chong & Treisman, 2003 ; Haberman & Whitney, 2009 ; Haberman et al., 2015 ; Li et al., 2016). Par exemple, Chong & Treisman (2003) ont présenté simultanément deux séries de cercles dans les champs visuels gauche et droit. Les deux tableaux de cercles étaient soit homogènes (12 cercles de la même taille dans chaque tableau), soit hétérogènes (12 cercles de 4 tailles différentes dans chaque tableau), soit uniques (il n'y avait qu'un seul cercle de chaque côté). Les participants devaient indiquer quel cercle (tableau) avait la taille la plus grande ou la taille moyenne la plus grande. Les résultats ont montré que la discrimination était comparable dans les trois conditions, ce qui suggère que l'extraction d'informations moyennes était aussi précise que l'extraction d'informations relatives à un seul élément. Selon le modèle holistique (Furtak et al., 2022), le cerveau extrait l'information globale avant l'extraction de l'information individuelle, et l'information globale peut influencer la perception ultérieure de l'information individuelle (Navon, 1977 ; Hochstein & Ahissar, 2002 ; Campana & Tallon-Baudry, 2013 ; Furtak et al., 2022). Toutefois, les informations individuelles spécifiques peuvent fausser la perception de l'ensemble. Par exemple, selon l'effet d'amplification constaté dans la perception d'ensemble, les éléments les plus saillants sont surreprésentés dans le codage d'ensemble (pour les caractéristiques de bas niveau, voir Kanaya et al., 2018 ; Iakovlev & Utochkin, 2021 ; Choi & Chong, 2020 ; pour les caractéristiques de haut niveau, voir

Goldenberg et al., 2021, 2022). Par exemple, Goldenberg et ses collègues ont étudié l'effet d'amplification des expressions faciales présentées simultanément (Goldenberg et al., 2021) et séquentiellement (Goldenberg et al., 2022). Dans leur étude, 1 à 12 visages heureux ou en colère avec différents niveaux d'intensité émotionnelle ont été présentés (les expressions heureuses et en colère n'ont été mélangées dans aucun essai), suivis d'un seul visage sonde avec une émotion neutre. Les participants devaient ajuster l'intensité émotionnelle du visage sonde à l'émotionnalité moyenne de l'ensemble des visages. Les résultats ont montré que les participants avaient tendance à surestimer l'émotionnalité de l'ensemble de visages, et que cet effet d'amplification était d'autant plus prononcé que l'ensemble était grand (c'est-à-dire qu'il y avait plus de visages présentés dans l'ensemble de visages). Le suivi oculaire a montré que les participants passaient plus de temps à regarder les visages les plus émotionnels de l'ensemble de visages, ce qui pourrait être à l'origine de l'effet d'amplification (Goldenberg et al., 2021).

6.3 La variance des distributions de stimuli module la performance d’ensemble

Des études antérieures ont montré que les individus peuvent extraire non seulement l'information moyenne mais aussi la variance de l'ensemble de stimuli (Solomon, 2010 ; Haberman et al., 2015). Les deux caractéristiques interagissent également pour aider les individus à obtenir une impression générale de la scène (Corbett et al., 2012 ; Im & Halberda, 2013). Comment les informations relatives à la moyenne et à la variance interagissent-elles lors du codage d'ensemble ? Chong & Treisman (2003) ont montré que la variance avait un léger effet sur la perception de la taille moyenne. Dans leur étude, la variance de l'ensemble de cercles a été manipulée en faisant varier la distribution de l'ensemble de cercles - normale, uniforme, à deux pics et homogène. Haberman et Whitney (2009) ont étudié la discriminabilité de l'émotion moyenne d'ensembles de visages homogènes et hétérogènes. Dans la condition homogène, l'expression faciale de chaque membre du visage était la même, tandis que dans la condition hétérogène, elles variaient les unes par rapport aux autres (c'est-à-dire qu'il y avait quatre expressions faciales uniques dans chaque ensemble de visages). Les résultats ont montré que la discriminabilité était comparable dans les deux conditions (les mêmes résultats ont été trouvés dans une étude sur la perception de l'attractivité faciale d'un ensemble, voir Luo & Zhou, 2018). Par ailleurs, certaines études ont démontré que la variance module l'effet de la taille de l'ensemble dans la perception de l'ensemble. Lorsque la variance de l'ensemble de visages était relativement importante, l'augmentation de la taille de l'ensemble entraînait une moins bonne performance en matière de calcul de la moyenne. Au contraire, lorsque la variance est relativement faible, la taille de l'ensemble n'a pas d'incidence sur les performances en matière de calcul de la moyenne (Ji & Pourtois, 2018 ; Im et al., 2017 ; Marchant et al., 2013). Les études futures devraient continuer à explorer les interactions entre la variance et la moyenne dans le traitement statistique sommaire.

6.4 La Différents mécanismes de pondération impliqués dans la perception d'ensemble

Bien que les individus puissent rapidement et relativement précisément extraire les informations statistiques sommaires de groupes d'objets similaires, le mécanisme sous-jacent fait encore l'objet d'un débat. Plusieurs mécanismes de pondération ont été mis en évidence par des études antérieures : (1) Le modèle de la moyenne générale : selon ce modèle, la perception d'un ensemble se produit automatiquement (Chong & Treisman, 2005). Il a été démontré que la performance de l'ensemble n'était pas affectée par la taille de l'ensemble (Ariely, 2001 ; Chong & Treisman, 2005 ; Haberman & Whitney, 2007, 2009), le mode de présentation (simultané ou successif ; Goldenberg et al., 2021, 2022), la durée du stimulus (de 50 ms à 2000 ms ; Chong & Treisman, 2003 ; Li et al, 2016) et la distribution de l'ensemble (homogène ou hétérogène ; Chong & Treisman, 2003 ; Haberman & Whitney, 2009 ; Luo & Zhou, 2018) appuient cette affirmation ; (2) le compte de la moyenne pondérée : ce compte soutient que les éléments de l'ensemble de stimuli n'ont pas le même poids dans le codage d'ensemble. Certains éléments ont plus de poids que d'autres dans les jugements d'ensemble. Par exemple, les chercheurs ont montré que les éléments présents (De Fockert & Marchant, 2008 ; Im et al., 2015), les éléments saillants (Kanaya et al., 2018 ; Cant & Xu, 2020 ; mais voir Epstein et al., 2020 ; Rosenbaum et al., 2021), les éléments antérieurs ou récents de la séquence (Hubert-Wallander & Boynton, 2015 ; Tong et al..., 2019 ; Goldenberg et al, 2019 ; Goldenberg et al., 2022), les éléments proches de la moyenne de la distribution (Vandormael et al., 2017 ; Ni & Stocker, 2022 ; Iakovlev & Utochkin, 2023), et les éléments à la fixation (Ji et al., 2014 ; Jung et al., 2017 ; Dandan et al., 2023a) pèsent davantage dans le codage d'ensemble.

6.5 Le rôle de l’attention sur l’ensemble perçu

Le système visuel peut calculer des informations statistiques avec des ressources attentionnelles limitées (Alvarez & Oliva, 2008 ; Sekimoto & Motoyoshi, 2022). Des études antérieures ont montré que les informations statistiques sommaires pouvaient être calculées à partir d'objets encombrés dans la périphérie visuelle (Parkes et al., 2001 ; Fischer & Whitney, 2011), avec un temps de présentation limité et une grande taille d'ensemble (Ariely, 2001 ; Chong & Treisman, 2003 ; Li et al., 2016). Ces résultats ont démontré que la perception d'ensemble pouvait se produire sans que l'attention focale ne soit trop sollicitée. La perception d'ensemble et l'attention focale ont également été suggérées comme étant deux mécanismes distincts traitant de la capacité limitée du système visuel (Baek & Chong, 2020) - la perception d'ensemble fournit l'essentiel d'une scène, tandis que l'attention focale sélectionne les informations importantes de la scène pour reconnaître quelques objets.

Même si la perception d'ensemble se produit avec des ressources attentionnelles limitées, les éléments présents ont plus de poids que les éléments non présents pendant le codage d'ensemble (De Fockert & Marchant, 2008 ; Im et al., 2015). Les participants ont tendance à surestimer la taille moyenne des ensembles de cercles lorsqu'on leur demande de fixer le membre le plus grand du cercle et vice versa (De Fockert & Marchant, 2008). Le biais d'entrée fovéale démontré dans la perception d'ensemble (voir ci-dessous) pourrait également s'expliquer par la répartition de l'attention sur l'entrée fovéale. Plus précisément, des études antérieures ont montré que le(s) visage(s) situé(s) à la fovéa se voit(vent) accorder plus de poids que les visages en périphérie dans l'ensemble perçu (Ji et al., 2014 ; Jung et al., 2017), révélant ainsi un biais d'entrée fovéal. Par exemple, Ji et ses collègues (2014) ont étudié la contribution relative des visages fovéaux et extrafovéaux dans l'ensemble perçu. Dans leur étude, les participants ont été présentés avec un groupe de visages, composé de 4 visages fovéaux situés dans la vision centrale et 12 visages extrafovéaux présentés dans la vision extrafovéale. Les visages fovéaux et extrafovéaux ont été manipulés pour être soit congruents, soit incongrus en termes d'expression émotionnelle. Dans la condition congruente, les visages fovéaux avaient la même émotion que les visages extrafovéaux, tandis que dans la condition incongrue, les visages fovéaux avaient une émotion différente de celle des visages extrafovéaux. Les résultats ont montré que l'ensemble perçu était moins précis dans la condition incongrue que dans la condition congruente. Les participants ont souvent rapporté l'émotion des visages fovéaux comme étant l'émotion moyenne de l'ensemble lorsque les visages fovéaux transmettaient une émotion différente de celle des visages extrafovéaux. L'ensemble de ces résultats suggère que l'attention focalisée peut faciliter la perception de l'ensemble.

Les études susmentionnées sur la façon dont les valeurs aberrantes biaisent la perception de l'ensemble (Kanaya et al., 2018 ; Cant & Xu, 2020) ont examiné la possibilité que la perception de l'ensemble se produise de manière involontaire et guidée par le stimulus. Cependant, on ne sait toujours pas comment le contrôle attentionnel descendant biaise la performance de l'ensemble. Dans les tâches de recherche visuelle, par exemple, lorsqu'il est demandé aux participants d'ignorer volontairement un élément saillant, celui-ci ne peut être ignoré que si la caractéristique de l'élément à ignorer et de la cible est constante pendant l'expérience (Theeuwes & Burger, 1998 ; Wang & Theeuwes, 2018). Par exemple, dans l'étude de Theeuwes et Burger (1998), le singleton saillant à ignorer (lettre E ou R) était placé dans une couleur différente (c'est-à-dire que le singleton était rouge tandis que les autres éléments, y compris la cible, étaient verts ou vice versa). La couleur du singleton et des autres éléments était intercalée entre les essais, ce qui signifie que la couleur du singleton était soit rouge soit verte entre les essais, et que les autres éléments avaient toujours une couleur différente de celle du singleton à chaque essai. Le singleton était soit congruent soit incongruent avec la cible. Dans la condition congruente, le singleton et la cible étaient identiques (c'est-à-dire que les deux lettres étaient Es ou Rs). Dans la condition incongrue, ils étaient différents (le singleton était E et la cible était R ou vice versa). Les participants devaient ignorer le singleton sur la base des informations de couleur et juger si un E ou un R était présent parmi les lettres non singulières. Les résultats ont montré que les participants passaient plus de temps à identifier la cible dans la condition incongrue que dans la condition congruente, ce qui suggère que les singletons à ignorer ne pouvaient pas être ignorés avec succès mais biaisaient la performance de recherche. Cependant, lorsque la couleur du singleton et de la cible était constante (c'est-à-dire que le singleton était constamment rouge alors que les autres éléments étaient constamment verts dans l'expérience ou vice versa), les TR étaient comparables dans les conditions congruentes et incongrues, ce qui démontre que lorsque la caractéristique du singleton à ignorer et de la cible était constante, les individus pouvaient volontairement ignorer le singleton très saillant. Dans les tâches d'ensemble, les chercheurs ont montré que le ou les éléments à ignorer contribuaient au processus de calcul de la moyenne (Oriet & Brand, 2013 ; Chen et al., 2021), ce qui suggère que le contrôle attentionnel descendant peut échouer dans la perception d'ensemble. Par exemple, Alvarez & Oliva (2008) ont utilisé une tâche d'attention divisée dans laquelle les participants suivaient le mouvement des cibles (nuages de points) tout en ignorant les distracteurs, puis localisaient soit l'emplacement d'un élément manquant, soit le centroïde de quatre éléments manquants. Les éléments manquants étaient soit des cibles, soit des distracteurs. Les résultats ont montré que si la discrimination d'un seul distracteur manquant était proche du niveau de chance, les participants étaient capables de discriminer avec précision le centroïde de quatre distracteurs manquants, ce qui indique que les distracteurs à ignorer étaient encore traités (voir également Oriet & Brand, 2013 ; Chen et al., 2021).

6.6 Le cadre de la présente thèse

La présente thèse se compose de trois études. L'étude 1 a examiné dans quelle mesure l'entrée fovéale peut biaiser la performance d'ensemble. Les stimuli consistaient en une matrice de 3 × 3 visages avec un visage fovéal présent ou absent, et 8 visages environnants identiques (expérience 1) ou variés (expérience 2) ("flankers"). Dans la condition congruente, le visage fovéal avait la même émotion que les flankers (les deux étaient heureux ou dégoûtés). Dans la condition incongrue, le visage fovéal avait une émotion différente de celle des flankers (le visage fovéal était heureux alors que les flankers étaient dégoûtés ou vice versa). Dans la condition d'absence centrale, il n'y avait pas de visage fovéal ; seuls les 8 flancs étaient présentés. Les performances nettement inférieures dans la condition incongrue par rapport à la condition congruente et à la condition centrale absente indiquent un biais d'entrée fovéal. En guise de contrôle, un seul visage a été présenté dans la zone fovéale. Les résultats ont montré que les performances étaient moins bonnes dans les conditions incongrues que dans les conditions congruentes et absentes au centre. Dans la condition incongrue, les participants ont fréquemment rapporté l'émotion du visage fovéal comme l'émotion moyenne de l'ensemble des visages, révélant un biais d'entrée fovéal prononcé. En outre, la performance de l'ensemble dans la condition d'absence centrale était aussi précise que la discrimination d'un seul visage, ce qui montre l'efficacité du codage d'ensemble (voir également Chong & Treisman, 2005 ; Haberman & Whitney, 2009 ; Li et al., 2016). Avec des flancs identiques, le regroupement des flancs en raison de leur similarité - et, par conséquent, le dégroupement des flancs du visage fovéal - aurait pu faire ressortir le visage fovéal des flancs, ce qui aurait faussé les réponses. Pour déterminer si le biais d'entrée fovéal constaté dans l'expérience 1 était dû à l'homogénéité des flankers, nous avons fait varier les émotions des flankers dans l'expérience 2. Le résultat a montré un biais d'entrée fovéale prononcé, suggérant que l'homogénéité des flankers n'était pas la raison du biais d'entrée fovéale. Dans l'ensemble, l'étude actuelle démontre que le visage fovéal peut fortement biaiser la performance de l'ensemble lorsqu'il a une émotion différente de celle des autres membres du visage.

Dans la deuxième étude (étude 2), j'ai cherché à savoir si le biais d'entrée fovéale se produisait lorsque le visage fovéal avait la même émotion mais une intensité différente de la moyenne. Pour ce faire, j'ai utilisé 11 niveaux d'intensité du visage fovéal et des flancs (de 100% dégoûté à 100% heureux ; 11 niveaux de visage fovéal × 11 niveaux de flancs) et j'ai demandé aux participants d'indiquer l'émotion moyenne de l'ensemble de visages en utilisant une échelle d'évaluation de 0 à 10 (avec "0" représentant "dégoûté", "5" représentant "neutre", et "10" représentant "heureux"). Là encore, il y avait une condition congruente, une condition incongruente et une condition centrale-absente. L'étude a reproduit les résultats du biais d'entrée fovéale de l'étude 1 : la performance était significativement réduite dans la condition incongrue par rapport à la condition congruente et à la condition centrale absente. Lorsque le visage fovéal présentait la même émotion que les flankers, mais une intensité émotionnelle différente, les participants ont déclaré que l'émotion moyenne était plus intense lorsque l'intensité émotionnelle du visage fovéal augmentait, ce qui révèle un biais d'entrée fovéal "au sein de la catégorie". En outre, la discrimination était moins précise lorsque le visage fovéal était dégoûté que lorsqu'il était heureux, en particulier lorsque l'émotion du visage fovéal était plus intense. Dans la condition de visage unique, la discrimination du visage dégoûté était plus précise que celle du visage heureux. Ce résultat est cohérent avec les conclusions précédentes sur l'effet de supériorité de la négativité, qui signifie que les expressions faciales négatives sont détectées en priorité par rapport aux expressions faciales positives (Gong & Smart, 2021 ; Gong & Li, 2022 ; Goldenberg et al., 2021, 2022). Enfin, j'ai constaté un biais de réponse significatif de tendance centrale - une tendance des observateurs à répondre fréquemment au point médian de l'échelle (c'est-à-dire "5") lorsqu'ils utilisent l'échelle d'évaluation comme format de réponse. Cependant, le biais d'entrée fovéal était prononcé malgré le fort biais de réponse de tendance centrale. Dans l'ensemble, les résultats suggèrent que le biais d'entrée fovéale se produit lorsque l'entrée fovéale est de la même catégorie émotionnelle que les autres membres du visage.

L'étude 3 a cherché à savoir si le biais d'entrée fovéale était omniprésent. En particulier, j'ai cherché à savoir si le biais d'entrée fovéale pouvait être surmonté par un contrôle attentionnel. Dans cette étude, j'ai demandé aux participants de juger l'émotion moyenne de l'ensemble des visages (expérience 1) ou d'ignorer le visage fovéal et de juger l'émotion moyenne des flancs (expérience 2). Comme dans les deux études susmentionnées, j'ai manipulé la congruence du visage fovéal et des flancs afin d'étudier l'éventuel biais d'entrée fovéal. J'ai constaté un biais d'entrée fovéal prononcé lorsqu'on leur a demandé de juger l'émotion moyenne de l'ensemble des visages, les participants ont fréquemment déclaré l'émotion du visage fovéal comme étant l'émotion moyenne de l'ensemble des visages. Cependant, le biais d'entrée fovéal disparaissait lorsqu'on demandait aux participants d'ignorer le visage fovéal.

Appendix: Supplementary material of Study 2

In Study 2, the participants performed the ensemble task with the same stimuli when the central face was centered at 3° and 8° eccentricity. The supplementary material here showed the detailed analysis and results.

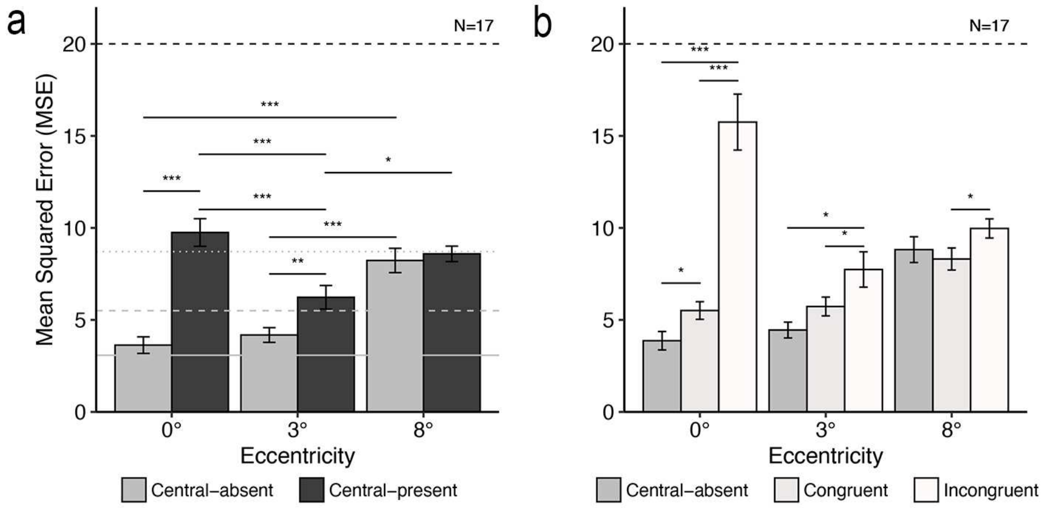
We compared participants' performance to identify the average emotion of the face set in central-present and central-absent conditions (Figure 1a). A 2×3 repeated-measures ANOVA with two factors was conducted: Central face (present vs. absent) and Eccentricity (0°, 3°, 8°). There was a significant main effect of Central face (F (1,16) = 35.50, $p < 0.001$, partial η^2 = 0.69) as well as Eccentricity (F (2,32) = 17.20, $p < 0.001$, partial η^2 = 0.52). There was also a significant interaction between Central face and Eccentricity (F (2,32) = 23.49, $p < 0.001$, partial η^2 = 0.60). First, we compared the effect of the central face in different visual fields. The difference between central-absent and central-present conditions was significant at 0° (absent (3.63 +/- 0.45) < present (9.75 +/- 0.75): $p < 0.001$) and 3° (absent (4.18 +/- 0.40) < present (6.23 +/- 0.64): $p < 0.01$), but not at 8° (absent (8.23 +/- 0.66) vs. present (8.59 +/- 0.42): p = 0.47). Second, we compared the effect of eccentricity in central-absent and central-present conditions. In the central-present condition, the performance was significantly better at 3° than 0° and 8° (3° < 0°: $p < 0.001$; 3° < 8°: $p < 0.05$), and there was no significant difference between 0° and 8° (p = 0.54). However, In the central-absent condition, the difference between 0° and 3° was not significant (p = 0.46), and participants' performance was better at 0° and 3° compared to 8° (0° < 8°: $p < 0.001$; 3° < 8°: $p < 0.001$).

To further investigate how the central face biased the performance, we divided trials in the central-present condition into two subsets: (1) the congruent condition where the central face had the same emotion as flankers, and (2) the incongruent condition where the central face had different emotion as that of flankers, and then we compared MSE in congruent, incongruent, and central-absent conditions (Figure 2b). A repeated-measures ANOVA with MSE as the dependent variable showed the main effects of Congruency (F(2, 32) = 40.60, p

< 0.001, partial $\eta^2 = 0.72$), Eccentricity ($F(2, 32) = 10.64$, $p < 0.001$, partial $\eta^2 = 0.40$), as well as an interaction between Congruency and Eccentricity ($F(4, 64) = 28.31$, $p < 0.001$, partial $\eta^2 = 0.64$). The ensemble performance was strongly impaired in the incongruent compared to the congruent and central-absent condition at 0° (congruent (5.51 +/- 0.48) < incongruent (15.75 +/- 1.52): $p < 0.001$; central-absent (3.87 +/- 0.50) < incongruent: $p < 0.001$; central-absent < congruent: $p < 0.05$). The performance at 3° eccentricity was similar to 0°, performance was significantly impaired in the incongruent condition compared to congruent and central-absent conditions (congruent (5.73 +/- 0.51) < incongruent (7.74 +/- 0.96): $p < 0.05$; central-absent (4.45 +/- 0.43) < incongruent: $p < 0.05$; congruent vs. central-absent: $p = 0.06$). At 8°, MSE was overall high and the performance was slightly better in congruent than incongruent conditions. There was no significant difference between congruent and central-absent conditions and between incongruent and central-absent conditions (congruent (8.31 +/- 0.60) < incongruent (9.97 +/- 0.52): $p < 0.05$; congruent vs. central-absent (8.82 +/- 0.70): $p = 0.75$; incongruent vs. central-absent: $p = 0.13$).

Figure 1

MSEs for emotion recognition of face ensembles at varied visual fields



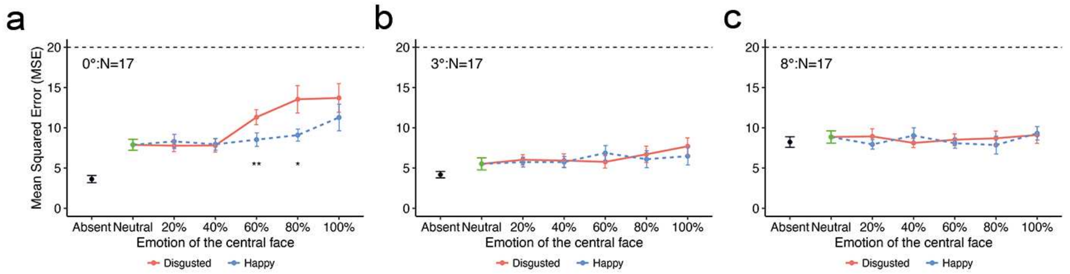
(a) MSE separated for face sets with (Central-present condition) and without (Central-absent condition) a central face. The gray horizontal lines represent MSEs in single-face conditions at 0° (solid line), 3° (dashed line), and 8° (dotted line). **(b)** MSE separated for face sets with congruent and incongruent central faces and without a central face (central-absent condition). The black dashed line represents the chance level. Asterisks indicate significance with alpha levels of 0.05 (*), 0.01 (**), and 0.001 (***). Error bars represent ± 1 *SEM*.

To assess whether the ensemble performance was also modulated by the emotional valence and intensity of the central face, we compared MSEs as a function of the emotional valence and intensity of the central face (Figure 2). The 2 × 5 repeated-measures ANOVA with two factors (Valence (happy vs. disgusted) × Intensity of the central face (20%, 40%, 60%, 80%, 100%)) was conducted at different eccentricities. At 0°, there was a significant main effect of Valence (F (1,16) = 4.66, $p < 0.05$, partial $\eta^2 = 0.23$) and Intensity (F (4,64) = 10.21, $p < 0.001$, partial $\eta^2 = 0.39$). There was also a significant Valence × Intensity interaction effect (F (4,64) = 3.45, $p < 0.05$, partial $\eta^2 = 0.18$). Specifically, the performance was less accurate with increased intensity of the foveal face (20%: 8.05 +/- 0.71; 40%: 7.88 +/- 0.73; 60%: 9.93 +/-

0.76; 80%: 11.31 +/- 0.88; 100%: 12.50 +/- 1.48). The performance difference between happy and disgusted foveal face conditions was significant in 60% (happy (8.53 +/- 0.84) < disgusted (11.32 +/- 0.93): $p < 0.01$) and 80% (happy (9.09 +/- 0.75) < disgusted (13.54 +/- 1.70): $p < 0.05$) intensity conditions but not others (20% happy (8.31 +/- 0.88) vs. 20% disgusted (7.78 +/- 0.73): $p = 0.50$; 40% happy (7.95 +/- 0.74) vs. 40% disgusted (7.81 +/- 0.83): $p = 0.81$; 100% happy (11.28 +/- 1.65) vs. 100% disgusted (13.71 +/- 1.78): $p = 0.18$). At 3° eccentricity, there was no significant main effect of Valence (F (1,16) = 0.60, $p = 0.45$, partial $\eta^2 = 0.04$), Intensity (F (4,64) = 1.05, $p = 0.39$, partial $\eta^2 = 0.06$), as well as interaction effect of Valence × Intensity (F (4,64) = 1.04, $p = 0.39$, partial $\eta^2 = 0.06$). At 8° eccentricity, neither a significant main effect of Valence (F (1,16) = 0.26, $p = 0.62$, partial $\eta^2 = 0.02$), Intensity (F (4,64) = 0.51, $p = 0.73$, partial $\eta^2 = 0.03$), nor the interaction effect of Valence × Intensity (F (4,64) = 0.53, $p = 0.71$, partial $\eta^2 = 0.03$) was found.

Figure 2

MSEs separated by the valence and intensity of the central face



MSE in the disgusted and happy central face conditions across the 5 emotional intensities of the central face (from 20% to 100%). “Absent” represents the condition where the central face was absent. “Neutral” represents the condition where the central face was neutral. The dashed line represents the chance level. Asterisks indicate significance with alpha levels of 0.05 (*), 0.01 (**). Error bars represent ± 1 *SEM*.

Acknowledgments

I would like to express my sincere gratitude to all those who have supported me throughout my journey towards completing this dissertation.

First and foremost, I would like to thank my supervisor, Prof. Bilge Sayim, for his guidance, support, and patience. His insightful comments and feedback have been invaluable in shaping my research. He has truly set an example for me of how to be a good researcher and a good supervisor. I am also grateful for the trust and freedom he gave me to pursue my research interests and ideas.

Meanwhile, I would like to express my appreciation to Prof. Amit Goldenberg for his supervision during my stay at Harvard University. Thank you for all the great support and encouragement you gave to me (*p.s.* you made excellent waffles).

I am also indebted to my family and friends, who have provided me with unwavering support. I know deep down that my parents are so proud of me, even though they have never said it to my face (that's typical of Chinese parents). Instead, whenever I get into trouble, my father always reminds me of how lucky I am. From him, I have learned to think more about what I have now and cherish it. My mother is a strong, independent, and selfless person whose life goal is to provide her three children with the opportunity to chase their dreams and enjoy their lives. I am so grateful for their dedication. As a PhD student who is self-funded, survival is one of the challenges I have had to face. For that, I would like to show my appreciation to Bilge, 钱兰英 (my master's supervisor), 赖丹凤 (my master's co-supervisor), and my friends 黄桂华，蔡璇青, Miko (陈奕雯), 刘敏，宋云萍，李淼，宋宗宝，王昕昕，张晋星，何杏彤，张润润，刘淋淋，崔静，罗娟，for their emotional and financial support. All the data in this dissertation was collected by my dear friend and collaborator Yunping Song （宋云萍）, without her, I would not have been able to finish my dissertation at this time. In addition, I would like to show my appreciation to my colleagues at SCALab. Miao, I would not have had the chance to know Bilge without you posing the PhD position. I also want to express my deepest gratitude for the incredible support you have given me. I am constantly amazed by your kindness, generosity, and selflessness. Jane, I am so lucky to have you as my best friend. Thank you for all the support you have given me, and for accepting my flaws and always loving me. Ângela, you are my angel. You cannot imagine how much strength you have given me. You are the only one who hugs me and tells me ‘I love you’ every time you see me. You make me feel that I am special, which I appreciate a lot. Sabrina, you are my role model, you never stopped fighting every time you got into trouble, and I know how much effort you have made. You deserve everything you have right now. Every time I faced trouble during my PhD, I used your story to encourage myself. Cristina, I really enjoyed my time working with you and the time we spent struggling with being poor. The journey of being poor is not lonely with your company. I sincerely hope you will find a PhD position with a salary soon, and I hope you can also enjoy your research journey. Amélie (Menut), thank you for all the special attention you have given me. In my mind, you are a cheerleader and I am a nerd, that’s why I feel so honored every time you talk to me and share everything with me. I can feel your big trust which I appreciate a lot. Amélie (Remy Neris) and Matjø, I always think that the three of us belong to the nerd group because we are shy, and we don’t socialize, even though Matjø has a different opinion about that. I really enjoyed our coffee break time, and thank you for your great support. Laura, many people from our lab feel warm and supported because of you, and I am one of them. I would also like to express my appreciation to Manu and Sabine, our lab managers. You are the backbone of our lab, and I cannot imagine what a mess our lab would be without the two of you. Angela (Bartolo) and Yannick, thank you for being my CST members and supporting me during the PhD. Angela, as you expected, we will not meet this year at the CST meeting, but I know you will be proud of me. I would like to show my appreciation to all my colleagues at SCALab (Chick & Chicks). It is an honor to join you and become a member of this team. I have really enjoyed my time here. Lastly, I also want to show my appreciation to my boyfriend Issam Lazaar. Thank you for your unwavering support and strength during a difficult time in my life. Your presence in my life has been like a rock, providing a firm foundation and a sense of stability that I could always rely on.

Bibliography

Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., … Schaller, M. (2006). They all look the same to me (unless they're angry). *Psychological Science*, *17*(10), 836–840. https://doi.org/10.1111/j.1467-9280.2006.01790.x

Allard, R., & Cavanagh, P. (2012). Different processing strategies underlie voluntary averaging in low and high noise. *Journal of vision*, *12*(11), 6-6. https://doi.org/10.1167/12.11.6

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, *74*(5), 810-815. https://doi.org/10.3758/s13414-012-0293-0

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., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*, 392–398. https://doi.org/10.1111/j.1467-9280.2008.02098.x

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, *106*(18), 7345-7350. https://doi.org/10.1073/pnas.0808981106

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

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Developmental Origins of the Other-Race Effect. *Current Directions in Psychological Science*, *22*(3), 173–178. https://doi.org/10.1177/0963721412474459

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological science*, *12*(2), 157-162. https://doi.org/10.1111/1467-9280.00327

Aston, S., Negen, J., Nardini, M., & Beierholm, U. (2022). Central tendency biases must be accounted for to consistently capture Bayesian cue combination in continuous response data. *Behavior Research Methods*, *54*(1), 508-521.

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

Baek, J., & Chong, S. C. (2020). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics*, *82*, 63-79. https://doi.org/10.3758/s13414-019-01827-z

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

Beck, D. M., & Lavie, N. (2005). Look Here but Ignore What You See: Effects of Distractors at Fixation. *Journal of Experimental Psychology: Human Perception and Performance, 31*(3), 592–607. https://psycnet.apa.org/doi/10.1037/0096-1523.31.3.592

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

Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of the effective stimulus in visual search. *Perception & Psychophysics*, *62*(3), 576-585. https://doi.org/10.3758/BF03212109

Bouma, H. (1970). Interaction Effects in Parafoveal Letter Recognition. *Nature, 226*(5241), 177-178. https://doi.org/10.1038/226177a0

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

Boari, G., & Nai Ruscone, M. (2015). A procedure simulating Likert scale item responses. *Electronic journal of applied statistical analysis*, *8*(3), 288-297. 10.1285/i20705948v8n3p288

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological science*, *22*(3), 384-392. https://doi.org/10.1177/0956797610397956

Bulakowski, P. F., Post, R. B., & Whitney, D. (2011). Reexamining the possible benefits of visual crowding: dissociating crowding from ensemble percepts. *Attention, Perception, & Psychophysics*, *73*, 1003-1009. https://doi.org/10.3758/s13414-010-0086-2

Campana, F., & Tallon-Baudry, C. (2013). Anchoring visual subjective experience in a neural model: the coarse vividness hypothesis. *Neuropsychologia*, *51*(6), 1050-1060. https://doi.org/10.1016/j.neuropsychologia.2013.02.021

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

Chambers, C. D., Allen, C. P., Maizey, L., & Williams, M. A. (2013). Is delayed foveal feedback critical for extra-foveal perception?. *Cortex*, *49*(1), 327-335. https://doi.org/10.1016/j.cortex.2012.03.007

Chan, J. C. (1991). Response-order effects in Likert-type scales. *Educational and Psychological Measurement*, *51*(3), 531-540.

Chen, Z., Zhuang, R., Wang, X., Ren, Y., & Abrams, R. A. (2021). Ensemble perception without attention depends upon attentional control settings. *Attention, Perception, & Psychophysics*, *83*(3), 1240-1250. https://doi.org/10.3758/s13414-020-02067-2

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, 153, 196–210. https://doi.org/10.1016/j.cognition.2016.04.018

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017). Rapid learning of visual ensembles. *Journal of Vision*, 17 (2): 21, 1–15. https://doi.org/10.1167/17.2.21

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

Chong, S. C., Joo, S. J., Emmanouil, T. A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics*, *70*(7), 1327-1334. https://doi.org/10.3758/PP.70.7.1327

Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement*, *56*(10), 15-23.

Contemori, G., Oletto, C. M., Cessa, R., Marini, E., Ronconi, L., Battaglini, L., & Bertamini, M. (2022). Investigating the role of the foveal cortex in peripheral object discrimination. *Scientific reports*, *12*(1), 1-13. https://doi.org/10.1038/s41598-022-23720-w

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta psychologica*, *138*(2), 289-301. https://doi.org/10.1016/j.actpsy.2011.08.002

Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition*, *20* (2), 211–231. https://doi.org/10.1080/13506285.2012.657261

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

Dandan, Y. R., Ji, L., Song, Y., & Sayim, B. (2023a). Foveal vision determines the perceived emotion of face ensembles. *Attention, Perception, & Psychophysics*, *85*(1), 209-221. https://doi.org/10.3758/s13414-022-02614-z

Dandan, Y. R., Song, Y., & Sayim, B. (2023b, in preparation). Foveal input bias in ensemble emotion perception.

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

Douven, I. (2018). A Bayesian perspective on Likert scales and central tendency. *Psychonomic bulletin & review*, *25*(3), 1203-1211. https://doi.org/10.3758/s13423-017-1344-2

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Elias, E., Dyer, M., & Sweeny, T. D. (2017). Ensemble perception of dynamic emotional groups. *Psychological Science*, *28*(2), 193-203.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*, *128*(2), 203. https://doi.org/10.1037/0033-2909.128.2.203

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & psychophysics*, *70*(6), 946-954. https://doi.org/10.3758/PP.70.6.946

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

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, *16*(1), 143-149. https://doi.org/10.3758/BF03203267

Fan, X., Wang, L., Shao, H., Kersten, D., & He, S. (2016). Temporally flexible feedback signal to foveal cortex for peripheral object recognition. *Proceedings of the National Academy of Sciences*, *113*(41), 11627-11632. https://doi.org/10.1073/pnas.1606137113

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, *106*(3), 1389-1398. http://dx.doi.org/10.1152/jn.00904.2010

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently?. *Cognition & emotion*, *14*(1), 61-92. https://doi.org/10.1080/026999300378996

Furtak, M., Mudrik, L., & Bola, M. (2022). The forest, the trees, or both? Hierarchy and interactions between gist and object processing during perception of real-world scenes. *Cognition*, *221*, 104983. https://doi.org/10.1016/j.cognition.2021.104983

Geringswald, F., Baumgartner, F., & Pollmann, S. (2012). Simulated loss of foveal vision eliminates visual search advantage in repeated displays. *Frontiers in human neuroscience*, *6*, 134. https://doi.org/10.3389/fnhum.2012.00134

Geringswald, F., & Pollmann, S. (2015). Central and peripheral vision loss differentially affects contextual cueing in visual search. *Journal of experimental psychology: learning, memory, and cognition*, *41*(5), 1485. https://psycnet.apa.org/doi/10.1037/xlm0000117

Givon, M. M. & Shapira, Z. Response to rating scales: A theoretical model and its application to the number of categories problem. J. Mark. Res. 21, 410–419 (1984).

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

Gong, M., & Smart, L. J. (2021). The anger superiority effect revisited: a visual crowding task. *Cognition and Emotion*, *35*(2), 214-224. https://doi.org/10.1080/02699931.2020.1818552

Gong, M., Li, X. (2022). Attentional demand induced by visual crowding modulates the anger superiority effect. *Attention, Perception & Psychophys*, *84*, 442–449.

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

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*,*15*(4):16, 1–11. https://doi.org/10.1167/15.4.16

Haberman, J., & Suresh, S. (2021). Ensemble size judgments account for size constancy. *Attention, Perception, & Psychophysics*, *83*(3), 925-933. https://doi.org/10.3758/s13414-020-02144-6

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

Hansmann-Roth, S., Chetverikov, A., & Kristjánsson, Á. (2019). Representing color and orientation ensembles: Can observers learn multiple feature distributions? *Journal of vision*, *19*(9), 2-2. https://doi.org/10.1167/19.9.2

Harvey, C. (2016). Binary Choice vs Ratings Scales: A behavioural science perspective. *International Journal of Market Research*, *58*(5), 647-648.

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

Hilbert, S., Kuechenhoff, H., Sarubin, N., Toyo Nakagawa, T. & Buehner, M. (2016). The influence of the response format in a personality questionnaire: An analysis of a dichotomous, a likert-type, and a visual analogue scale. Test. Psychom. Methodol. Appl. Psychol. 23, 25.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791-804. https://doi.org/10.1016/S0896-6273(02)01091-7

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

Iakovlev, A. U., & Utochkin, I. S. (2023). Ensemble averaging: What can we learn from skewed feature distributions?. *Journal of Vision*, *23*(1), 5-5. https://doi.org/10.1167/jov.23.1.5

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception & Psychophysics*, *75* (2), 278–286, https://doi.org/10.3758/s13414-012-0399-4.

Im, H. Y., Park, W. J., and Chong, S. C. (2015). Ensemble statistics as units of selection. *Journal of Cognitive Psychology*, 27(1): 114-127.

Im, H. Y., Chong, S. C., Sun, J., Steiner, T. G., Albohn, D. N., Adams, R. B., & Kveraga, K. (2017). Cross-cultural and hemispheric laterality effects on the ensemble coding of emotion in facial crows. *Culture and Brain*, *5*(2), 125–152. https://doi.org/10.1016/j.physbeh.2017.03.040

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

Ji, L., Chen, W., Loeys, T., & Pourtois, G. (2018). Ensemble representation for multiple facial expressions: Evidence for a capacity limited perceptual process. *Journal of vision*, *18*(3), 1-19. https://doi.org/10.1167/18.3.17

Ji, L., & Pourtois, G. (2018). Capacity limitations to extract the mean emotion from multiple facial expressions depend on emotion variance. *Vision Research*, *145*, 39-48. https://doi.org/10.1016/j.visres.2018.03.007

Ji, L., Pourtois, G., & Sweeny, T. D. (2020). Averaging multiple facial expressions through subsampling. *Visual Cognition*, *28*(1), 41-58.

Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision*, *9*(9), 1-12.

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

Kacin, M., Gauthier, I., & Cha, O. (2021). Ensemble coding of average length and average orientation are correlated. *Vision Research*, *187*, 94-101. https://doi.org/10.1016/j.visres.2021.04.010

Kang, H. (2021). Sample size determination and power analysis using the G* Power software. *Journal of educational evaluation for health professions*, *18*.

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

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The Other-Race Effect Develops During Infancy: Evidence of Perceptual Narrowing. *Psychological Science*, 18(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Khvostov, V. A., & Utochkin, I. S. (2019). Independent and parallel visual processing of ensemble statistics: Evidence from dual tasks. *Journal of Vision*, *19*(9):3, 1-18.

Khayat, N. & Hochstein, S. (2019). Relating categorization to set summary statistics perception. *Attention, Perception & Psychophysics*, *75*(2), 278-286

Khayat, N., Fusi, S., & Hochstein, S. (2021). Perceiving ensemble statistics of novel image sets. Attention, *Perception, & Psychophysics*, *83*(3), 1312-1328.

Kouider, S., Berthet, V., & Faivre, N. (2011). Preference is biased by crowded facial expressions. *Psychological Science*, *22*(2), 184-189. https://doi.org/10.1177/0956797610396226

Kroell, L. M., & Rolfs, M. (2022). Foveal vision anticipates defining features of eye movement targets. *Elife*, *11*, e78106. https://doi.org/10.7554/eLife.78106

Kunz, T. (2015). Rating scales in web surveys. A test of new drag-and-drop rating procedures.

Kuru, O., & Pasek, J. (2016). Improving social media measurement in surveys: Avoiding acquiescence bias in Facebook research. *Computers in Human Behavior*, *57*, 82-92.

Kwon, D., & Chong, S. C. (2023). The relationship between ensemble representations of facial information. *Vision Research*, *203*, 108156. https://doi.org/10.1016/j.visres.2022.108156

Lepsien, J., & Nobre, A. C. (2007). Attentional modulation of object representations in working memory. *Cerebral cortex*, *17*(9), 2072-2083. https://doi.org/10.1093/cercor/bhl116

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision research*, *48*(5), 635-654. https://doi.org/10.1016/j.visres.2007.12.009

L-Miao, L., Dandan, Y.R., Reynvoet, B., & Sayim, B. (in preparation). Redundancy masking of faces reveals a substantial failure to detect socially relevant information.

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

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

Lin, Z., Gong, M., & Li, X. (2022). On the relation between crowding and ensemble perception: Examining the role of attention. *PsyCh Journal*. https://doi.org/10.1002/pchj.559

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

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta psychologica*, *142*(2), 245-250. https://doi.org/10.1016/j.actpsy.2012.11.002

Maule, J., & Franklin, A. (2016). Accurate rapid averaging of multihue ensembles is due to a limited capacity subsampling mechanism. *JOSA A*, *33*(3), A22–A29.

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

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*, *9*(3), 353-383. https://doi.org/10.1016/0010-0285(77)90012-3

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

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, *128*, 56–63. http://dx.doi.org/10.1016/j.cognition.2013.03.006

Ni, L., & Stocker, A. A. (2023). Efficient sensory encoding predicts robust averaging. *Cognition*, *232*, 105334.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*(3), 466–478. https://doi.org/10.1037/0096-3445.130.3.466

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research*, *79*, 8-16. https://doi.org/10.1016/j.visres.2012.12.004

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

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

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, *4*(12):12, 1136–1169. https://doi.org/10.1167/4.12.12

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11*(10), 1129-1135. https://doi.org/10.1038/nn.2187

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

Pimentel, J. L. (2010). A note on the usage of Likert Scaling for research data analysis. *USM R&D Journal*, *18*(2), 109-112.

Qasem, M. A. N., & Gul, S. B. A. (2014). Effect of items direction (positive or negative) on the factorial construction and criterion related validity in Likert scale.

Ray, J. J. (1990). Acquiescence and problems with forced-choice scales. *The Journal of Social Psychology*, *130*(3), 397-399.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science*, *206*(4417), 468-469.

Rivera-Garrido, N., Ramos-Sosa, M. P., Accerenzi, M., & Brañas-Garza, P. (2022). Continuous and binary sets of responses differ in the field. *Scientific reports*, 12(1), 1-10.

Rosenbaum, D., de Gardelle, V., & Usher, M. (2021). Ensemble perception: Extracting the average of perceptual versus numerical stimuli. *Attention, Perception, & Psychophysics*, *83*(3), 956-969.

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

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin*, *115*(1), 102.

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

Sayim, B., & Wagemans, J. (2017). Appearance changes and error characteristics in crowding revealed by drawings. *Journal of vision, 17*(11), 8-8. https://doi.org/10.1167/17.11.8

Sayim, B., & Taylor, H. (2019). Letters lost: capturing appearance in crowded peripheral vision reveals a new kind of masking. *Psychological Science, 30*(7), 1082-1086. https://doi.org/10.1177/0956797619847166

Sekimoto, T., & Motoyoshi, I. (2022). Ensemble perception without phenomenal awareness of elements. *Scientific Reports*, *12*(1), 1-8. https://doi.org/10.1038/s41598-022-15850-y

Shasteen, J. R., Sasson, N. J., Pinkham, A. E., & Yovel, G. (2014). Eye tracking the face in the crowd task: Why are angry faces found more quickly? *PLoS One*, *9*(4), e93914. https://doi.org/10.1371/journal.pone.0093914

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

Theeuwes, J., & Burger, R. (1998). Attentional control during visual search: the effect of irrelevant singletons. *Journal of Experimental Psychology: Human perception and performance*, *24*(5), 1342. https://psycnet.apa.org/doi/10.1037/0096-1523.24.5.1342

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta psychologica*, *135*(2), 77-99.

Thomas, P. M., Jackson, M. C., & Raymond, J. E. (2014). A threatening face in the crowd: effects of emotional singletons on visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 253. https://psycnet.apa.org/doi/10.1037/a0033970

Thornton, I. M., Srismith, D., Oxner, M., & Hayward, W. G. (2019). Other-race faces are given more weight than own-race faces when assessing the composition of crowds. *Vision Research*, *157*, 159-168.

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

Tong, K., Dubé, C., & Sekuler, R. (2019). What makes a prototype a prototype? Averaging visual features in a sequence. *Attention, Perception, & Psychophysics*, *81*(6), 1962-1978.

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, *32* (7), 1349–1357. https://doi.org/10.1016/0042-6989(92)90227-A

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

Tripathy, S. P., Cavanagh, P., & Bedell, H. E. (2014). Large crowding zones in peripheral vision for briefly presented stimuli. *Journal of Vision*, *14*(6), 11-11. https://doi.org/10.1167/14.6.11

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

Vandormael, H., Herce Castañón, S., Balaguer, J., Li, V., & Summerfield, C. (2017). Robust sampling of decision information during perceptual choice. *Proceedings of the National Academy of Sciences*, *114*(10), 2771-2776. https://doi.org/10.1073/pnas.1613950114

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500-503. https://doi.org/10.1038/nature04171

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436. https://psycnet.apa.org/doi/10.1037/0096-1523.32.6.1436

Wang, B., & Theeuwes, J. (2018). Statistical regularities modulate attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 13.

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

Weldon, K. B., Woolgar, A., Rich, A. N., & Williams, M. A. (2020). Late disruption of central visual field disrupts peripheral perception of form and color. *PloS one*, *15*(1), e0219725.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*, *15*(4), 160-168. https://doi.org/10.1016/j.tics.2011.02.005

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

Williams, M. A., Baker, C. I., Op de Beeck, H. P., Mok Shim, W., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature neuroscience*, *11*(12), 1439-1445.

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

Yeh, S. L., He, S., & Cavanagh, P. (2012). Semantic priming from crowded words. *Psychological science*, *23*(6), 608-616. https://doi.org/10.1177/0956797611434746

Yildirim, F. Z., Coates, D. R., & Sayim, B. (2020). Redundancy masking: The loss of repeated items in crowded peripheral vision. *Journal of vision,* *20*(4), 14-14. https://doi.org/10.1167/jov.20.4.14

Yildirim, F. Z., Coates, D. R., & Sayim, B. (2021). Hidden by bias: how standard psychophysical procedures conceal crucial aspects of peripheral visual appearance. *Scientific Reports,* *11*(1), 4095. https://doi.org/10.1038/s41598-021-83325-7

Yildirim, F. Z., Coates, D. R., & Sayim, B. (2022). Atypical visual field asymmetries in redundancy masking. *Journal of vision*. https://doi.org/10.1167/jov.20.4.14

Ying, H. (2022). Attention modulates the ensemble coding of facial expressions. *Perception*, *51*(4), 276-285.

Yörük, H., & Boduroglu, A. (2020). Feature-specificity in visual statistical summary processing. *Attention, Perception, & Psychophysics*, *82*, 852-864. https://doi.org/10.3758/s13414-019-01942-x

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