

Effects of Subjective Similarity and Culture on Ensemble Perception of Faces

Shenli Peng^{1#}, Chang Hong Liu^{2#}, and Ping Hu³

¹ College of Education, Hunan Agricultural University, Changsha, China

² Department of Psychology, Bournemouth University, Poole, UK

³ Department of Psychology, Renmin University of China

These two authors contribute equally to this article.

Author Note

This research was supported by the National Social Science Foundation of China (Major Program) (19ZDA021). The authors declare no conflict of interest to disclose. Correspondence should be addressed to Ping Hu, Department of Psychology, Renmin University of China, Zhongguancun Street, Haidian District, Beijing 100872, P.R. China. Telephone number: 010-82509716, E-mail: huping@ruc.edu.cn.

Abstract

It is well established that ensemble coding is regulated by physical similarity and variance in a set of stimuli. For example, observers are more accurate at judging the mean size of objects in a set if the overall size variance in the set is small. However, sometimes similarity among set members can be purely subjective. For example, faces from another race tend to look more similar than faces from one's own race. Very little is known about whether such subjective similarity also regulates ensemble coding in the same manner as objective similarity. To investigate this question, we had British and Chinese participants view sets of four faces that were of either own-race or other-race, own-gender or other-gender. After viewing each set the task was to judge whether a test face was presented in the set. Our results showed that, as demonstrated in prior research, participants often mistook a morphed set average to be a member of the set. Critically, this tendency to average a face set was not stronger for other-race faces. Hence contrary to objective similarity, subjectively perceived similarity in the other-race faces does not facilitate ensemble coding. The results in our British group also replicated de Fockert and Gautrey's (2013) own-gender effect, where observers showed more averaging for own-gender faces. However, our Chinese displayed same level of averaging for both genders. This suggests a cultural difference in ensemble coding, where the own-gender bias may be overridden by a stronger tendency to employ ensemble coding in Chinese participants.

Keywords: ensemble coding, subjective similarity, in-group vs. out-group faces, cultural difference.

People are good at rapidly extracting statistical properties such as a set average and variance from an array of stimulus ensemble. This ability, known as ensemble coding, has been demonstrated with a variety of stimuli ranging from simple geometric shapes (Ariely, 2001), colors (Maule & Franklin, 2015) to complex stimuli such as faces (Haberman & Whitney, 2007; de Fockert & Wolfenstein, 2009) in both adults and children (Rhodes et al., 2018), and people influenced by different cultures (Li et al., 2016).

Research suggests that ensemble coding of a set average is modulated by the physical similarity among its set members (Corbett, Wurnitsch, Schwartz, & Whitney, 2012; Michael, De Gardelle, & Summerfield, 2014; Maule & Franklin, 2015; Sweeny, Haroz, & Whitney, 2013; Tong, Ji, Chen, & Fu, 2015). For example, smaller color variation among set members makes it easier for observers to judge the mean color of the set (Maule & Franklin, 2015). Estimations of the mean size improve if there is less heterogeneity of sizes in a set (Marchant, Simons & de Fockert, 2013; Im & Halberda, 2013; Utochkin & Tiurina, 2014). Judgments of the mean shape or color is improved if the feature variance presented in a preceding set matched that of the test set (Michael et al., 2014). The adaptation to the mean is weaker when ensembles contain more variance (Corbett et al., 2012). These results suggest that physical similarity is a crucial factor in ensemble coding. This also applies to more complex stimuli. For example, extracting a mean face identity is easier when the faces in a set is made more similar to each other by morphing a percentage of each face with the average of these faces (Cabeza, Bruce, Kato, & Oda, 1999). The mean facial expression in a set is more accurately perceived when the set variance is low (Goldenberg, Sweeny, Shpigel, & Gross, 2020).

Whilst most studies focused on objective inter-item similarity by manipulating

the physical dimensions of the stimuli such as the size gradient of circles in a set, de Fockert and Gautrey (2013) studied the effect of subjective or perceived similarity. Based on the reports that people are better at recognizing faces of their own-gender than other-gender (e.g., Cross et al., 1971), de Fockert and Gautrey predicted a stronger tendency to extract a mean identity from a set of other-gender faces, because of the outgroup homogeneity effect (Park & Judd, 1990; Byatt & Rhodes, 2004), that is, other-gender faces should be subjectively perceived as being more similar to each other than same-gender faces, and hence are less discriminable. To test this prediction, they adapted the paradigm from de Fockert and Wolfenstein (2009), in which participants judged whether a test face was “present” or “absent” in the prior display consisted of a set of four faces. The test face could be a morphed average of the preceding set of four faces (matching average) or another set of four faces (nonmatching average), or a member of the preceding set (matching member) or another face set (nonmatching member). The main result, which has subsequently been replicated by a number of studies (de Fockert & Gautrey, 2013; Peng, Kuang, & Hu, 2019), was that the matching average was often misidentified as a matching member. The study provided the evidence that observers automatically extract an identity average from a set of faces. The new manipulation in the de Fockert and Gautrey (2013) study was that their participants were shown both own-gender or other-gender face sets. To their surprise, their results revealed an opposite pattern to the prediction, i.e., averaging was stronger for faces of own gender. The authors reasoned that the lack of support for the original hypothesis could be due to gender being insufficiently perceived as a factor for in-group and out-group classification. They speculated that categories such as ethnicity or race might produce a stronger sense of in-group and out-group distinction.

The present study aimed to explore such possibility by studying the effect of other-race homogeneity on the extraction of average identity. Unlike other-gender faces whose effect on perceived similarity was based on an untested hypothesis, there is already evidence that other-race faces are perceived as being more similar than the own-race faces (e.g., Byatt & Rhodes, 2004; Papesh & Goldinger, 2010; Laurence, Zhou, & Mondloch, 2016). For example, when a group of Caucasian participants were asked to rate the similarity between paired Chinese or Caucasian faces, they rated the Chinese as more similar to each other than Caucasian faces (Byatt & Rhodes, 2004). Thus, compared to other-gender faces, other-race faces would provide a more suitable ground for testing de Fockert and Gautrey's initial hypothesis that subjectively perceived similarity should produce more averaging for out-group faces.

Like de Fockert and Gautrey, the key purpose of our study was to investigate the effect of perceived rather than physical inter-item similarity on ensemble perception. A main difference in our study design was that our British and Chinese participants were shown both own-race and other-race faces, whereas their study used only own-race faces. We used female and male faces to replicate their finding on own-gender averaging effect. In order to reduce the complexity of research design (please refer to Design below), we only tested female participants. If subjectively perceived homogeneity could enhance mean extraction just like the effects of objective homogeneity, the out-group faces should create stronger averaging for other-race faces. Namely, Chinese participants would produce more "present" responses for a set average of Caucasian faces, while the British participants would produce more "present" responses for a set average of Chinese faces. However, the unexpected finding by de Fockert and Gautrey also led to an alternative prediction, which was more averaging responses for own-race faces just as what was found for own-gender

faces.

By comparing effects of in-group vs. out-group of face race and gender in two ethnic groups, we necessarily had to consider the potential influence from the cultural dimension. It is well known that East Asians often have a stronger tendency to process visual information holistically relative to Westerners (e.g., Masuda & Nisbett, 2001). There is also evidence that ensemble perception requires holistic processing (Chong & Treisman, 2003). Together, these may explain some preliminary findings that East Asians tend to perceive a set average more often than Westerners do (Im et al., 2017; Peng et al., 2020). Such cross-cultural difference could modulate the ways ensemble coding of in-group and out-group set members in this study. Specifically, while both cultural groups would be influenced by subjective similarity in out-group faces (perceived out-group homogeneity), Chinese participants would show a stronger tendency to endorse a set average than British participants.

Finally, in de Fockert and Gautrey (2013) and other similar studies, the out-group homogeneity effect in their stimuli was often assumed without being verified. To extend this literature, we also verified whether perceived similarity in our face stimuli actually varied across in-group and out-group face categories in a second experiment.

Experiment 1

Method

Participants. Thirty-four Chinese female college students (19.8 ± 1.1 years old) from Renmin University of China and 34 Caucasian female college students (19.5 ± 1.3 years old) from Bournemouth University of the United Kingdom took part in this study. All reported normal or corrected-to-normal vision. This research was approved

by the local Institutional Review Board. Written consent was obtained from each participant.

Materials. A set of face images with a neutral emotional expression was adopted from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). The database contained 52 Asian Males, 57 Asian Females, 93 Western males, 90 Western females. From these we selected 56 East Asian and 56 Caucasian faces, which amounted to a total of 112 faces. Within each face race, there were equal numbers of male and female faces. The faces were selected with the constraints that they all came from the same age group and ethnicity, which excluded a few elderly and South Asian (such as Indian and Pakistani) faces. Apart from this, the process of the stimulus selection was random.

The external features of each face (e.g. hair, clothing) were occluded via an oval mask through Adobe Photoshop CS6. Each face race of the same gender consisted of seven sets of four faces, which amounted to a total of 28 sets (2 face race \times 2 face gender \times 7 sets). For each set, we created a morph from the four faces using Abrosoft FantaMorph 5. There was therefore a total of 28 morphed faces. Each facial image (170 \times 220 px), presented against a black background, subtended 4.8° \times 7.4° of visual angle at a viewing distance of around 60cm. Both experiments in this study were programmed in E-prime 2.0.

Procedure. Figure 1 illustrates the procedure in each trial, which was identical to de Fockert and Wolfenstein (2009). Each trial began with a set of four faces presented centrally for 2000ms after a 500ms fixation. This was then followed by a test face. Participants were asked to judge whether the test face was presented in the preceding set. They pressed the “f” key with left index finger for “present” response, and the “j” key with right index finger for “absent” response. No feedback or time limits were

given. The test face was from one of the four conditions: a morph of the four faces in the preceding set (matching average), a morph of a different set of faces with matched gender and race (nonmatching average), a member of the preceding set (matching

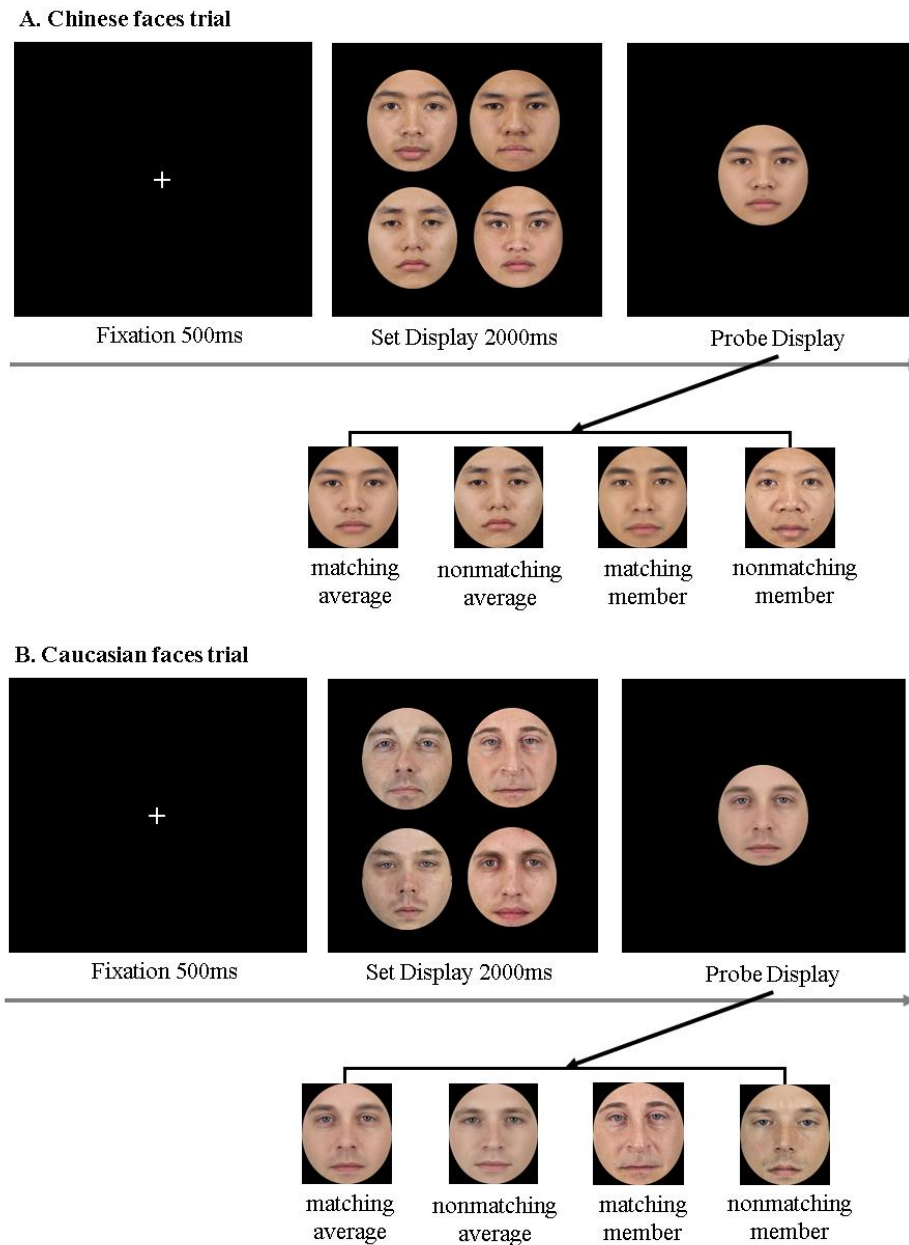


Figure 1. Illustration of the trial procedure in this study. A and B show two example trials for Chinese and Caucasian face stimuli respectively.

member), or another set of the same race and gender (nonmatching member). The test face type and face gender were counterbalanced within block, while race was counterbalanced between blocks. Thus, the experiment consisted of 2 blocks with each face race, and each block included 56 trials (2 face gender \times 7 sets \times 4 test face types). The two experimental blocks were run after a practice block.

Design. We employed a mixed design. Observer Race (Chinese vs. Caucasian) was a between-participants variable, whereas Face Gender (female, male), Face Race (Chinese vs. Caucasian), Face Type (set average vs. set member), Match (matching vs. nonmatching) were within-participant variables. Following de Fockert and Wolfenstein (2009), and de Fockert and Gautrey (2013), we used the proportion of “present” response as the dependent variable.

Results and discussion

The proportion and standard deviation of present responses for each condition are given in Table 1. A five-way repeated-measures analysis of variance (ANOVA) was conducted on the data. As the ANOVA results in Table 2 indicate, the significant main effects of Face Race and Match were qualified by a number of significant interactions involving these variables. The significant two-way interactions between Face Gender and Face Type, and Face Type and Match were implicated in the significant four-way interactions between these variables and Face Race. Therefore, our subsequent analyses were focused on identifying the source of the two significant four-way and three-way interactions.

Table 1.

The Mean Proportion of “Present” Responses as a Function of Observer Race, Face Gender, Face Race, Face Type, and Match. Values in Parentheses Represent Standard Deviations.

Observer Race	Face Race	Face Gender	Match			
			Matching		Nonmatching	
			Set average	Set member	Set average	Set member
Chinese	Chinese	Female	.62 (.19)	.59 (.14)	.33 (.15)	.21 (.14)
		Male	.67 (.20)	.67 (.16)	.30 (.17)	.24 (.12)
	Caucasian	Female	.62 (.20)	.60 (.16)	.29 (.16)	.18 (.15)
		Male	.62 (.20)	.60 (.15)	.25 (.16)	.23 (.14)
Caucasian	Chinese	Female	.62 (.18)	.55 (.09)	.31 (.14)	.22 (.12)
		Male	.63 (.15)	.57 (.18)	.26 (.14)	.22 (.11)
	Caucasian	Female	.57 (.18)	.58 (.10)	.26 (.17)	.21 (.12)
		Male	.53 (.19)	.61 (.13)	.19 (.10)	.27 (.11)

We first conducted simple effects analyses for the four-way interaction between Face Gender, Face Type, Face Race, and Match by separate ANOVAs for the two matching conditions. For the matching test face condition, the only significant effect was the two-way interaction between Face Gender and Face Type, $F(1, 67) = 14.95$, $p < .001$, $\eta_p^2 = .18$. Figure 2A illustrates this interaction, which shows that female set members ($M = .578$) were recognized less accurately than male set members ($M = .629$), $F(1, 67) = 12.38$, $p = .001$, $\eta_p^2 = .156$, while the proportions of endorsement for female and male matching averages (M 's = .607 and .589, respectively) were comparable, $F(1, 67) = 3.23$, $p = .130$, $\eta_p^2 = .03$.

Table 2.
ANOVA of the Proportion of “Present” Responses

Source	<i>df</i>	<i>F</i>	<i>p</i>	η^2_p
Face Race (FR)	1	6.50	.013*	.09
Face Gender (FG)	1	0.74	.393	.01
Face Type (FT)	1	2.23	.141	.03
Match (M)	1	1132.07	<.001***	.95
Observer Race (OR)	1	1.35	.250	.02
FG × OR	1	3.53	.065	.05
FR × OR	1	0.38	.538	.01
FT × OR	1	2.76	.101	.04
M × OR	1	2.59	.112	.04
FG × FR	1	0.60	.441	.01
FG × FT	1	37.25	<.001***	.36
FR × FT	1	3.76	.057	.05
FG × M	1	3.33	.073	.05
FR × M	1	0.67	.415	.01
FT × M	1	20.93	<.001***	.24
FG × FR × OR	1	2.14	.149	.03
FG × FT × OR	1	5.12	.027*	.07
FR × FT × OR	1	1.40	.241	.02
FG × FR × FT	1	0.40	.534	.01
FG × M × OR	1	0.68	.412	.01
FR × M × OR	1	0.17	.684	<.01
FG × FR × M	1	3.53	.065	.05
FT × M × OR	1	0.10	.750	<.01
FG × FT × M	1	0.41	.522	.01
FR × FT × M	1	3.33	.073	.05
FT × FR × FT × OR	1	0.01	.929	<.01
FG × FR × M × OR	1	1.86	.177	.03
FG × FT × M × OR	1	2.96	.090	.04
FR × FT × M × OR	1	0.06	.811	<.01
FG × FR × FT × M	1	7.22	.009**	.10
FG × FR × FT × M × OR	1	2.57	.113	.04
Error	66			

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

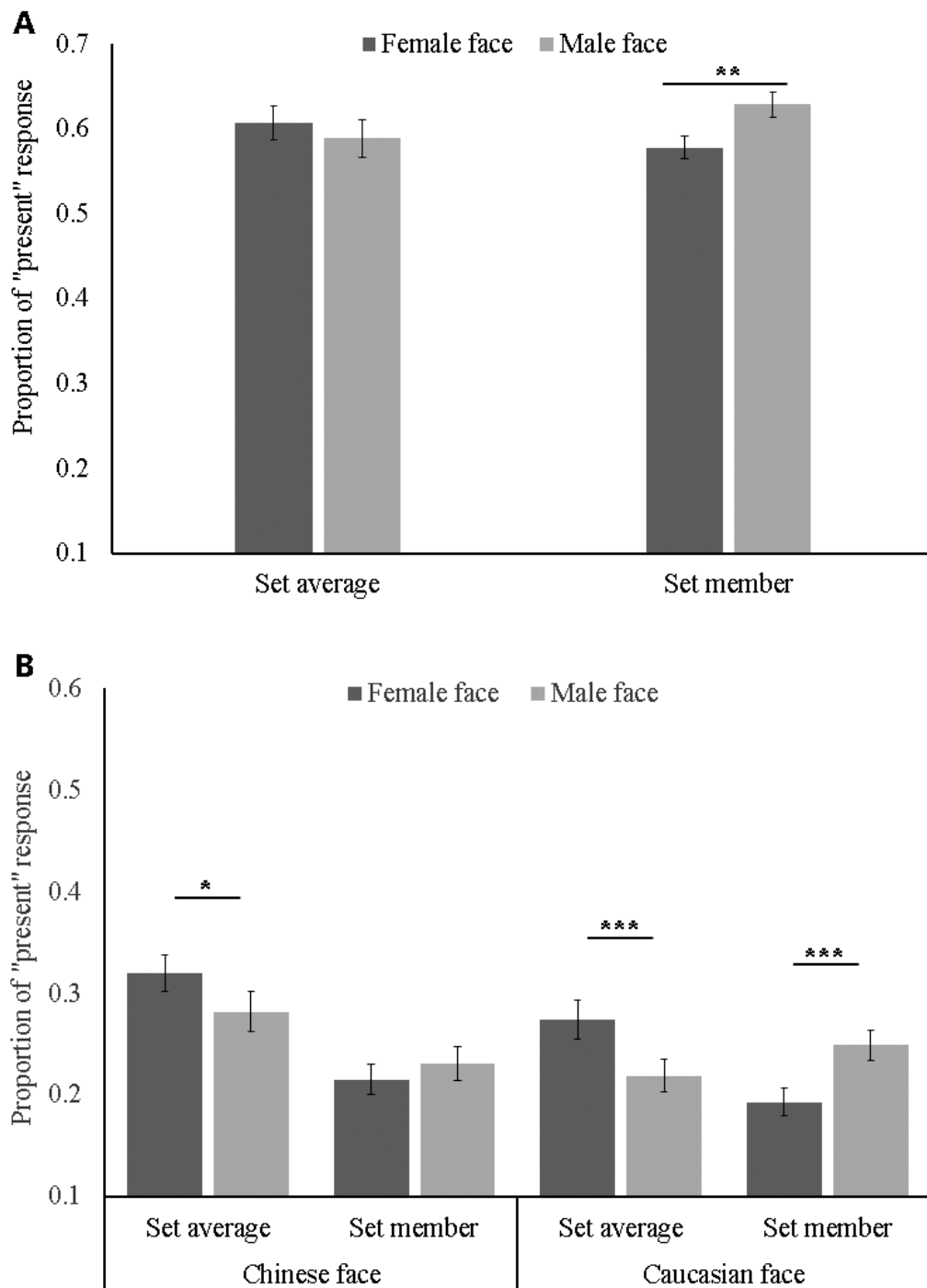


Figure 2. A. Proportion of “present” response as a function of Face Gender and Face Type in the matching condition. B. Proportion of “present” response as a function of Face Gender, Face Type and Face Race in the nonmatching condition. Error bars show standard errors. * $p < .05$, ** $p < .01$, *** $p < .001$.

For the nonmatching test face condition, ANOVA revealed main effects of Face Race, $F(1, 66) = 8.04, p = .006, \eta_p^2 = .109$, and Face Type, $F(1, 66) = 11.31, p = .001, \eta_p^2 = .146$, and significant two-way interactions between Face Gender and Face Type, $F(1, 66) = 24.64, p < .001, \eta_p^2 = .272$, and between Face Race and Face Type, $F(1, 66) = 7.45, p = .008, \eta_p^2 = .101$. However, these main effects and two-way interactions were qualified by a significant three-way interaction, $F(1, 66) = 5.15, p = .026, \eta_p^2 = .072$. Further analysis of this interaction (illustrated in Figure 2B) showed that for set members, Caucasian male faces ($M = .249$) were more frequently endorsed than Caucasian female faces ($M = .193$), $p < .001$, 95% CI [-.085, -.026]. This was not the case for Chinese female and male faces ($M = .215$ and $.231$, respectively), $p = .277$, 95% CI [-.044, .013]. For set averages, participants produced more “present” responses for Chinese female averages ($M = .320$) than Chinese male averages ($M = .282$), $p = .030$, 95% CI [.004, .073]. This was also the case for Caucasian female and male faces (M 's = $.274$ and $.219$ respectively), $p < .001$, 95% CI [.026, .085].

It is worth noting that nonmatching set average has a higher response rate than the nonmatching set member (see Figure 2B). This demonstrates a bias for choosing an average even when it contained no information about the target faces. Such bias is consistent with the De Fockert and Wolfenstein (2009)'s suggestion that averaging is a “default mode”, which could explain why even a nonmatching average were preferred relative to a nonmatching member. It is important to note, however, that the proportion of matching set average was much higher than the proportion of nonmatching set average.

We next conducted simple effects analysis for the three-way interaction among Observer Race, Face Gender and Face Type in Table 2. Figure 3 illustrates this interaction. While Caucasian observers endorsed the female averages (own gender) more often ($M = .439$) relative to male averages (other gender, $M = .381$), $p < .001$, 95% CI [.036, .080], Chinese observers showed no difference between endorsing female averages (own gender, $M = .465$) and male averages (other gender, $M = .458$), $p = .543$. Additionally, higher endorsement rates for male member faces relative to female member faces were found in both Chinese, $p = .004$, 95% CI [-.068, -.013], and Caucasian participants, $p = .001$, 95% CI [-.072, -.018].

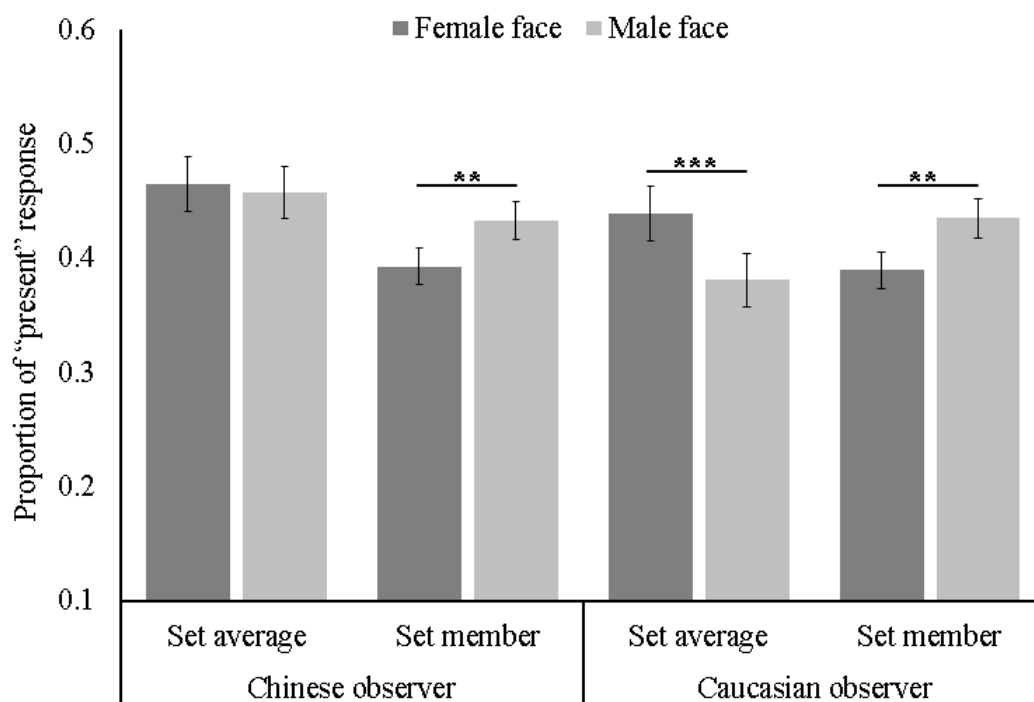


Figure 3. Proportion of "present" response as a function of Observer Race, Face Gender and Face Type. Error bars show standard errors. ** $p < .01$, *** $p < .001$.

The results in Figure 3 showed more averaging for own-gender faces in the Caucasian observers. This replicates the own-gender effect in de Fockert and

Gautrey's (2013). However, the Chinese observers displayed a discrepant pattern, showing a same level of averaging for both face genders.

On the whole, participants in this experiment tended to mistake a set average as a set member, although this effect was not stronger for other-race faces. This suggests that subjectively perceived similarity in the other-race faces does not facilitate ensemble coding. However, despite the existing evidence that other-race faces tend to be perceived as being more similar than own-race faces (e.g. Byatt & Rhodes, 2004), it was unknown whether it was also true of the face stimuli used in the current study. It was also not known whether other-gender faces were perceived as being more similar to each other. Given the importance of the out-group homogeneity assumption for this study, we next conducted a rating experiment to measure whether observers actually judge the out-group faces as being more similar than the in-group faces.

Experiment 2

Method

Participants. This was a group of Chinese female college students, who did not participant in the first experiment ($N = 60$, 19.7 ± 1.3 years old).

Materials. These were identical as Experiment 1. The 112 faces were shown in 56 pairs taking from the 28 face sets from Experiment 1. The face race and gender were matched in each pair. The pairings and the left-right positions of each pair were randomized for each participant.

Procedure. In each trial, a pair of faces was presented on the center of the screen. Participants were instructed to rate the similarity between the two faces on a 7-point scale ($1 = \textit{very dissimilar}$, $7 = \textit{very similar}$). The response was given on a numeric keypad. The pair of stimuli stayed on the screen until the participant

responded, upon which a new pair was presented on the screen until participant rated all 56 face-pairs.

Design. This was a within-participant design. The two independent variables were Face Race (Chinese vs. Caucasian) and Face Gender (female vs. male). The dependent variable was rating scores of similarity.

Results and discussion

Results are shown in Figure 4. The inter-rater reliability of the results was high (Cronbach's $\alpha = .86$). A two-way repeated measures ANOVA showed a significant main effect of Face Race, $F(1, 59) = 406.95, p < .001, \eta_p^2 = .87$, where Caucasian face-pairs ($M = 4.18$) were rated more similar than that of Chinese face-pairs ($M = 2.64$), 95% CI = [1.39, 1.70]. There was also a significant main effect of Face Gender, $F(1, 59) = 86.60, p < .001, \eta_p^2 = .60$, where Male face-pairs ($M = 3.63$) were rated more similar than Female face-pairs ($M = 3.19$), 95% CI = [.35, .54]. The two-way interaction was not significant, $F(1, 59) = .79, p = .377, \eta_p^2 = .01$.

The results showed that the Caucasian faces looked more similar to each other than Chinese faces to Chinese participants in this study. This effect was consistent with previous findings (Byatt & Rhodes, 2004; Laurence et al., 2016). Additionally, our results also showed that male faces looked more similar to each other than female faces to our female participants. This is perhaps the first direct evidence that the out-group homogeneity effect can also be observed in out-group gender. In conclusion, this experiment confirmed the key assumption of our study that participants perceived greater similarity among out-group faces in our stimuli.

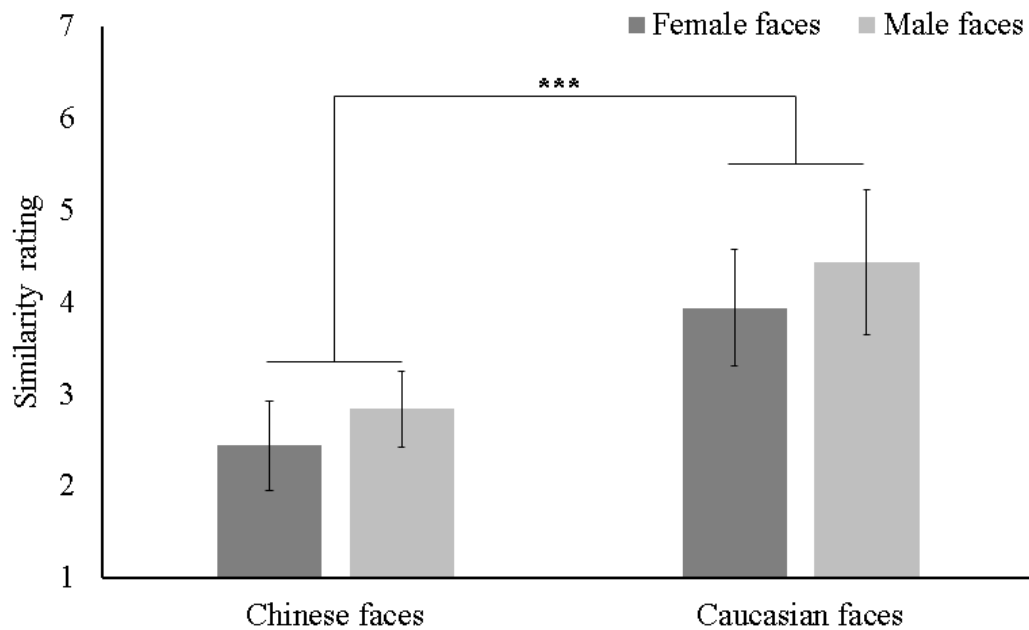


Figure 4. Similarity rating data for Chinese/Caucasian and female/male faces. Error bars show standard errors. *** $p < .001$.

General Discussion

The main purpose of this study was to investigate whether ensemble coding was driven by subjectively perceived similarity as it is by objective similarity demonstrated in the literature. According to one prediction (out-group homogeneity effect), participants should be more likely to endorse a set average for other-race faces than for own-race faces, because other-race faces are perceived more similar to each other than own-race faces (e.g., Byatt & Rhodes, 2004). Our results showed no evidence for this prediction. Both Chinese and British participants produced the same level of “present” responses for the own-race and other-race set average, even though our subsequent rating experiment was able to confirm that participants judged out-group faces to be more similar. This result also ruled out the prediction that the own-gender bias for set average found in de Fockert and Gautrey (2013) would generalize

to an own-race bias for set average.

The rest of our findings were largely consistent with the existing literature. Observers in both Chinese and British groups mistook a matching average as a member of the faces in the preceding set equally often as they correctly identified a set member. This replicates the findings of several studies (Rhodes et al., 2018; Neumann, Schweinberger, & Burton, 2013; Neumann, De Bonis, Rhodes, & Palermo, 2015; Neumann, Ng, Rhodes, & Palermo, 2017), although de Fockert and Wolfenstein (2009) found a higher endorsement rate for matching averages than for matching members.

Our results also partially replicated the own-gender effect for female participants in de Fockert and Gautrey (2013), who found higher endorsement rates for own-gender set averages relative to other-gender averages. Because they did not find an interaction between Match and other factors, this own-gender effect was found in both matching and nonmatching conditions. In our results, however, Match interacted with Face Gender, Face Race, and Face Type in a four-way interaction. When simple-effect analyses were conducted on the matching conditions, it became evident that the own-face gender effect was absent for matching averages, although there was a similar effect of poorer identification for own-gender matching members like that reported in the Fockert and Gautrey study (see Figure 2A). However, the own-gender bias was found in our nonmatching conditions (see Figure 2B), where a nonmatching female average was more likely to be reported as a set member of the preceding set relative to a nonmatching male average. Observer Race did not interact with these variables, meaning that both Chinese and British participants produced the same own-gender effect in the nonmatching conditions, although the effect was relatively stronger for the Caucasian faces. The partial discrepancy between our results and de Fockert and

Gautrey's could be due to the larger sample size in our study as a result of employing two ethnic groups.

Another additional finding of our study was the cross-cultural difference between the ways in which Face Gender and Face Type affected Chinese and British participants. The results in Figure 3 showed that the own-gender bias only existed for the British participants, although both groups showed higher endorsement for set members of other-gender than own-gender faces. Because Match did not interact with these factors here, it means that the effects were similar for matching and nonmatching conditions. It is worth noting that Match also did not interact with Observer Gender and Face Type in de Fockert and Gautrey's (2013) study. Thus the pattern of results in our British group was consistent with that in their study, which also employed a British sample. Both studies consistently showed a stronger endorsement for own-gender set average, and a weaker endorsement for own-gender exemplar. The results of our Chinese group, however, showed no own-gender bias for set average, although a similar effect was found for own-gender exemplar.

These results suggest that although the two cultural groups processed set *members* of the in-group versus out-group gender in a similar fashion, they showed a difference in the way the set *average* of in-group and out-group gender was treated. This is the first time that a cross-cultural difference is demonstrated for the own-gender effect. East Asians are known have a stronger tendency to process visual information holistically relative to Westerners (e.g., Masuda & Nisbett, 2001; Miyamoto, 2013). Given the evidence that ensemble perception requires holistic processing (Chong & Treisman, 2003; Im et al., 2017; Peng et al., 2020), perhaps the stronger holistic processing mode of our Chinese participants have propelled them to extract set averages from both face genders, and this could have in turn reduced their

own-gender effect. Our finding is consistent with Im et al. (2017), who showed a stronger tendency in Koreans than in Americans to extract a mean emotion from a set of faces.

An important implication in this study is that subjectively perceived similarity in a set of stimuli may be fundamentally different from objectively defined similarity in their effects on ensemble perception. While objective similarity facilitates ensemble representation, subjective similarity has little effect on forming such representation. De Fockert and Gautrey (2013) were the first to demonstrate this. They showed that although the subjective out-group homogeneity effect would predict increased endorsement of set average for the other-gender faces, their results showed the exactly the opposite, where higher endorsement was found for the own-gender faces. They pointed out that their finding was an exception to the own-gender advantage in the face recognition literature, where female observers are better at recognizing faces of their own gender (Lewin & Herlitz, 2002; Rehnman & Herlitz, 2006). There are also reports that male observers are better at recognizing faces of their own-gender as well (Wright & Sladden, 2003). The typical interpretation of this effect is that out-group faces look more similar to each other relatively to in-group faces. Based on the findings in our study, this explanation clearly does not apply to averaging in ensemble perception. When discussing the difference between the effects of perceived similarity on face recognition and ensemble averaging, de Fockert and Gautrey (2013) speculated that gender might not be sufficiently perceived in-group and out-group identifications. They suggested that perhaps ethnic group could produce a stronger sense of in-group and out-group distinction to create an effect in visual averaging. Our study has shown that this is not the case. Instead, this study has provided new evidence that unlike objective similarity, subjectively perceived similarity is

independent of averaging in ensemble perception. Thus our results further confirm that the influence of subjective similarity on ensemble perception is different from that on identity recognition.

In conclusion, evidence from the present study have consolidated the view that perceived similarity due to the out-group homogeneity effect does not facilitate extraction of average identity from a group of unique faces. This may be surprising given the well-established observation that objective similarity can strongly affect the process. Our results further demonstrate that subjectively perceived similarity plays different roles in identity recognition and ensemble coding. Another important finding from the current study is that the predisposition to extract a mean identity from a set of faces can be modulated culturally. The own-gender effect on ensemble coding found in Western participants may diminish in East Asian participants due to their varying degrees of processing visual information holistically.

Data Availability

The data of the reported experiments are available via the Open Science Framework (OSF): <https://osf.io/xzebh/>. The experiment reported here was not preregistered.

References

- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157-162.
- Byatt, G., & Rhodes, G. (2004). Identification of own-race and other-race faces: Implications for the representation of race in face space. *Psychonomic Bulletin & Review, 11*(4), 735-741.
- Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition: Extension and limits. *Memory & Cognition, 27*(1), 139-151.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393-404.
- Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition, 20*(2), 211-231.
- Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics, 10*(6), 393-396.
- de Fockert, J. W., & Gautrey, B. (2013). Greater visual averaging of face identity for own-gender faces. *Psychonomic Bulletin & Review, 20*(3), 468-473.
- de Fockert, J. W., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology, 62*(9), 1716-1722.
- 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), 445-460.
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*(17), R751-R753.
- Im, H. Y. , Chong, S. C. , Sun, J. , Steiner, T. G. , Albohn, D. N. , & Adams, R. B. , et

- al. (2017). Cross-cultural and hemispheric laterality effects on the ensemble coding of emotion in facial crowds. *Culture and Brain*, 5(2), 125-152.
- 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.
- Laurence, S., Zhou, X., & Mondloch, C. J. (2016). The flip side of the other-race coin: They all look different to me. *British Journal of Psychology*, 107(2), 374–388.
- Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition: women's faces make a difference. *Brain and Cognition*, 50(1), 121–128.
- 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.
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47(4), 1122-1135.
- 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.
- Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, 81(5), 922-934.
- Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*, 15(4):6, 1-18.
- 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.
- Miyamoto, Y. (2013). Culture and analytic versus holistic cognition: Toward multilevel analyses of cultural influences. *Advances in Experimental Social Psychology*, *47*, 131-188.
- Neumann, M., De Bonis, F., Rhodes, G., & Palermo, R. (2015). The role of similarity in coding ensemble identity of face groups. *Journal of Vision*, *15*(12), 705-705.
- Neumann, M. F., Ng, R., Rhodes, G., & Palermo, R. (2017). Ensemble coding of face identity is not independent of the coding of individual identity. *The Quarterly Journal of Experimental Psychology*, 1-27.
- Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, *128*(1), 56-63.
- Papesh, M. H., & Goldinger, S. D. (2010). A multidimensional scaling analysis of own- and cross-race face spaces. *Cognition*, *116*(2), 283–288.
- Park, B., & Judd, C. M. (1990). Measures and models of perceived group variability. *Journal of Personality and Social Psychology*, *59*(2), 173-191.
- Peng, S., Kuang, B., & Hu, P. (2019). Memory of ensemble representation was independent of attention. *Frontiers in Psychology*, *10*, 228.
- Peng, S., Liu, C. H., Yang, X., Li, H., Chen, W., & Hu, P. (2020). Culture variation in the average identity extraction: The role of global vs. local processing orientation. *Visual Cognition*, *28*(3), 180–191.
- Rehman, J., & Herlitz, A. (2006). Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias. *Memory*, *14*(3), 289-296.
- Rhodes, G., Neumann, M., Ewing, L., Bank, S., Read, A., Engfors, L. M., ... & Palermo, R. (2018). Ensemble coding of faces occurs in children and develops

dissociably from coding of individual faces. *Developmental Science*, 21(2), e12540.

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

Tong, K., Ji, L., Chen, W., & Fu, X. (2015). Unstable mean context causes sensitivity loss and biased estimation of variability. *Journal of Vision*, 15(4):15, 1-12.

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica*, 146, 7-18.

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica*, 114(1), 101-114.