Do 3D face images capture cues of strength, weight, and height better than 2D face images do?

Iris J Holzleitner (Institute of Neuroscience & Psychology, University of Glasgow)

Alex L Jones (Department of Psychology, Swansea University)

Victor Shiramizu (Institute of Neuroscience & Psychology, University of Glasgow)

Kieran J O'Shea (Institute of Neuroscience & Psychology, University of Glasgow)

Vanessa Fasolt (Institute of Neuroscience & Psychology, University of Glasgow)

Benedict C Jones (Institute of Neuroscience & Psychology, University of Glasgow)

Lisa M DeBruine (Institute of Neuroscience & Psychology, University of Glasgow)

Author Note
This research is supported by a European Research Council grant awarded to LMD (#647910 KINSHIP).
Do 3D face images capture cues of strength, weight, and height better than 2D face images do?

Abstract
A large literature exists investigating the extent to which physical characteristics (e.g., strength, weight, and height) can be accurately assessed from face images. While most of these studies have employed two dimensional (2D) face images as stimuli, some recent studies have used three dimensional (3D) face images because they may contain cues not visible in 2D face images. We know of no direct empirical tests of this claim and the equipment required for 3D face images is considerably more expensive than that required for 2D face images. Consequently, we will test whether 3D face images capture cues of strength, weight, and height better than 2D face images do by directly comparing the accuracy of strength, weight, and height ratings of 2D and 3D face images taken simultaneously.

Keywords: face perception, attributions, accuracy, formidability

Introduction
There now exists a large literature examining the extent to which physical characteristics, such as formidability and health, can be accurately assessed from face images (for reviews of this literature see De Jager et al., 2018 and Puts, 2010). The results of such studies have implications for evolutionary theories of the signal value of facial characteristics in both human mate choice and intrasexual competition (De Jager et al., 2018; Puts, 2010).

Studies on this topic typically employ two dimensional (2D) face images as stimuli (Coetzee et al., 2009, 2010; Re et al., 2013; Tinlin et al., 2013; Sell et al., 2008). However, some studies have used three dimensional (3D) face images because they may contain cues that are not captured well in 2D face images and, therefore, are likely to have greater ecological validity (Holzleitner et al., 2014; Re et al., 2012). The equipment needed to obtain high quality 3D face images is considerably more expensive than the equipment needed to obtain high quality 2D face images. However, it is not known whether this additional cost is warranted (i.e., it is not known whether 3D face images capture cues of physical characteristics better than 2D face images do). The current study investigates this issue by comparing the accuracy of perceptions of strength, weight, and height from 2D and 3D face images.

Several studies have reported positive correlations between measures of upper-body strength, such as handgrip strength, and perceptual judgments (i.e., ratings) of strength from both 2D face photographs (Sell et al., 2008) and 3D face images (Holzleitner & Perrett, 2016). Studies have also reported positive correlations between body mass index (BMI) and ratings of weight from
2D face photographs (Coetzee et al., 2009, 2010; Tinlin et al., 2013) and 3D face images (Holzleitner et al., 2014). Other studies have reported positive correlations between height and ratings of height from 2D face photographs (Re et al., 2013) and 3D face images (Re et al., 2012; but see also Holzleitner et al., 2014). These studies suggest strength, weight, and height can be judged somewhat accurately from face images, but no studies have yet directly compared the effect of stimulus type (2D versus 3D face images) on the accuracy of perceptions of these (or any other) traits. While some previous work has used both 2D and 3D images, these studies have focused on testing attractiveness and formidability ratings only (Tigue et al., 2012; Trebicky et al., 2018), or similarities in face recognition (e.g., Eng et al., 2017) and morphological measurements (e.g., Hill et al., 2017).

In light of the above, we will compare the effect of stimulus type (2D versus 3D face images) on the accuracy of perceptions of strength, weight, and height from face images. We will test three specific hypotheses.

Hypothesis 1. Handgrip strength will be positively and significantly correlated with strength ratings of face images and this correlation will be significantly stronger for ratings of 3D than 2D images.

Hypothesis 2. Height will be positively and significantly correlated with height ratings of face images and this correlation will be significantly stronger for ratings of 3D than 2D images.

Hypothesis 3. BMI will be positively and significantly correlated with weight ratings of face images and this correlation will be significantly stronger for ratings of 3D than 2D images.

Methods

Face stimuli
Stimuli will be 2D and 3D face images of 125 women and 64 men (mean age=24.0 years, SD=8.3 years). These men and women first cleaned their face with hypoallergenic face wipes to remove any make-up. Face photographs were taken a minimum of 15 minutes later against a constant background and under standardized diffuse lighting conditions. The men and women were instructed to pose with a neutral expression. Camera-to-head distance (90 cm) and camera settings were held constant. Six photographs of each individual were taken simultaneously from different angles using a DI3D system (www.di4d.com) with six standard digital cameras (Canon EOS100D with Canon EF 50 mm f/1.8 STM lenses). More information on the image collection procedure, including an example of the collected image data as well as a schematic drawing of the set up can be found at https://osf.io/gs5wm/. The front-view face images as captured by the top middle camera will be used as
the 2D images. 3D images will be generated using DI3Dview (version 6.8.9), which creates both a texture map in the BMP file format (exported at a resolution of 1MP) as well as a 3D mesh from the raw data that will be exported in the Wavefront OBJ file format. Both 2D and 3D images will be Procrustes-aligned prior to rating based on 132 landmarks and 55 landmarks respectively, to remove differences in alignment and size. 3D face images will be rendered with a perspective/field of view equivalent to that of 2D images and presented as short GIFs of size 600x800 pixel in which the face rotates laterally from -40 degrees to +40 degrees (in steps of 2.25°, displayed at 12 frames per second; see https://osf.io/3f6hn/wiki/home/ for examples using pairs of scans of children's faces). 2D and 3D images will be masked so that hairstyle and clothing are not visible and presented on a black background. These images have been collected as part of an ongoing project on 2D and 3D kinship cues. The 189 images we have of each type mean that we have 80% power to detect correlations (r) of .20 with alpha set at .05.

**Body measures**

Height and weight were measured from each participant and used to calculate BMI for each participant. Height was measured using a wall-mounted stadiometer. Weight was measured using a medical-grade seca 761 flat scale. Participants’ handgrip strength was measured from their dominant hand three times using a T. K. K. 5001 Grip A dynamometer. Following Fink et al. (2007) and Han et al. (2018), the highest recording from each participant will be used in analyses. These body measures were taken when the images were obtained.

**Ratings**

2D male faces, 3D male faces, 2D female faces, and 3D female faces will be presented in separate blocks of trials. Trial order in each block will be fully randomized and block order will also be fully randomized. Participants will be randomly allocated to rate the faces for strength, weight, or height using 1 (not very strong/heavy/tall) to 7 (very strong/heavy/tall) scales. Note that each participant will rate the images for one trait only but will rate both the 2D and 3D images for that trait. Twenty raters will rate each trait (i.e., we will test 60 raters in total, approximately half men and half women, all aged between 16 and 40 years of age). Simulations suggest that this number of raters is sufficient to produce high Cronbach’s alphas and stable averages for ratings of most traits (DeBruine & Jones, 2018). Raters will be recruited from the University of Glasgow's Psychology Participant Panel and complete the ratings in the lab on 21.5 inch iMacs.

**Data quality checks**

Data will only be analyzed for a given trait if the Cronbach’s alphas for ratings of 3D and 2D images are both greater than 0.8. Where Cronbach’s alpha is less than 0.8, more raters will be added until Cronbach’s alpha is greater than 0.8.
(with a maximum number of raters per trait of 50). Ratings from participants who give the same score for more than 75% of trials will be removed from the dataset prior to any analyses. Ratings from participants who indicated that they recognized any of the faces will also be removed from the dataset prior to any analyses.

**Analyses**
Handgrip strength, height, and BMI will be analyzed in separate multilevel models (see [https://osf.io/wz5nc/](https://osf.io/wz5nc/) for full analysis code). In each analysis, the relevant rating (strength, height, or weight), averaged across all raters, will be the outcome variable. The corresponding body measurement (handgrip strength, height, and BMI), centered on their means and scaled to range between -5. and +.5, will be included as a predictor, along with sex of face (male or female, effect-coded) and stimulus type (2D or 3D image, effect-coded). The interaction between stimulus type and the relevant body measurement will also be included in the model. The model will also include a by-stimulus random intercept and by-stimulus random slope for stimulus type. For each hypothesis, average ratings more than three standard deviations above or below the mean for the relevant sex will be excluded from the regression analysis. Should any of the tested models run into convergence issues, we will try to resolve these by using different optimizers. As the focus of this study lies in investigating accuracy of judgments across 2D and 3D stimuli and we see no strong rationale to predict sex differences in accuracy of the investigated judgments, we will not analyze effects of rater sex.

**Hypothesis 1. Handgrip strength will be positively and significantly correlated with strength ratings of face images and this correlation will be significantly stronger for ratings of 3D than 2D images.**

Hypothesis 1 will be supported if there is (a) a significant positive effect of handgrip strength on strength ratings and (b) a significant interaction between handgrip strength and stimulus type whereby the effect of handgrip strength is stronger for ratings of 3D than 2D images.

**Hypothesis 2. Height will be positively and significantly correlated with height ratings of face images and this correlation will be significantly stronger for ratings of 3D than 2D images.**

Hypothesis 2 will be supported if there is (a) a significant positive effect of height on height ratings and (b) a significant interaction between height and stimulus type whereby the effect of height is stronger for ratings of 3D than 2D images.
*Hypothesis 3.* BMI will be positively and significantly correlated with weight ratings of face images and this correlation will be significantly stronger for ratings of 3D than 2D images.

Hypothesis 3 will be supported if there is (a) a significant positive effect of BMI on weight ratings and (b) a significant interaction between BMI and stimulus type whereby the effect of BMI is stronger for ratings of 3D than 2D images.

*Robustness checks*
To check that potential ‘carry over’ effects on ratings that could be introduced by having participants rate both 2D and 3D versions of the same faces do not affect results, we will repeat our three main analyses using ratings only from the first block of trials for each rater.

*Conflict of interest*
On behalf of all authors, the corresponding author states that there is no conflict of interest.
References


Society B: Biological Sciences, 276(1656), 575-584.  
https://doi.org/10.1098/rspb.2008.1177

doi: https://doi.org/10.1167/12.12.3.

https://doi.org/10.1111/j.2044-8295.2012.02117.x

https://doi.org/10.3389/fpsyg.2018.02405