Gender Biases in the Accuracy of Facial Judgments: Facial Attractiveness and Perceived Socioeconomic Status

Yue Qi[1] and Jia Ying[2]*

[1] *Department of Psychology, Renmin University of China, Beijing, China,* [2] *Graduate School of Education, University of Pennsylvania, Philadelphia, PA, United States*

Many studies demonstrate that people form their first impression of a stranger based on facial appearance, and these impressions influence their subsequent decisions and behaviors. However, much less research has examined the factors that moderate the accuracy of first impressions based on a photo of face. The present study included three experiments to explore gender differences in the accuracy of impressions based on faces. The results showed that people judge facial attractiveness more accurately for female faces than for male faces while giving more accurate wealth judgments for male faces than for female faces. Interestingly, although we did not find a significant correlation between confidence ratings and the accuracy of wealth rating, we recognized a significant moderate correlation between confidence ratings and the accuracy of attractiveness ratings when female participants rated male faces. To our knowledge, the present study is the first to reveal gender biases in the accuracy of impression judgments based on facial appearance. These findings imply a significant influence of traditional gender roles on accurate facial judgments.

Keywords: facial attractiveness, gender bias, impression, socioeconomic status, facial judgment

INTRODUCTION

When interacting with a stranger, people may form their first impression based on limited available information (e.g., facial appearance), and these judgments can subsequently and indirectly influence social decision making (Qi et al., 2018, 2021; Li et al., 2021). Many studies demonstrate that facial attractiveness has an impact on various social decisions, such as friendship and mating choices (Thornhill and Gangestad, 1999), monetary decision-making (Pandey and Zayas, 2021), and hiring (Luxen and Van De Vijver, 2006). People judge facial attractiveness based on common aesthetic or affective attributes of different genders (Rhodes, 2006). According to the owner hypothesis, facial attractiveness is a stable characteristic of those with faces (Chen et al., 1997; Little and Perrett, 2002). Researchers have explored some facial features that affect facial attractiveness judgments, such as averageness (Komori et al., 2009), symmetry (Baudouin and Tiberghien, 2004), sexual dimorphism (Perrett et al., 1998; Russell, 2003), and vitality (Zheng and Zhou, 2021). The observer hypothesis argues the importance of the beholder on facial attractiveness perception and emphasizes the characteristics of the observer, such as the observer's age (Little et al., 2010),

personality (Welling et al., 2009), and sociocultural factors (Little et al., 2002). Attractiveness can be a sign of health, and highly attractive faces can induce positive and pleasant emotional experiences (Rhodes, 2006; Zhang et al., 2021), which are rewarding to individuals (Aharon et al., 2001). Previous studies have found that the reward value of facial attractiveness can be influenced by the gender of the perceiver (Cloutier et al., 2008; Levy et al., 2008).

Facial gender is another impact factor in attractiveness processing. Mitrovic et al. (2018) found that both males and females looked longer at female faces, especially attractive female faces. This is in accordance with the "female beauty captures the mind" hypothesis (Maner et al., 2003). From an evolutionary perspective, males and females will emphasize the different characteristics of potential mates. Males pay more attention to characteristics related to reproductive potential, such as physical attractiveness, while females pay more attention to characteristics that signal resource acquisition, such as status and dominance (Buunk et al., 2002). Furthermore, attractive female faces capture more behavioral attention (Slater et al., 1998; Maner et al., 2003), bring more rewards (Collins and Missing, 2003; Colwell, 2007; Wang et al., 2015), and cause more brain activation in neural mechanisms (Zhang et al., 2012; Ru et al., 2017). In other words, attractive female faces capture more attention and are more visible than attractive male faces.

To date, many studies have discussed the accuracy of facial judgments (e.g., Todorov et al., 2015; Walker and Vetter, 2016). However, much less research has examined the factors that moderate the accuracy of first impressions when viewing a photo of a face (Alaei and Rule, 2016). Previous studies have investigated self-other agreement on traits in face-to-face contexts and found that extroversion and openness can be accurately judged (e.g., Borkenau et al., 2009; Back et al., 2010; Moritz and Roberts, 2018). However, neuroticism is the least accurately judged trait in online contexts (Gosling et al., 2007; Back et al., 2010). These findings can be explained by the trait visibility effect (Funder and Dobroth, 1987); that is, the more relevant and frequent the behaviors the trait elicits, the more accurate the judgments that are made will be (Watson et al., 2000), because perceivers can acquire more valid cues to judge the trait.

Moreover, previous studies have revealed that the longer people know each other, the more accurately they rate each other's traits. Compared with strangers who observed behaviors for only a few minutes, acquaintances predicted behavior better and were more consistent with their reports of observed behavior (Biesanz et al., 2007). For example, married couples have higher self-other agreements on most affectivities and personalities than friendship dyads or dating couples do (Watson et al., 2000). Increased acquaintanceship is accompanied by more trait-relevant messages; thus, perceivers can make more accurate judgments of the target (Funder, 1995; Funder et al., 1995). Considering that facial attractiveness carries additional significance for women (Luxen and Van De Vijver, 2006), people may be more accustomed to evaluating women's attractiveness in everyday life. Thus,

we expected gender bias in the accuracy of attractiveness judgment from faces.

The present study was designed to explore the influence of gender factors on the accuracy of people's judgments of facial attractiveness. In the review of Tsankova and Tair (2021), the accuracy of first impressions refers to "the correspondence between the subjective perception of the interaction partners and some more objective criterion (e.g., Funder and West, 1993; Brauer and Proyer, 2020)." Thus, previous research commonly uses the term "accuracy" to illustrate the agreement between actual cooperative behaviors or self-reported personality and perceived personality from others (e.g., Funder, 1995; Borkenau et al., 2009; Chan et al., 2010; Todorov et al., 2015; Alaei and Rule, 2016). However, with regard to attributes without objective criteria, such as self-reported stress (Little et al., 2011), researchers employ self-other agreement or distinctive self-other agreement (Human et al., 2013) to measure facial judgment accuracy. According to the above definitions of accuracy, in the current research, the accuracy of facial attractiveness was calculated by self-other agreement.

According to the trait visibility effect (Funder and Dobroth, 1987; Watson et al., 2000) and the acquaintanceship effect (Funder, 1995), we hypothesized that people tend to give more accurate ratings of the facial attractiveness of female faces than of male faces. These gender differences arise because across many cultures, a woman's attractiveness is important (Li et al., 2002; Shackelford et al., 2005), whereas a man's status and resources are more crucial than his attractiveness (Buss and Schmitt, 1993; Sprecher et al., 1994). Therefore, in Studies 1 and 2, we explored the gender differences in judgment accuracy and metaperception accuracy on facial attractiveness. Study 3 was designed to investigate the cognitive mechanism of these gender biases. Participants were asked to give their ratings on the perceived wealth of the person depicted in each photo in Study 3. The accuracy of perceived economic status in Study 3 was calculated by the correspondence between the participants' subjective perception of faces and the actual wealth ranking group.
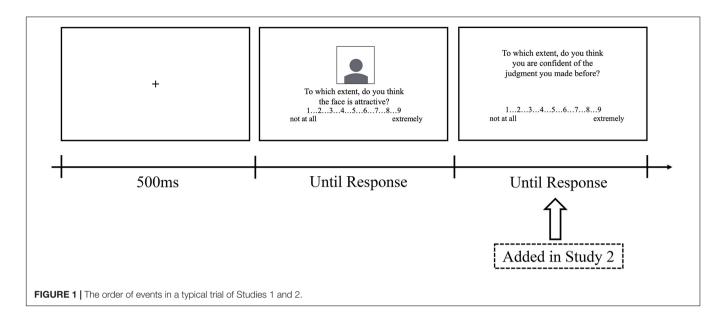
STUDY 1

This experiment was designed to explore the influence of the perceiver's gender on accuracy in judging facial attractiveness. Considering that facial attractiveness carries additional significance for women (Luxen and Van De Vijver, 2006), people may be more accustomed to evaluating women's attractiveness in everyday life, which motivates women to pay more attention than men to their attractiveness. Thus, we hypothesized that (1) people tend to give more accurate ratings of facial attractiveness for female faces than for male faces and (2) women tend to assess people's facial attractiveness more accurately than men.

Methods
Participants
A total of 90 students participated in Study 1 for payment, including 41 males (M_{age} = 24.32, SD_{age} = 3.66) and 49 females (M_{age} = 24.86, SD_{age} = 4.68). This study was approved by the

FIGURE 1 | The order of events in a typical trial of Studies 1 and 2.

internal review board of the Department of Psychology, Renmin University of China. Each participant signed an informed consent form and received monetary compensation for his or her time.

Stimuli

Another 119 undergraduate students (58 male and 61 female, age range 18–25 years) were recruited to have frontal shoulder-up pictures taken with a digital camera in front of a white background for use as stimuli. Before the photos were taken, they removed all accessories except glasses (if they could not finish the task without them). We asked the students to maintain a natural (neutral emotion) expression. After the photos were taken, they were asked to rate their attractiveness in the eyes of others of the same gender and different genders using a 9-point scale ranging from 1 (not attractive at all) to 9 (extremely attractive). We did not find a difference between their self-ratings of attractiveness in the eyes of others of the same gender and others of a different gender $t(236) = -0.42$, $p = 0.678$. All photographed participants consented to the use of their photos for our research purposes, including showing their pictures to other participants. All the faces were adjusted to the same size, 295×295 pixels.
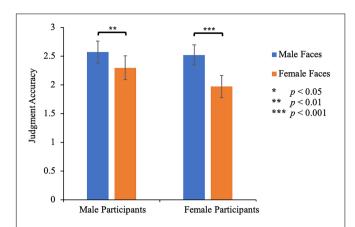
Apparatus and Procedure

The experiment was conducted on a computer with E-prime 2.0. Participants were told to use their gut feeling to rate the attractiveness of each face photo. In a typical trial of the study, a fixation point was presented for 500 ms, and then a face photo was shown with a 9-point rating scale below it. Participants were asked to give their rating on the attractiveness of each face photo from 1 (not attractive at all) to 9 (extremely attractive). The experiment contained two blocks with a total of 238 trials, and each face photo was presented once in a block. The order of the photos was random. Participants started with a practice block of 8 trials to familiarize them with the task. Between the two blocks, the participants were allowed to

take a break and started the next block on their own if they thought they were ready for it (see **Figure 1**). Considering that one participant rated the same target twice during the study, we used the mean rating for each face as the other-rating of the face.

In the review of Devos et al. (2013), self-other agreement is a relative phenomenon that refers to a degree of discrepancy between self-ratings and other-ratings. In previous research, self-other agreement was operationalized as the absolute difference of self and other ratings (Atwater and Yammarino, 1992, 1997; Bernieri et al., 1994; Lee and Carpenter, 2018; Kim et al., 2019) in addition to correlation (Borkenau and Liebler, 1993; Rogers et al., 2018). In the present research, we standardized the ratings of attractiveness for each face by subtracting other-ratings from self-ratings. Specifically, when a participant rated the face of someone of the same gender, the other-rating of attractiveness for this face was subtracted from the self-rating in the eyes of others with the same gender and vice versa. Thus, the standardized rating scores, which refer to rating accuracy, ranged from −8 to 8, with higher scores indicating that participants rated the target's attractiveness lower than the target's self-ratings. The absolute value indicates the difference between self-rating scores and other-rating scores. To be more specific, a higher absolute value indicates that participants rated the target's attractiveness lower than the target's self-ratings. Positive or negative values suggest whether participants underestimated or overestimated facial attractiveness compared to self-ratings. All subsequent analyses were based on the standardized data.

Results

Mean standardized ratings were submitted to a 2 (participant's gender: male, female) × 2 (facial gender: male, female) mixed-design measures ANOVA with face gender as a within-subject factor (**Figure 2**). The main effect of face gender was significant, $F(1, 88) = 62.07$, $p < 0.001$, $\eta_p^2 = 0.414$, indicating that female

FIGURE 2 | Rating accuracy of facial attractiveness in Study 1. Standardized rating scores as a subtraction of self-rating and other-rating facial attractiveness in Study 1. Error bars represent 1 S.E. of the means.

faces (2.14 ± 0.14) were judged more accurately than male faces (2.55 ± 0.13). The main effect of participant's gender was not significant, $F(1, 88) = 0.50$, $p = 0.483$, $\eta_p^2 = 0.006$, indicating that the influence of the participant's gender on judgment accuracy was relatively limited. More importantly, there was a significant interaction between the participant's gender and facial gender, $F(1, 88) = 6.87$, $p = 0.010$, $\eta_p^2 = 0.072$. Male participants rated female faces (2.30 ± 0.21) more accurately than male faces (2.57 ± 0.19), $p = 0.001$. Female participants also rated female faces (1.97 ± 0.19) more accurately than male faces (2.52 ± 0.17), $p < 0.001$. These results indicated that compared to the self-ratings of the targets, the participants' ratings tended to underestimate the targets' attractiveness. More importantly, all participants showed higher rating accuracy in judging the attractiveness of female faces. In addition, for male faces, male participants and female participants had similar rating accuracy (2.57 ± 0.19 vs 2.30 ± 0.021, $p = 0.845$), while for female faces, male participants and female participants also had similar rating accuracy (2.52 ± 0.17 vs 1.97 ± 0.19, $p = 0.255$).

STUDY 2

Study 1 found that participants tend to rate female faces' attractiveness more accurately than male faces, which confirms hypothesis 1. Thus, Study 2 was designed to retest these findings. Moreover, to explore whether participants were aware of their rating accuracy, we added a confidence-rating task to the experiment and calculated the correlation between confidence rating and rating accuracy.

Methods
Participants
A total of 50 students at Nankai University participated in Study 2, including 25 males ($M_{age} = 20.88$, $SD_{age} = 1.92$) and 25 females ($M_{age} = 20.68$, $SD_{age} = 1.22$). Each participant signed an informed consent form and received monetary compensation for his or her time.

Stimuli, Apparatus, and Procedure
The stimuli and procedure were identical to those in Study 1, with one exception. Participants were asked to rate their confidence after giving their attractiveness rating in each trial to explore whether raters were aware of the accuracy of their judgments. Specifically, after rating the attractiveness of the target face photo, a 9-point scale was shown with the question, "To what extent do you think you are confident of the judgment you made before?" Participants could respond from 1 (not confident at all) to 9 (extremely confident).

Results
Attractiveness Rating
Mean standardized ratings were submitted to a 2 (participant's gender: male, female) × 2 (facial gender: male, female) mixed-design measures ANOVA with facial gender as a within-subject factor (**Figure 3**). The main effect of facial gender was significant, $F(1, 48) = 57.30$, $p < 0.001$, $\eta_p^2 = 0.544$, indicating that female faces (1.61 ± 0.15) were judged more accurately than male faces (2.21 ± 0.14). The main effect of participant's gender was not significant, $F(1, 48) = 1.67$, $p = 0.203$, $\eta_p^2 = 0.034$. There was a significant interaction between the participant's gender and facial gender, $F(1, 48) = 24.50$, $p < 0.001$, $\eta_p^2 = 0.338$. Specifically, male participants rated female faces (1.62 ± 0.22) almost as accurately as male faces (1.83 ± 0.20), $p = 0.070$. Female participants rated female faces (1.59 ± 0.22) more accurately than male faces (2.60 ± 0.20), $p < 0.001$. These results confirmed that compared to the self-ratings of the targets, the participants' ratings tended to underestimate the facial attractiveness of male and female target faces. Similar to Study 1, all participants showed higher accuracy in judging the attractiveness of female faces. In addition, for male faces, male participants gave more accurate ratings (1.83 ± 0.20) than female participants (2.60 ± 0.20, $p = 0.010$), while for female faces, male participants and female participants had similar rating accuracy (1.62 ± 0.22 vs 1.60 ± 0.22, $p = 0.307$).
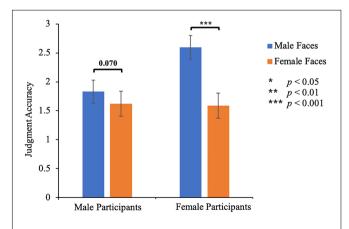

FIGURE 3 | Rating accuracy of facial attractiveness in Study 2. Error bars represent 1 S.E. of the means.

Confidence Rating

We conducted a 2 (participant's gender) $\times$ 2 (facial gender) mixed-design measures ANOVA with facial gender as a within-subject factor. The main effect of facial gender was not significant, $F(1, 48) = 2.41$, $p = 0.191$, $\eta_p^2 = 0.048$. The main effect of participant's gender was also not significant, $F(1, 48) = 1.10$, $p = 0.307$, $\eta_p^2 = 0.022$. The interaction between the participant's gender and facial gender was not significant, $F(1, 48) = 3.79$, $p = 0.058$, $\eta_p^2 = 0.073$. These results suggested that male and female participants were not aware of their own accuracy in judging the attractiveness of others based on the facial appearance of different genders. Moreover, there was a significant but weak positive correlation between rating confidence and judgment accuracy ($r = 0.27$, $p = 0.006$). Specifically, the correlation between confidence ratings and rating accuracy was significant when female participants rated male faces ($r = 0.52$, $p = 0.008$). In other words, when male participants rated male ($r = 0.23$, $p = 0.275$) and female faces ($r = 0.21$, $p = 0.323$) or female participants rated female faces ($r = 0.28$, $p = 0.173$), they lacked a clear awareness of their rating accuracy.

STUDY 3

Studies 1 and 2 found that participants rated the attractiveness of female faces more accurately than that of male faces. There are two possible explanations: (1) the self-ratings of males are higher than those of females, or (2) the more people care about females' facial attractiveness, the more accurately they are able to rate this trait. Thus, we compared the self-ratings of male and female faces and found that there was no significant difference between males' and females' self-ratings in the eyes of others of the same gender [$t(117) = 0.23$, $p = 0.985$] or a different gender [$t(117) = -0.74$, $p = 0.246$]. We conducted Study 3 to test the second explanation. Elder (1969) and Udry (1977) found that physical attractiveness has a "market value" for females, while males tend to be evaluated in terms of their status (Wade, 1988). This indicates that people pay more attention to the attractiveness of females, while they care more about the social status or wealth of males. Therefore, this experiment was designed to test whether participants could rate the wealth of males more accurately than that of females based on facial appearance.

Methods

Participants

Another 50 students at Nankai University participated in Study 3, including 25 males ($M_{age} = 21.48$, $SD_{age} = 2.63$) and 25 females ($M_{age} = 20.12$, $SD_{age} = 1.90$). Each participant signed an informed consent form and received monetary compensation for his or her time.

Stimuli

Using the Chinese Rich List (China Fuhao List, 2019), we selected 72 photos of faces (36 male and 36 female) as the targets and excluded all famous people, such as Jack Ma. Considering that the age of faces might affect the wealth ratings, we balanced the

TABLE 1 | Number of each age group of targets in Study 3.

Age group	Number of female/male faces
30–39	9/9
40–49	9/9
50–59	10/10
60 above	8/8

TABLE 2 | Average fortune of each group of targets in Study 3.

Group	Average fortune (billion RMB)	Average fortune of females (billion RMB)	Average fortune of males (billion RMB)
1	143.32	147.38	141.96
2	95.22	89.55	97.11
3	59.22	60.56	58.41
4	42.35	40.34	43.56
5	28.91	28.50	29.57
6	18.25	18.96	17.83
7	13.26	13.29	13.00
8	8.95	8.93	9.00
9	4.20	2.80	6.53

numbers of male and female faces in each age group and selected faces across a variety of age groups (see **Table 1**).

All the facial photos of wealthy people were found online. In the photos, their eyes look straight ahead. All the photos were manipulated to show the person from the shoulder up, against a white background and of the same size (295×295 pixels). Unlike in Studies 1 and 2, we used an objective standard of wealth rather than subjective self-reported attractiveness to compare with the participants' ratings when examining rating accuracy. In Study 3, based on wealth, we sorted wealthy people from high to low and divided them into nine equal groups (1-most wealthy group, 9-least wealthy group). By using the chi-square test, we ensured that in each wealth group, neither age, $\chi^2(264) = 278.40$, $p = 0.260$, nor gender made a difference, $\chi^2(8) = 13.00$, $p = 0.112$. We conducted a gender (facial gender: male, female) $\times$ group ANOVA and found that the main effect of facial gender was not significant, $F(1, 54) = 0.02$, $p = 0.878$, $\eta_p^2 = 0.000$. The main effect of group was significant, $F(8, 54) = 41.60$, $p < 0.001$, $\eta_p^2 = 0.860$. The interaction of facial gender and group was also not significant, $F(8, 54) = 0.07$, $p > 0.999$, $\eta_p^2 = 1.000$. These results suggested that the gender differences in the fortune of each group are negligible (see **Table 2**).

Procedure and Design

Guessing people's wealth based on facial appearance is impossible and lacks objective standards; thus, the participants saw an overview picture containing all the faces arranged randomly before the rating began. All the faces were arranged randomly into six lines, with 12 pictures per line. We made 10 versions of the overview pictures, one of which was shown randomly in one experiment. The picture was presented for 8,000 ms so that participants could establish an overall impression of these faces. The duration of the display was determined by a pilot study in which we asked people to look at the photos and react when

they thought they were ready for the next step. Then, they were told, "Each face belongs to 1 of the 9 richness wealth groups; please use your gut feeling to guess which group each face photo is in". Participants started with a practice block with eight trials to familiarize them with the task. The eight pictures used in the practice block (four male and four female) were chosen from the Chinese Rich List, which was the same source as the 72 experimental figures. However, they were not included in the 72 experimental figures, and the ratings of the eight faces were not included in the following analysis.
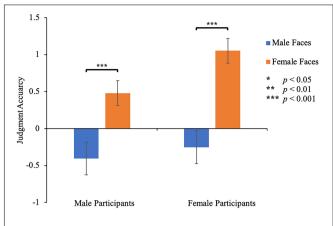
In a typical trial of the study, a fixation point was presented for 500 ms, and then a face photo was shown with a 9-point rating scale below it. Participants were asked to give their ratings on the wealth of the person depicted in each photo from 1 (most wealthy group) to 9 (least wealthy group). They were then asked to rate their confidence in the judgment they had made. Finally, they were asked to answer the question, "Are you familiar with this person?" by pressing 1 (no) or 9 (yes). The experiment contained two blocks with a total of 152 trials, and each face photo was presented once in a block. The order of the presentation was random. Between the two blocks, the participants were allowed to take a break and start the next block on their own if they thought they were ready for it. Considering that one participant rated the same target twice during the study, we used the mean rating for each face as the other-rating of the face. In the present research, we standardized the ratings of wealth for each face by subtracting other-ratings from objective ratings. Thus, standardized rating scores ranged from −8 to 8, with higher scores indicating that participants rated the target's wealth lower than the actual rating. The absolute value indicates the difference between objective rating scores and other-rating scores. To be more specific, a higher absolute value indicates that participants rated the target's wealth ranking lower than the target's actual wealth ranking. Positive or negative values suggest whether participants underestimated or overestimated the wealth rankings compared to an objective standard. All subsequent analyses were based on the standardized data.

To avoid the influence of familiarity, we excluded the data if the participant recognized the face in both blocks. Thus, some trials (2.71%) were not included in the following analysis.

Results
Richness Rating
Mean standardized ratings were submitted to a 2 (participant's gender) × 2 (facial gender) mixed-design measures ANOVA with facial gender as a within-subject factor (**Figure 4**). The main effect of facial gender was significant, $F(1, 48) = 84.17$, $p < 0.001$, $\eta_p^2 = 0.631$, indicating that people give more accurate ratings of wealth for male faces (-0.33 ± 1.10) than for female faces (0.77 ± 0.88). The main effect of the participant's gender was not significant, $F(1, 48) = 2.08$, $p = 0.156$, $\eta_p^2 = 0.042$. There was no significant interaction between the participant's gender and facial gender, $F(1, 48) = 3.08$, $p = 0.086$, $\eta_p^2 = 0.060$. Specifically, male participants rated male faces (-0.41 ± 0.22) more accurately than female faces (0.48 ± 0.17), $p < 0.001$. Female participants also rated male faces (-0.25 ± 0.22) more accurately than female

FIGURE 4 | Rating accuracy of wealth in Study 3. Error bars represent 1 S.E. of the means.

faces (1.05 ± 0.17), $p < 0.001$. These results indicated that all participants showed higher rating accuracy in judging the wealth of male faces than that of female faces. Additionally, for male faces, male raters and female raters had similar accuracy (-0.41 ± 0.22 vs -0.25 ± 0.22, $p = 0.626$), while for female faces, male raters gave more accurate ratings (0.48 ± 0.17) than female raters (1.05 ± 0.17, $p = 0.020$).

Confidence Rating
We conducted a 2 (participant's gender) × 2 (facial gender) mixed-design measures ANOVA with facial gender as a within-subject factor. The main effect of facial gender was not significant, $F(1, 48) = 0.02$, $p = 0.902$, $\eta_p^2 = 0.000$. The main effect of participant's gender was not significant, $F(1, 48) = 0.99$, $p = 0.325$, $\eta_p^2 = 0.020$. The interaction of the participant's gender and facial gender was also not significant, $F(1, 48) = 0.70$, $p = 0.407$, $\eta_p^2 = 0.014$. These results suggested that male and female participants were not aware of their own accuracy in judging others' wealth based on the facial appearance of different genders. Moreover, the correlation between confidence ratings and rating accuracy was not significant, $r = 0.01$, $p = 0.962$. Specifically, whether male participants rated male ($r = -0.05$, $p = 0.831$) and female faces ($r = 0.28$, $p = 0.174$) or female participants rated male ($r = 0.09$, $p = 0.662$) and female faces ($r = -0.13$, $p = 0.532$), they had low awareness of their rating accuracy.

GENERAL DISCUSSION

The present study showed that people give more accurate judgments of the facial attractiveness of female faces than of male faces and give more accurate wealth judgments for male faces than for female faces. To our knowledge, the current research is the first to show gender biases in the accuracy of impressions formed from faces. This indicates an important role of facial gender in shaping accurate first impressions.

The differences in judgment accuracy of male and female faces may be caused by differences in traditional gender roles. From an evolutionary perspective, these gender biases have been linked to

the production and survival of offspring. A man's reproductive potential is related more to his (economic) resources. In contrast, a woman's reproductive potential is associated more closely with her health, which may be related to physical attractiveness (Luxen and Van De Vijver, 2006). Thus, females might be more familiar with others' evaluations of their own facial attractiveness and thus achieve a higher level of consistency on self-other agreement. These results are also consistent with previous findings that facial gender is a salient facial cue in face processing and has an effect on other types of information (e.g., expression) processing (Liu et al., 2017). Moreover, Maner et al. (2003) found that both male and female observers selectively focus on physically attractive female targets according to the targets' facial photos, suggesting that people care more about female facial attractiveness than male facial attractiveness. The more attention that is paid to female facial attractiveness, the more accurate the judgments that can be made based on facial appearance.

In contrast to the findings about female faces in Studies 1 and 2, Study 3 revealed that people tend to rate perceived socioeconomic status (SES) more accurately for male faces than for female faces. In mate selection, SES is of great significance to males since females are more attentive to resources that can be invested in themselves and their offspring (Wang et al., 2018). Thus, on the one hand, males will expend more effort to increase their SES and recognize SES differences between themselves and competitors so that they can attract potential mates. On the other hand, females will seek as much evidence as possible to confirm their judgment of males' SES to help them "make a good choice". Moreover, because the number of male billionaires is larger than that of females all over the world (Wai, 2014; Forbes, 2022) and there is more media news or information related to wealthy males than to wealthy females, people may learn more useful cues to help them rate males' SES, even using only faces. Therefore, people's gender stereotypes are enhanced when SES is highly correlated with males in society. Similar gender bias is also found in research on how masculine facial cues play a key role in competence impressions (Oh et al., 2019). When people evaluate traits or personalities, the more evidence they accumulate and the more information they have observed and mastered, the higher the accuracy of their judgments and evaluations will be (Watson et al., 2000; Biesanz et al., 2007). These findings provide cross-validation of our hypothesis that people may pay more attention to the characteristics that are consistent with gender roles (e.g., the attractiveness of women, the socioeconomic status of men), thus accumulating more evidence that helps them make more accurate judgments.

The current findings regarding gender bias show the great social influences on gender differences. The higher accuracy of judgments of the facial attractiveness of female faces and of the wealth of male faces indicates that people can make relatively accurate judgments about these factors based only on faces. More importantly, it suggests that when the characteristics are consistent with gender stereotypes and are emphasized by society, people assign more attention to the characteristics of the gender. As a result, by accumulating more experience and evidence, people can make more accurate judgments. On the positive side, people can quickly establish a relatively accurate

impression of some characteristics that fit gender stereotypes to benefit their daily life interactions. However, the restricted accuracy of impressions based on face photos should receive more research attention. On the negative side, people put little effort into learning about characteristics that conflict with gender stereotypes, which might aggravate gender stereotypes across society. In addition, in Studies 2 and 3, we found that males rated characteristics that conflict with gender stereotypes more accurately than females did, which suggests that males might be affected less by gender stereotypes. This finding could be further examined in future research.

The analysis of confidence ratings implies that although the participants were able to make relatively accurate judgments, they may have struggled to be aware of their judgment accuracy. Participants might not realize whether they have extracted useful information from faces to help them make judgments. In addition, it is possible that they might not be sure of the gap between their own standards and external standards while giving their ratings. However, in Study 2, the significant moderate correlation between confidence ratings and rating accuracy when female participants rated male faces is interesting and is in line with research showing that females exhibit higher levels of interpersonal sensitivity than males (Chan et al., 2010). Despite female participants' higher accuracy when rating female faces, they had a clearer awareness when rating male faces. When rating male faces, even though male participants rated them more accurately, they failed to recognize their rating accuracy. However, we did not find a similar result in Study 3. Overall, these results show that although gender bias exists in terms of judgment accuracy, people do not have a relatively clear awareness of their rating behaviors and the gender bias of their judgments. This means that during the rating process, people might have underlying evaluation references that they are unaware of, which could be explored more thoroughly in the future.

CONCLUSION

The present study demonstrates that people evaluate females' attractiveness and males' perceived SES more accurately when looking at faces. Thus, we conclude that people evaluate the traits that they pay attention to more accurately based on facial appearance. In sum, these results reveal the effect of gender stereotypes on the judgment accuracy of impressions from faces. Accurate first impressions have a long-term effect on social relationship development (Human et al., 2013). The causes of this effect require more research. On the one hand, the present study illustrates that just by looking at faces, people can form relatively accurate impressions about traits that fit gender stereotypes. On the other hand, it shows the long-term and intensive impacts of social attitudes such as gender stereotypes on our daily life and social interactions.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://osf.io/dzfvp.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the internal review board of the Department of Psychology, Renmin University of China. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YQ designed and performed the experiments and wrote the manuscript. JY performed the experiments, analyzed the data, and wrote the manuscript. Both authors contributed to the article and approved the submitted version.

REFERENCES

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., and Breiter, H. C. (2001). Beautiful faces have variable reward value. *Neuron* 32, 537–551. doi: 10.1016/S0896-6273(01)00491-3

Alaei, R., and Rule, N. O. (2016). "Accuracy of perceiving social attributes," in *The Social Psychology of Perceiving Others Accurately*, eds J. A. Hall, M. Schmid Mast, and T. V. West (Cambridge: Cambridge University Press), 125–142. doi: 10.1017/CBO9781316181959.006

Atwater, L. E., and Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychol.* 45, 141–164. doi: 10.1111/j.1744-6570.1992.tb00848.x

Atwater, L. E., and Yammarino, F. J. (1997). "Self-other rating agreement: a review and model," in *Research in Personnel and Human Resources Management*, ed. G. R. Ferris (Greenwich, CT: Elsevier Science), 121–174.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., et al. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychol. Sci.* 21, 372–374. doi: 10.1177/0956797609360756

Baudouin, J.-Y., and Tiberghien, G. (2004). Symmetry, averageness, and feature size in the facial attractiveness of women. *Acta Psychol.* 117, 313–332. doi: 10.1016/j.actpsy.2004.07.002

Bernieri, F. J., Zuckerman, M., Koestner, R., and Rosenthal, R. (1994). Measuring person perception accuracy: another look at self-other agreement. *Personali. Soc. Psychol. Bull.* 20, 367–378. doi: 10.1177/0146167294204004

Biesanz, J. C., West, S. G., and Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *J. Personali. Soc. Psychol.* 92, 119–135. doi: 10.1037/0022-3514.92.1.119

Borkenau, P., and Liebler, A. (1993). Consensus and self-other agreement for trait inferences from minimal information. *J. Personali.* 61, 477–496. doi: 10.1111/j.1467-6494.1993.tb00779.x

Borkenau, P., Brecke, S., Möttig, C., and Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *J. Res. Personali.* 43, 703–706. doi: 10.1016/j.jrp.2009.03.007

Brauer, K., and Proyer, R. T. (2020). Judging dispositions toward ridicule and being laughed at from short self-descriptions at zero-acquaintance: testing self-other agreement, consensus, and accuracy. *J. Res. Personali.* 89:104016. doi: 10.1016/j.jrp.2020.104016

Buss, D. M., and Schmitt, D. P. (1993). Sexual strategies theory: an evolutionary perspective on human mating. *Psychol. Rev.* 100, 204–232. doi: 10.1037/0033-295X.100.2.204

Buunk, B. P., Dijkstra, P., Fetchenhauer, D., and Kenrick, D. T. (2002). Age and gender differences in mate selection criteria for various involvement levels. *Personal Relation.* 9, 271–278. doi: 10.1111/1475-6811.00018

Chan, M., Rogers, K. H., Parisotto, K. L., and Biesanz, J. C. (2010). Forming first impressions: the role of gender and normative accuracy in personality perception. *J. Res. Personali.* 45, 117–120. doi: 10.1016/j.jrp.2010.11.001

Chen, A. C., German, C., and Zaidel, W. (1997). Brain asymmetry and facial attractiveness: facial beauty is not simply in the eye of the beholder. *Neuropsychologia* 35, 471–476. doi: 10.1016/S0028-3932(96)00065-6

China Fuhao List (2019). Available online at: https://www.phb123.com/renwu/fuhao/px_CN.html (Accessed October 2019)

Cloutier, J., Heatherton, T. F., Whalen, P. J., and Kelley, W. M. (2008). Are attractive people rewarding? sex differences in the neural substrates of facial attractiveness. *J. Cogn. Neurosci.* 20, 941–951. doi: 10.1162/jocn.2008.20062

Collins, S. A., and Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behav.* 65, 997–1004. doi: 10.1006/anbe.2003.2123

Colwell, J. (2007). An empirical test of sex differences in the emphasis on physical attractiveness in mate selection. *Percept Mot. Skills* 105, 326–338. doi: 10.2466/pms.105.1.326-338

Devos, G., Hulpia, H., Tuytens, M., and Sinnaeve, I. (2013). Self-other agreement as an alternative perspective of school leadership analysis: an exploratory study. *School Effect. School Improv.* 24, 296–315. doi: 10.1080/09243453.2012.693103

Elder, G. H. (1969). Appearance and education in marriage mobility. *Am. Sociol. Rev.* 34, 519–533. doi: 10.2307/2091961

Forbes, B. (2022). *The Richest People in the World (2022)*. Available online at: https://www.forbes.com/billionaires/ (accessed February 2, 2022)

Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychol. Rev.* 102, 652–670. doi: 10.1037/0033-295X.102.4.652

Funder, D. C., and Dobroth, K. M. (1987). Differences between traits: properties associated with interjudge agreement. *J. Personali. Soc. Psychol.* 52, 409–418. doi: 10.1037/0022-3514.52.2.409

Funder, D. C., and West, S. G. (1993). Consensus, self-other agreement, and accuracy in personality judgment: an introduction. *J. Personali.* 61, 457–476. doi: 10.1111/j.1467-6494.1993.tb00778.x

Funder, D. C., Kolar, D. C., and Blackman, M. C. (1995). Agreement among judges of personality: interpersonal relations, similarity, and acquaintanceship. *J. Personali. Soc. Psychol.* 69, 656–672. doi: 10.1037/0022-3514.69.4.656

Gosling, S. D., Gaddis, S., and Vazire, S. (2007). Personality impressions based on facebook profiles. *ICWSM* 7, 1–4.

Human, L. J., Sandstrom, G. M., Biesanz, J. C., and Dunn, E. W. (2013). Accurate first impressions leave a lasting impression: the long-term effects of distinctive self-other agreement on relationship development. *Soc. Psychol. Personali. Sci.* 4, 395–402. doi: 10.1177/1948550612463735

Kim, H., Di Domenico, S., and Connelly, B. (2019). Self–other agreement in personality reports: a meta-analytic comparison of self- and informant-report means. *Psychol. Sci.* 30, 129–138. doi: 10.1177/0956797618810000

Komori, M., Kawamura, S., and Ishihara, S. (2009). Averageness or symmetry: which is more important for facial attractiveness? *Acta Psychol.* 131, 136–142. doi: 10.1016/j.actpsy.2009.03.008

Lee, A., and Carpenter, N. C. (2018). Seeing eye to eye: a meta-analysis of self-other agreement of leadership. *Lead. Quart.* 29, 253–275. doi: 10.1016/j.leaqua.2017.06.002

Levy, B., Ariely, D., Mazar, N., Chi, W., Lukas, S., and Elman, I. (2008). Gender differences in the motivational processing of facial beauty. *Learn. Motivat.* 39, 136–145. doi: 10.1016/j.lmot.2007.09.002

Li, N. P., Bailey, J. M., Kenrick, D. T., and Linsenmeier, J. A. W. (2002). The necessities and luxuries of mate preferences: testing the tradeoffs. *J. Personali. Soc. Psychol.* 82, 947–955. doi: 10.1037/0022-3514.82.6.947

Li, Y., Chen, Z., Liu, X., and Qi, Y. (2021). Perceiving the facial trustworthiness: facial age, emotional expression, and attractiveness. *Quart. J. Exp. Psychol.* 2021:174702182110471. doi: 10.1177/17470218211047176

Little, A. C., and Perrett, D. I. (2002). Putting beauty back in the eye of the beholder. *Psychologist* 15, 28–32.

Little, A. C., Jones, B. C., Penton-Voak, I. S., Burt, D. M., and Perrett, D. I. (2002). Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape. *Proc. R. Soc. Lond. B.* 269, 1095–1100. doi: 10.1098/rspb.2002.1984

Little, A. C., McPherson, J., Dennington, L., and Jones, B. C. (2011). Accuracy in assessment of self-reported stress and a measure of health from static facial information. *Personali. Indiv. Diff.* 51, 693–698. doi: 10.1016/j.paid.2011.06.010

Little, A. C., Saxton, T. K., Roberts, S. C., Jones, B. C., DeBruine, L. M., Vukovic, J., et al. (2010). Women's preferences for masculinity in male faces are highest during reproductive age range and lower around puberty and post-menopause. *Psychoneuroendocrinology* 35, 912–920. doi: 10.1016/j.psyneuen.2009.12.006

Liu, C., Liu, Y., Iqbal, Z., Li, W., Lv, B., and Jiang, Z. (2017). Symmetrical and asymmetrical interactions between facial expressions and gender information in face perception. *Front. Psychol.* 8:1383. doi: 10.3389/fpsyg.2017.01383

Luxen, M. F., and Van De Vijver, F. J. R. (2006). Facial attractiveness, sexual selection, and personnel selection: when evolved preferences matter. *J. Organiz. Behav.* 27, 241–255. doi: 10.1002/job.357

Maner, J. K., Kenrick, D. T., Becker, D. V., Delton, A. W., Hofer, B., Wilbur, C. J., et al. (2003). Sexually selective cognition: beauty captures the mind of the beholder. *J. Personali. Soc. Psychol.* 85, 1107–1120. doi: 10.1037/0022-3514.85.6.1107

Mitrovic, A., Goller, J., Tinio, P. P. L., and Leder, H. (2018). How relationship status and sociosexual orientation influence the link between facial attractiveness and visual attention. *PLoS One* 13:e0207477. doi: 10.1371/journal.pone.0207477

Moritz, D., and Roberts, J. E. (2018). Self-other agreement and metaperception accuracy across the big five: examining the roles of depression and self-esteem. *J. Pers.* 86, 296–307. doi: 10.1111/jopy.12313

Oh, D., Buck, E. A., and Todorov, A. (2019). Revealing hidden gender biases in competence impressions of faces. *Psychol. Sci.* 30, 65–79. doi: 10.1177/0956797618813092

Pandey, G., and Zayas, V. (2021). What is a face worth? Facial attractiveness biases experience-based monetary decision-making. *Br. J. Psychol.* 112, 934–963. doi: 10.1111/bjop.12495

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., et al. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature* 394, 884–887. doi: 10.1038/29772

Qi, Y., Li, Q., and Du, F. (2018). Are rich people perceived as more trustworthy? Perceived socioeconomic status modulates judgments of trustworthiness and trust behavior based on facial appearance. *Front. Psychol.* 9:512. doi: 10.3389/fpsyg.2018.00512

Qi, Y., Luo, Y., Feng, Y., Chen, Z., Li, Q., Du, F., et al. (2021). Trustworthy faces make people more risk-tolerant: the effect of facial trustworthiness on risk decision-making under gain and loss conditions. *PsyCh J.* 11, 43–50. doi: 10.1002/pchj.498

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annu. Rev. Psychol.* 57, 199–226. doi: 10.1146/annurev.psych.57.102904.190208

Rogers, K. H., Wood, D., and Furr, R. M. (2018). Assessment of similarity and self-other agreement in dyadic relationships: a guide to best practices. *J. Soc. Personal Relat.* 35, 112–134. doi: 10.1177/0265407517712615

Ru, T., Fan, H., Chen, Q., Li, J., and Zhou, G. (2017). Electrophysiological correlates of processing female facial attractiveness and its influence on male honesty. *J. Psychol. Sci.* 40, 1428–1432. doi: 10.16719/j.cnki.1671-6981.20170623

Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception* 32, 1093–1107. doi: 10.1068/p5101

Shackelford, T. K., Schmitt, D. P., and Buss, D. M. (2005). Universal dimensions of human mate preferences. *Personal. Indiv. Diff.* 39, 447–458. doi: 10.1016/j.paid.2005.01.023

Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., et al. (1998). Newborn infants prefer attractive faces. *Infant Behav. Dev.* 21, 345–354. doi: 10.1016/S0163-6383(98)90011-X

Sprecher, S., Sullivan, Q., and Hatfield, E. (1994). Mate selection preferences: gender differences examined in a national sample. *J. Pers Soc. Psychol.* 66, 1074–1080. doi: 10.1037//0022-3514.66.6.1074

Thornhill, R., and Gangestad, S. W. (1999). The scent of symmetry a human sex pheromone that signals fitness? *Evolu. Hum. Behav.* 20, 175–201. doi: 10.1016/S1090-5138(99)00005-7

Todorov, A., Olivola, C. Y., Dotsch, R., and Mende-Siedlecki, P. (2015). Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annu. Rev. Psychol.* 66, 519–545. doi: 10.1146/annurev-psych-113011-143831

Tsankova, E., and Tair, E. (2021). Meta-accuracy of very first impressions: a mini review. *Front. Psychol.* 12:736534. doi: 10.3389/fpsyg.2021.736534

Udry, J. R. (1977). The importance of being beautiful: a reexamination and racial comparison. *Am. J. Sociol.* 83, 154–160. doi: 10.1086/226511

Wade, R. (1988). *Village Republics: Economic Conditions for Collective Action in South India.* Cambridge: Cambridge University Press.

Wai, J. (2014). Investigating the world's rich and powerful: Education, cognitive ability, and sex differences. *Intelligence* 46, 54–72. doi: 10.1016/j.intell.2014.05.002

Walker, M., and Vetter, T. (2016). Changing the personality of a face: perceived big two and big five personality factors modeled in real photographs. *J. Personali. Soc. Psychol.* 110, 609–624. doi: 10.1037/pspp0000064

Wang, G., Cao, M., Sauciuvenaite, J., Bissland, R., Hacker, M., Hambly, C., et al. (2018). Different impacts of resources on opposite sex ratings of physical attractiveness by males and females. *Evolu. Hum. Behav.* 39, 220–225. doi: 10.1016/j.evolhumbehav.2017.12.008

Wang, Y., Yao, P., and Zhou, G. (2015). The influence of facial attractiveness and personality labels on men and women's mate preference. *Acta Psychol. Sin.* 47:108. doi: 10.3724/SP.J.1041.2015.00108

Watson, D., Hubbard, B., and Wiese, D. (2000). Self-other agreement in personality and affectivity: the role of acquaintanceship, trait visibility, and assumed similarity. *J. Personali. Soc. Psychol.* 78, 546–558. doi: 10.1037/0022-3514.78.3.546

Welling, L. L. M., DeBruine, L. M., Little, A. C., and Jones, B. C. (2009). Extraversion predicts individual differences in women's face preferences. *Personali. Indiv. Diff.* 47, 996–998. doi: 10.1016/j.paid.2009.06.030

Zhang, L., Chen, W., Liu, M., Ou, Y., Xu, E., and Hu, P. (2021). Light makeup decreases receivers' negative emotional experience. *Sci. Rep.* 11:23802. doi: 10.21203/rs.3.rs-572162/v1

Zhang, Y., Kong, F., Guo, Y., Ren, J., and Chen, H. (2012). Explicit memory or implicit memory: an ERP study on memory bias toward female facial attractiveness. *Beijing Daxue Xuebao (Ziran Kexue Ban)/Acta Scientiarum Naturalium Universitatis Pekinensis* 48, 160–168.

Zheng, R., and Zhou, G. (2021). Vitality makes dynamic faces more attractive than static faces. *J. Vis.* 21:2361. doi: 10.1167/jov.21.9.2361