**Results**

*3.1 Recognising ability by direct/explicit judgement*

In the discrimination task, our aim was to check subjects’ ability to identify synthetic from real faces by direct/explicit decision. Collapsed across all images, 48.3% were labelled correctly either as “real” or “synthetic”. A t-test was used to examine the difference between the accuracy rates of real and synthetic faces. The accuracy rate of real faces (*M* = 50.9%, *SD* = 20.3%) was significantly higher than the accuracy rate of synthetic faces (*M* = 45.7%, *SD* = 19.7%), *t* (356) = 3.86, *p* < 0.001, with a small effect size (Cohen’s *d* = .20). The evidence showed that by direct/explicit judgement, subjects could slightly better recognize real faces than synthetic faces. But accuracy rates on real and synthetic faces were both close to chance (50%), suggesting that subjects’ ability to identify synthetic from real faces was poor.

*3.2 Recognising by trustworthiness perception*

In the trustworthiness rating task, one of aims was to check whether there was a difference of subjects’ trustworthiness perception between real and fake faces. Subjects viewed 16 faces (including real and synthetic faces) and rated them how trustworthy they thought each face looked on a trustworthiness continuum where 1 indicated extremely untrustworthy and 7 indicated extremely trustworthy. The experimental results show that, the mean trustworthiness rating score for all faces was (*M =* 4.2, *SD* = 0.6). There was a statistically significant difference of mean trustworthiness rating scores between real and synthetic faces, *t* (356) = -10.69, *p* < 0.001, for real faces (*M* = 3.9, *SD* = 0.8), for synthetic faces (*M* = 4.4, *SD* = 0.8), with a medium effect size (Cohen’s *d* = .57), indicating that target face authenticity had a medium effect on trustworthiness perception. This might be an implication that there is a specific mechanism existing in the synthesis process of StyleGAN2 that enhances the level of perceived trustworthiness from synthetic faces, resulted in that synthetic faces received higher trustworthiness scores than real faces. Here, typicality is speculated to be related to this specific mechanism. This will be discussed further in the Discussion section.

*3.3 Gender bias of trustworthiness perception*

The experimental analysis further employed ANOVAs to examine whether there was an interaction effect between three independent variables (subject gender, target face gender, target face authenticity) on trustworthiness perception. Excluding those who identified as non-binary (N = 10) and those who did not provide gender information (N = 6), 341 subjects (114 males, 227 females) were included in this analysis part. At this part, target face authenticity (real/synthetic) was thought as the focal factor, and subject gender (male/female) and target face gender (male/female) were thought of as moderator factors, namely that the effect of target face authenticity on trustworthiness perception depended on subject gender and target face gender. Due to the small sample size of non-binary subjects, their data along with data for those selecting “prefer not to say” on the gender question were excluded from this part of analysis.

*3.3.1 Three-way Interaction*

A three-way factorial ANOVA was conducted with one between group factor [subject gender (male/female)] and two within group factors [target face gender (male/female) × target face authenticity (real/synthetic)]. The ANOVA analysis indicated that there was no significant three-way interaction between these three factors, . Also, there was no simple two-way interaction between any two of the three factors. There was no significant interaction between subject gender and target face gender,, indicating that the effect of subject gender on trustworthiness perception did not depend on the level of target face gender. There was no significant interaction between subject gender and target face authenticity,, indicating that the effect of subject gender on trustworthiness perception did not depend on the level of target face authenticity There was no significant interaction between target face gender and target face authenticity,, indicating that the effect of target face gender on perceived trustworthiness did not depend on the level of target face authenticity.

Furthermore, the result showed that there was no significant simple main effect of subject gender, , namely that male subjects and female subjects perceived a similar level of trustworthiness at each level of target face gender. This indicated that male subjects and female subjects show similar perceptual characteristics in trustworthiness for human faces. Neither gender was more inclined to perceive people’s faces as more trustworthy. However, the result showed a significant main effect of target face authenticity, ; and a significant main effect of target face gender, .

*3.3.2 Gender bias of target faces*

As reported above that no significant interaction between target face gender and target face authenticity, the analysis compared the difference between real and synthetic faces respectively in both genders. The post hoc tests used Tukey HSD to check the main effect of two factors at the same time [target face gender (male/female) and target face authenticity (real/synthetic)]. For the main effect of target face gender, male target faces (*M* = 4.0, *SD* = 0.7) received significantly lower trustworthiness rating scores than female target faces (*M* = 4.3, *SD* = 0.7) with a *p-*value of < 0.001, yet the effect was small (Cohen’s *d* = 0.11), indicating that female target faces were on average perceived as more trustworthy than male target faces. For the main effect of target face authenticity, the results showed male synthetic faces (*M* = 4.2, *SD* = 0.9) were perceived as more trustworthy than male real faces (*M* = 3.8, *SD* = 0.9), and female synthetic faces (*M* = 4.5, *SD* = 0.9) were perceived as more trustworthy than female real faces (*M* = 4.0, *SD* = 0.9). To conclude, for both real and synthetic faces, the effect of target face gender was similar. Synthetic faces in both genders displayed a certain extent of enhancement at the aspect of perceived trustworthiness as compared to real faces (see Fig. 1).

图表, 条形图

描述已自动生成

**Fig. 1.** Differences in mean trustworthiness rating scores by target face gender and authenticity. Synthetic faces received higher trustworthiness scores than real faces. Female faces received higher trustworthiness scores than male faces.

*3.4 Ethnicity bias of trustworthiness perception*

ANOVAs were also conducted to better interpret and understand the effect of ethnicity on the perception of trustworthiness for human faces. Due to the small sample size of other ethnicities (Black, Hispanic, Middle Eastern, South Asian, Others), only data for East Asian subjects and White subjects (*N*=329: 237 East Asian, 92 White) were included in the subject ethnicity bias analysis.

*3.4.1 Three-way interaction*

Again, a three-way factorial ANOVA was conducted with one between group factor [subject ethnicity (East Asian/White)] and two within group factors [target face ethnicity (Black, East Asian, South Asian, White) × target face authenticity (real/synthetic)]. The result of the three-way ANOVA analysis indicated that there was no statistically significant three-way interaction, . However, there were two-way interactions between any two of the three factors. There was significant interaction between subject ethnicity and target face ethnicity,, indicating that the effect of subject ethnicity was different on different levels of target face ethnicity. There was a significant interaction between subject ethnicity and target face authenticity,, indicating that the effect of subject ethnicity was different on real and synthetic faces. Also, there was a significant interaction between target face ethnicity and target face authenticity,, indicating that the effect of target face ethnicity was different on real and synthetic faces. Furthermore, the result showed that all three factors had significant simple main effects on trustworthiness rating score: for subject ethnicity, ; for target face authenticity, ; for target face ethnicity, . Further, the post hoc test was conducted to better interpret the interaction effects.

*3.4.2* *Interaction between subject ethnicity and target face ethnicity*

Follow-up comparisons were performed using Tukey HSD test to analyse the interaction between two factors [subject ethnicity (East Asian/White) × target face ethnicity (Black/East Asian/South Asian/White)]. The mean perceived trustworthiness scores of different combinations of the two factors were mutually compared. The results indicated that the mean trustworthiness score of White subjects rating Black faces (*M* = 4.7, *SD* = 0.9) was significantly higher than that of they rating South Asian faces (*M* = 4.3, *SD* = 0.6; *p* = .012) and White faces (*M* = 4.0, *SD* = 0.8; *p* < .001), also significantly higher than that of East Asian rating Black faces (*M* = 4.0, *SD* = 1.0; *p* < .001). The mean trustworthiness score of East Asian subjects rating East Asian faces (*M* = 3.9, *SD* = 0.9) was significantly lower than that of they rating South Asian faces (*M* = 4.1, *SD* = 0.9; *p* = .03) and White faces (*M* = 4.2, *SD* = 0.9; *p* < .001), also significantly lower than that of White subjects rating East Asian faces (*M* = 4.4, *SD* = 0.7; *p* < .001). Whereas the effect size calculations produced indications of a small effect of subject ethnicity (Cohen’s *d* = 0.10) and a small effect of target face ethnicity (Cohen’s *d* = 0.06). In general conclusion, White and East Asian subjects had different trustworthiness perception from human faces of various ethnicities. Noticeably, both White and East Asian subjects rated their own ethnicities as the lowest level of trustworthiness, which in a certain extent presented the own-ethnicity bias as a negative impact on trustworthiness perception (see Fig. 2).

图表, 箱线图

描述已自动生成

**Fig. 2.** Distribution of trustworthiness rating scores by subject ethnicity and target face ethnicity. The lines across the middle of the boxes indicate medians. The upper and lower border lines indicate the third quartile and the first quartile. The ends of the lines extending from the boxes indicate the maximum and minimum of values in datasets. Both White and East Asian subjects rated their own ethnicities as the least trustworthy.

*3.4.3 Interaction between target face ethnicity and target face authenticity*

To further examine the characteristics of synthetic faces on trustworthiness perception, the multiple comparisons using Tukey HSD test to analyse the interaction between two factors [target face authenticity (real/synthetic) × target face ethnicity (Black/East Asian/South Asian/White)]. The mean perceived trustworthiness scores of different combinations of the two factors were compared and summarized as follows. For real type of target face authenticity, the mean trustworthiness score of Black faces (*M* = 4.1, *SD* = 1.2) was only significantly higher than that of South Asian faces (*M* = 3.8, *SD* = 1.2; *p* < .001) and White faces (*M* = 3.7, *SD* = 1.2; *p* < .001), with target face ethnicity showing a small effect size (Cohen’s *d*= 0.14). For synthetic type of target face authenticity, inversely, the mean trustworthiness score of Black faces (*M* = 4.3, *SD* = 1.1) was significantly lower than that of South Asian faces (*M* = 4.6, *SD* = 1.1; *p* = .018) and White faces (*M* = 4.6, *SD* = 1.1; *p* = .002); and the mean trustworthiness score of East Asian faces (*M* = 4.1, *SD* = 1.2) was also significantly lower than that of South Asian faces (*p* < .001) and White faces (*p* < .001), with target face ethnicity again showing a small effect size (Cohen’s *d* = 0.04). For the comparisons between real and synthetic types across target face ethnicities, South Asian faces had a significant difference between real type and synthetic type (*p* < .001), and White faces also had a significant difference between real type and synthetic type (*p* < .001); yet neither East Asian faces (*p* = .99) nor Black faces (*p* = .83) had significant difference between real and synthetic types, with target face ethnicity showing a small effect size (Cohen’s *d* = 0.04). In conclusion, the differences in perceived trustworthiness between real and synthetic faces varied across the four experimental ethnicities (Black/East Asian/South Asian/White), specifically that South Asian synthetic faces seemed more trustworthy than South Asian real faces, and White synthetic faces seemed more trustworthy than White real faces, yet neither East Asian faces nor Black faces revealed a similar discrepancy (see Fig. 3).

*图表, 箱线图

描述已自动生成*

**Fig. 3.** Distribution of trustworthiness rating scores by target face ethnicity and target face authenticity. The lines across the middle of the boxes indicate medians. The upper and lower border lines indicate the third quartile and the first quartile. The ends of the lines extending from the boxes indicate the maximum and minimum of values in datasets.