**Data Acquisition Form 2021**

**Project Title:** Do trustworthiness judgements help people to recognise synthetic faces?

**Creator Name:** Haisa Shan

**Date:**  8 September 2021

**Collection**

Questionnaire-Based Study

**Abstract**

Recent advances in digital image generative models have allowed for artificial creation of fake imagery such as synthesising highly photorealistic human faces. Style-based Generative Adversarial Networks (StyleGAN) is one of the most state-of-the-art generative models in this field, and has been widely used on facial image generation. However, with the increasing ease of using such image generative models, the security in many domains, such as forensic, border control and mass media, is vulnerable in front of the potential threats resulted from the misuse of image generative technologies. To date there has only been limited empirical research into the facial characteristics of StyleGAN-generated faces to support the design of detection methods against such synthetic faces. This study used StyleGAN2 (an improved version of StyleGAN) to generate faces and invited people to complete two facial image evaluation tasks, 1) Discrimination task, 2) Trustworthiness rating task. The study results demonstrated that, in the discrimination task, subjects had trouble recognising synthetic faces by direct/explicit judgement; while in the trustworthiness rating task, subjects perceived the synthetic faces as significantly more trustworthy than real faces. The study further analysed gender bias and ethnicity bias on the perception of facial trustworthiness, with results showing some differences between different levels of gender and ethnicity. In conclusion, people’s ability to recognise synthetic faces is poor, but it is possible that people rely on the perception of facial trustworthiness to discriminate synthetic from real faces. The findings in this study have implications for the development of detection methods against digitally generated faces.

**Keywords**

StyleGAN, synthetic face, trustworthiness perception, facial trustworthiness

**Method**

Subjects and design

Three hundred and fifty-seven subjects (114 males, mean age = 25.2, SD = 5.8; 227 females, mean age = 25.0, SD = 6.3; 10 non-binary, mean age = 23.6, SD = 8.93) were recruited to complete an online survey test delivered on [www.qualtrics.com](http://www.qualtrics.com). The responses of subjects who started but did not complete the online survey were eliminated to avoid distorting the research results. We used computer-synthesised facial images in this research as fake faces, mixed with real faces to examine people’s ability to detect fake faces and perceptual differences of trustworthiness between real/fake faces. Subjects did not get rewards for their participation, though they could see the test score of their performances at the end of the survey. The Qualtrics survey was based on a within-subjects design in which all subjects viewed the same two sets of adult facial images and completed each of the two tasks. To eliminate the effect of between-sets difference, the use of each image sets was counterbalanced in the individual test for each subject. Before the survey started, all subjects provided informed consent and completed a demographic questionnaire about their age, gender, ethnicity. In terms of the experimental power of 0.8 and significance level of 0.05, with a small effect, the power calculation indicated that the study needed at least 198 subjects.

Stimuli

A total of thirty-two human facial images (1024×1024 resolution), including 16 real and 16 synthetic faces, were used as stimuli in the survey. All real faces were taken from a publicly available dataset for high-quality human facial images, Flickr-Faces-HQ (FFHQ), which is created as a benchmark for GAN (see <https://github.com/NVlabs/ffhq-dataset>), and all synthetic faces were gained from the dataset of the generative image modeling, StyleGAN2 (see <https://github.com/NVlabs/stylegan2>). To ensure a diverse dataset, in each of the two sets of faces, there were 4 Black, 4 East Asian, 4 South Asian, 4 White, and 2 males and 2 females for each ethnicity. Among the sixteen faces of each set, half of them were real and half were synthetic, but this was unknown to subjects.

Procedure

First, subjects completed a short questionnaire for demographic information (age, gender, ethnicity), and subjects had to be 18 years of age or older to take part. Prior to the main body of test, there was an example of real and synthetic faces presented to provide subjects with a general impression of what real and synthetic faces look like. Subjects then were asked to complete two face evaluating tasks, 1) Discrimination Task, 2) Trustworthiness Rating Task. The two tasks were presented to subjects in a counterbalanced order to check for any possible order effects. Before the start of each task, participants were informed that they would see a series of 16 facial images, and that they had to carry out their evaluation following the instructions provided. In both tasks, only one image was presented at a time and individual images appeared in a random order.

In the discrimination task, participants made their decision between two choices, “real” or “synthetic”, to classify the 16 faces according to whether they thought the presented faces were real or not. Subjects did not receive immediate feedback during the task on the correctness of their classifications. In this task, subjects relied on direct/explicit judgments. In the trustworthiness rating task, subjects were required to rate how trustworthy they thought each of 16 faces looked using a 7-point Likert scale (1 = extremely untrustworthy; 4 = neither untrustworthy nor trustworthy; 7 = extremely trustworthy). We instructed subjects that they did not need to consider face authenticity in this task, and they could just assume that the faces shown to them were all of real people. Although there was no time limit to respond for trustworthiness rating, we encouraged subjects to rely on their intuitions and provide their responses to work as quickly as possible. In this task, we expected to trigger a relatively indirect/implicit approach to evaluate faces as compared to direct/explicit judgement on face authenticity, specifically by trustworthiness perception. At the end of the survey, subjects saw a result report of their own mean trustworthiness rating scores for real and synthetic faces, and their mean accuracy in classifying real and synthetic faces in the discrimination task.

**Format of data**

data/Excel.csv

**Rights**

Open

**Relation**

None

**Language**

Enghlish

**Supervisor Name:** Sophie Nightingale

**Project Level**

Undergraduate

Masters

**Topic**

*Please select all the sub-categories of Psychology that the project falls under*

Clinical

Clinical, Developmental

Cognitive, Developmental

Cognitive, Perception

Cognitive, Psychopharmacology

Developmental

Developmental, Neuropsychology

Forensic

Historical

Marketing

Modelling (Computational)

Neuropsychology

Psychopharmacology

Psychophysics

Social

Other

*If you selected “Other” and the topic of your project is not listed above, please state your topic(s) in this box*

**Sample Size:** 357 Participants

**Statistical Analysis Type**

*Please select all the types of statistical analysis used in your project*

ANCOVA

ANOVA

Bayesian Analysis

Chi-Squared

Correlation

Factor Analysis

Latent Class Analysis

Linear Mixed Effects Modelling

Log-Linear Analysis

MANCOVA

Monte Carlo Simulations

Multi-Dimensional Scaling

Power Analysis

Psychometrics

Qualitative

Qualitative (Conversation Analysis)

Qualitative (Discourse Analysis)

Qualitative (Thematic Analysis)

Regression

Signal Detection Theory

Structural Equation Modelling

T-Test

Other

*If you selected “Other” and the statistical analysis used in your project is not listed above, please state your analysis/analyses in this box*

**Data Files**

*Please ensure that you attach the documents listed below alongside your data acquisition form and provide the names of the files for ease of identification.*

**Blank Consent Form:** 文本, 信件

描述已自动生成

**Data File(s):** Raw\_Data.csv, Processed\_Data.csv, gender\_gender\_authenticity.csv, ethnicity\_ethnicity\_anthenticity.csv, R\_analysis.R, Results.docx

**Codebook:** Codebook.docx