Computers in Human Behavior Reports 14 (2024) 100397

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Using computer-generated faces in experimental psychology: The role of realism and exposure

Anna Flavia Di Natale[a,b,**], Stefania La Rocca[b,c], Matilde Ellen Simonetti[d], Emanulea Bricolo[b,c,*]

[a] Department of Psychology, Università Cattolica del Sacro Cuore, Largo Gemelli 1, 20123, Milan, Italy
[b] Mind and Behavior Technological Center, Università Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126, Milan, Italy
[c] Department of Psychology, Università Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126, Milan, Italy
[d] Institute of Psychology, RWTH Aachen University, Jaegerstr. 17/19, D52066, Aachen, Germany

ARTICLE INFO

Keywords:
Computer-generated faces
Face perception
Uncanny valley
Realism
Exposure

ABSTRACT

In psychology, researchers often rely on face photographs to study face perception. However, finding suitable face stimuli for experiments is often challenging. Computer-generated (CG) faces have emerged as a potential solution due to their flexibility and controllability. However, it has been suggested that these stimuli are evaluated and processed differently from real faces, and their suitability as alternatives in experimental settings remains unclear. To address this, two studies were conducted to examine the impact of CG faces' realism and observers' self-reported exposure to CG faces on faces appraisals (Study 1) and processing (Study 2).

In Study 1 (n = 97), we assessed perceptions of both real and CG faces. Findings indicated that participants generally viewed CG faces less favourably, especially when these faces lacked realism. This trend was particularly pronounced among individuals less exposed to digital characters.

In Study 2 (n = 33), we examined the recognition accuracy of these faces in a memory task. The data revealed that CG faces, especially those less realistic, were less accurately recognized. However, this discrepancy was primarily observed among individuals with limited exposure to digital characters, while those more familiar with such characters showed no significant difference in recognition.

Overall, this work confirmed that, to date, CG faces are not an adequate alternative to real faces and that researchers should be cautious when using these stimuli in experiments involving face processing. However, as digital graphics improve and as digital characters become more commonplace in daily life, the perceptual gap between CG and real faces may diminish.

1. Introduction

Faces are essential to many research areas in psychology as they provide valuable cues about the people we interact with. Faces indeed reveal information about age, gender, ethnicity, emotional state, and many other attributes that we can use to guide our behaviour during verbal and non-verbal communication (McKone & Robbins, 2011).

Photographs or video clips portraying faces are commonly used in psychological research paradigms to explore how individuals respond to these stimuli. However, researchers investigating human face perception often find it challenging to retrieve adequate face stimuli that they can use in experimental settings to answer their research questions appropriately. Finding high-quality, standardized pictures or videos that fit different experimental goals may be challenging. Available face databases traditionally provide researchers with images of faces (Gross, 2005). These databases contain photographs of individuals with unique identities but possibly different physical attributes and characteristics, including age and ethnicity. Sometimes models are captured in various poses and display various emotional expressions. However, these databases are often tailored to the researchers' individual needs and do not always meet the needs of other experimenters. Also, images from different databases are difficult to merge as they differ in several technical details and are usually validated on distinct psychological and perceptual dimensions. Finding a way to control stimuli and conduct

* Corresponding author. Department of Psychology, Università Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126, Milan, Italy.
** Corresponding author. Department of Psychology, Università Cattolica del Sacro Cuore, Largo Gemelli 1, 20123, Milan, Italy.
E-mail address: emanuela.bricolo@unimib.it (E. Bricolo).

https://doi.org/10.1016/j.chbr.2024.100397
Received 27 December 2023; Received in revised form 28 February 2024; Accepted 29 February 2024
Available online 6 March 2024

systematic investigations into how faces are perceived may turn out to be an arduous task.

In recent years, the use of computer-generated (CG) faces has increased in psychological experimental settings (Dawel, Miller, Horsburgh, & Ford, 2021). CG faces are generated through computer software from scratch or by converting real pictures into 3D head models. Today, there are numerous graphical software available, such as FaceGen (FaceGen Modeller, 2021), Poser (*Poser*, 2021), and Character Creator (*Character Creator*, 2022). These programs allow researchers to design highly realistic CG faces quickly and modify them to meet the requirements of specific studies by manipulating facial expressions, facial characteristics, and other socio-cognitive aspects (La Rocca, Gobbo, Tosi, Fiora, & Daini, 2023). These programs indeed provide integrated tools that enable researchers to control various aspects of the faces, including physical appearance, facial expression, and morphological features. Additionally, plugins can be integrated to fine-tune technical details such as shading and rendering. Consequently, these stimuli can be rapidly generated and manipulated without requiring advanced technical expertise, making them adaptable to meet the objectives of different research projects.

Thus, CG faces have the potential to serve as versatile and flexible alternatives for conventional pictures. A recent systematic review (Dawel et al., 2021) found that approximately 7.3% of the publications focusing on psychological research and employing faces as stimuli used CG faces in their experiments. The authors highlighted two distinct applications of CG faces in their study. A few studies examined the unique responses elicited by CG faces compared to human faces. Most of these studies, however, used CG faces as substitutes for human faces. Therefore, it is crucial to investigate their appropriateness and suitability as alternatives to real faces in psychological research, by directly verifying any changes in how observers appraise and process these stimuli. Central to this, are two intertwined factors: “realism” and “self-reported exposure".

1.1. Realism

Thanks to the rapid advances in digital graphics, it is now possible to create highly realistic CG faces that have the potential to serve as versatile and flexible alternatives for conventional pictures. We here define *realism* as the accuracy of the details of the physical appearance. Lighting, shading, surface details, and resolution are all included in this category, which describes the degree of visual realism reached by different render details (Grewe, Liu, Kahl, Hildebrandt, & Zachow, 2021; Nowak & Fox, 2018). Therefore, the degree of realism of a CG face is defined by the quality of its technical details and can be manipulated in terms of facial proportions, levels of details, skin appearance, lighting, and shading (Burleigh, Schoenherr, & Lacroix, 2013; Diel, Weigelt, & MacDorman, 2021; Green, MacDorman, Ho, & Vasudevan, 2008; Zell et al., 2015). Depending on the software, level of detail, and rendering techniques, the realism of these faces can vary widely. This variation has implications for how individuals appraise and process these stimuli.

For instance, research suggests that CG faces elicit different emotional responses and first impressions than real faces (Miller, Foo, Mewton, & Dawel, 2023). This difference may be partially attributed to the level of realism in CG faces. The uncanny valley (UV) hypothesis (Mori, 1970) posits that as CG faces become more realistic, our emotional affinity towards them increases until a certain threshold, after which our response becomes negative or eerie. However, most studies investigating individuals’ responses to CG faces failed to find support for the UV hypothesis (Di Natale et al., 2023a). Some studies even propose an alternative, the uncanny slope hypothesis (Kätsyri, de Gelder, & Takala, 2019), suggesting a continuous positive response as realism increases. This ongoing debate underscores the need to fully understand the relationship between CG faces’ realism and human responses. Beyond emotional responses, research has also shown that CG faces tend to have more unfavourable perceptions of social qualities than real ones (Miller et al., 2023).

Another critical aspect to consider is the impact of realism on face processing (e.g., identity recognition). Different studies revealed that CG faces are less efficiently discriminated than real faces (Balas & Pacella, 2015; Crookes et al., 2015; Kätsyri, 2018). One possible explanation relies on the realism of CG faces. These stimuli display lower amounts of discriminating information than real faces (Crookes et al., 2015). Specifically, due to the limitations of existing graphical programs, CG faces lack surface cues (e.g., colour, pigmentation, reflectance, albedo) that help an individual recognize a face. Consequently, the less realistic CG faces are, the more similar they become to each other and the harder they are to remember (Goldstein & Chance, 1980; Valentine, 1991). As a result, the level of realism can impact how individuals perceive and categorize CG faces, impacting how well they can recognize and distinguish them.

1.2. Exposure

Exposure, the second factor, also plays a crucial role in how individuals appraise and process CG faces. In this context, we refer to exposure as the frequency with which individuals encounter specific categories of stimuli, such as faces. It has been observed that self-reported frequency of exposure to certain stimulus types is correlated with increased familiarity and potentially enhanced processing capabilities for these stimuli (Lu, Hua, Huang, Zhou, & Dosher, 2011). It has been suggested that exposure to particular groups of stimuli shapes our expertise with these kinds of stimuli. This process is often called perceptual narrowing (Nelson, 2001). According to perceptual narrowing, the demands of our surrounding environment shape our perceptual abilities. This means that our perceptual system specializes in processing and recognizing the most frequently encountered stimuli. Among these stimuli, we can definitively find human faces. For example, there is evidence of a decrease in the capacity to distinguish between two faces within a category that is infrequently seen, such as other-race faces (e.g., Kelly et al., 2007), other-species (e.g., Scott & Monesson, 2009), and other-age faces (e.g., Kuefner, Macchi Cassia, Picozzi, & Bricolo, 2008).

In daily life, most of us encounter real faces far more frequently than computer-generated (CG) faces. This disparity in exposure has led researchers to speculate that our unfamiliarity with CG faces could influence our reactions to them. For instance, studies have suggested that people may find CG faces ‘uncanny’ or strange due to this lack of familiarity them (Chattopadhyay & MacDorman, 2016; Kätsyri, 2018; Kätsyri et al., 2019). However, increased exposure to CG faces can mitigate these feelings, resulting in a more favourable impression of the character (Burleigh et al., 2013; Cheetham, Suter, & Jäncke, 2011, 2014; Yamada, Kawabe, & Ihaya, 2013). This unfamiliarity may also impact our perceptions of other attributes of CG faces, such as trustworthiness or attractiveness.

Lastly, while there is limited research on how we process CG faces, initial findings suggest that people tend to be less accurate in recognizing CG faces than real ones (Balas & Pacella, 2015; Crookes et al., 2015; Kätsyri, 2018). Crookes et al. (2015) suggested that, beyond CG faces’ realism, a possible explanation also relies on the fact that we are less exposed to CG faces in everyday life, and therefore, we are less exposed to these stimuli. As a result, CG faces may be susceptible to the other-race effect (ORE). According to the ORE, we perceive more quickly and accurately the faces of our ethnic group. One explanation of such an effect is that individuals are less efficient in processing other-race faces since they are less exposed, and thus less trained, with this kind of stimuli (Rhodes, Brake, Taylor, & Tan, 1989).

1.3. Present work

As the use of CG faces in psychological research continues to grow, understanding the intertwined roles of realism and self-reported exposure becomes increasingly crucial. These factors influence how we perceive and process CG faces and have profound implications for the validity and generalizability of research findings. The goal of the present studies is to see whether face appraisals and identity recognition are affected by the realism of the faces and by the self-reported exposure of participants to CG characters.

Study 1 delves into the appraisal of CG faces, focusing on aspects such as the UV and first impressions. We aim to understand how realism and self-reported exposure might influence these evaluations, potentially offering insights into the responses elicited by CG faces compared to real ones.

Study 2 shifts the focus to the processing of CG faces, particularly on face identity recognition. Here, we explore how individuals process and remember CG faces and whether their ability to do so is influenced by the realism of the CG face or their prior experience with such stimuli.

With these studies, we aim to address the following research questions:

- *RQ1: Are there notable differences in how individuals appraise and process real faces compared to CG faces?*
- *RQ2: How does the realism of CG faces affect participants' appraisals and their ability to process these faces?*
- *RQ3: Does prior experience or familiarity with CG faces influence how participants appraise and process them?*

2. Study 1: CG faces appraisals

2.1. Introduction

Faces deliver essential social and emotional information used for verbal and nonverbal communication. Human beings are good at detecting and using such information in social interactions. It has been suggested that, as soon as a face is perceived, humans tend to rapidly generate first impressions based on only a minimum amount of information (Marzi, Righi, Ottonello, Cincotta, & Viggiano, 2014; Todorov & Uleman, 2004; Willis & Todorov, 2006). Specifically, when meeting someone new, individuals tend to form an automatic social evaluation of that person based on facial appearance (Todorov, Mende-Siedlecki, & Dotsch, 2013; Willis & Todorov, 2006; Zebrowitz, 2017). Information deriving from a facial appearance is used to make several other social evaluations, including, for example, estimates of attractiveness, trustworthiness, aggressiveness, and dominance (Bar, Neta, & Linz, 2006; Mende-Siedlecki, Said, & Todorov, 2013; Olson & Marshuetz, 2005).

In previous works, Balas and colleagues (Balas, Tupa, & Pacella, 2018) examined whether CG faces' judgments are evaluated within the same social face space identified for real faces (Oosterhof & Todorov, 2008). According to the face space model faces are evaluated on two main dimensions: valence and dominance. In their study, Balas and colleagues (Balas et al., 2018) found that CG and real faces were assessed within the same two-factor model (valence-dominance) but further revealed that CG faces were positioned differently within that face space. According to the authors, this result suggests that although observers apply the same strategy to judge CG and real faces, CG faces' appraisals may be disrupted. This is in line with studies showing differences in CG and real faces' evaluations of, for example, likeability, trustworthiness, and attractiveness, finding CG faces to be rated more negatively (Balas & Pacella, 2017; Schindler, Zell, Botsch, & Kissler, 2017; Zell et al., 2015).

The first aim of our study was to confirm whether real faces are evaluated differently from CG faces. To this end, we collected participants' assessments of real and CG faces. Within the study, we considered some psychologically-relevant features of face perception that span different aspects: human-likeness, eeriness, likability, attractiveness, and trustworthiness. We selected only six dimensions to ensure the task was not too repetitive. However, we specifically decided to measure human-likeness, likability, attractiveness, and trustworthiness as they are the most widely explored traits in the current literature but have never been measured together on the same images. Specifically, we hypothesized that:

H1. Real faces are rated more human-like, likeable, attractive, trustworthy and less eerie than CG faces.

A second goal was to explore the specific role of CG faces' realism in face appraisals. Given that social evaluations are generated very fast based on facial appearance, it is plausible that observers' appraisals of CG faces largely rely on their realism and the resulting perception of human-likeness. According to the more recent uncanny slope hypothesis (Kätsyri et al., 2019), when a CG face looks more realistic, it can be perceived as more familiar and more likely to elicit a positive response from the viewer. We, therefore, presented participants with two different degrees of realism. Realism was manipulated by modifying the texture resolution of the virtual characters (Diel et al., 2021). We included two CG variants of the original real faces representing a CG face with low realism (CG-low) and a CG face with high realism (CG-high). We predicted that:

H2. The level of realism in CG faces influence observers' appraisals of those stimuli, with CG-high faces being rated as more human-like, likeable, attractive, trustworthy, and less eerie compared to CG-low faces.

Finally, our tertiary goal was to investigate the role of self-reported exposure to CG faces on their appraisals. It has been suggested that CG faces might be differently evaluated depending on individuals' self-reported exposure to these entities (Chattopadhyay & MacDorman, 2016; Kätsyri et al., 2019). Specifically, it has been argued that, compared to real faces, CG faces are less familiar to human observers since they are less exposed to these stimuli in everyday life. As a result, individuals' judgments of CG faces might differ from those of real faces. However, to the best of our knowledge, no one has so far directly investigated the influence of self-reported exposure on face social evaluations. To this end, we asked participants to report how often they are exposed to or interact with realistic virtual characters. We then used this measure to examine how people's responses to virtual characters' faces differed depending on how frequently they were exposed to these characters. Specifically, we hypothesized that:

H3. Self-reported exposure to CG faces influence CG faces 'appraisals, with individuals reporting higher exposure rating both CG-high and CG-low faces as more human-like, likeable, attractive, trustworthy, and less eerie than individuals reporting lower exposure. However, this self-reported exposure does not influence real faces' appraisals.

2.2. Methods

2.2.1. Materials

In this study, we used 120 photographs of Caucasian females ($n = 61$) and males ($n = 59$) young adult models retrieved from the Chicago Face Database (CFD; Ma, Correll, & Wittenbrink, 2015). The photographs were chosen among those that did not present hairs on the face (e.g., fringe or tuft). We then used these pictures to generate two CG variants for each face using Character Creator 3 software (*Character Creator*, 2022). Specifically, the two CG variants were created with different degree of realism in terms of texture resolution (Diel et al., 2021). To do so, we used two different tools available in the software. We used the Auto Mode tool to generate CG faces with low-resolution (1K) textures (*CG-low* faces), and the Pro Mode, to generate high-resolution (4K) textures CG faces (*CG-high* faces). In both steps, we used image matching and sculpt morph tools to match the CG faces as much as possible with the original faces. All faces were cropped to conceal hairs and ears and resized using an image editor program. The final pictures measured 123x123 pixels with 123 dpi. Luminance means, and standard deviations were standardized across pictures using SHINE_color toolbox in Matlab (Dal Ben, 2021). Fig. 1 shows an example of the resulting stimuli.

Fig. 1. Example of the stimuli used in the experiment.
The cropped and luminance matched real face (center), low-resolution CG face (left) and the high-resolution CG face (right).

Due to copyright restrictions, we were unable to publicly share all of the experiment stimuli on an open science platform. However, we would like to inform readers that the original real faces used in our study are accessible on the Chicago Face Database website (https://www.chicagofaces.org), while the modified versions specifically created for our research can be found in the Additional Resources section of the same website (https://www.chicagofaces.org/resources/, Di Natale, La Rocca, Simonetti, & Bricolo, 2023b). We encourage interested readers to visit these sources for further reference.

2.2.2. Participants

Based on previous studies (Langner et al., 2010), we aimed to have about twenty participants rate each stimulus. Therefore 139 participants were invited to our study, but only 115 participants (females = 78; mean age = 26.5 ± 6.5) completed the task. All participants were informed about the characteristics of the study, and before the start of the experiment, they were required to provide informed consent. The study was approved by the local ethical committee of the University of Milano-Bicocca (Protocol RM-2021-407). After giving their informed written consent, participants were asked to share socio-demographic information, including age and gender. Second, they were asked to evaluate how frequently they are exposed to realistic CG characters (e.g., in video games, digital media or other contexts). Participants were presented with an example of a realistic CG character unrelated to the stimuli used in the experiment. They were asked to rate how often they interacted or were exposed to these kinds of stimuli on a 5-point Likert scale where 1 = Never, 2 = Less than once a month, 3 = Less than once a week, 4 = More than once a week, and 5 = Daily. We then divided participants into two groups: participants reporting limited exposure (limited-exposure participants, or LEPs) and participants reporting high-exposure (high-exposure participants, or HEPs). LEPs were considered participants answering either 1 (never) or 2 (less than once a month), while HEPs were considered those who answered either 4 (more than once a week) or 5 (daily). Out of 115 participants, 55 were *LEPs*, and 42 were HEPs. Analyses were then performed on a total of 97 participants.

2.2.3. Procedure

The study was programmed and conducted using Qualtrics on an MSI P75 Creator 9SE with 1920 x 1080 resolution. All faces were presented in color and cropped for the experiment (see 2.2.1. Materials section for details). At the viewing distance of approximately 50 cm, stimuli subtended a visual angle of approximately 6.3° × 8.57°. Given the high number of stimuli to be rated, they were divided into six blocks of 60 faces each. Stimuli were pseudo-randomized into each subset, made of 60 different identities. Twenty pictures were real, 20 were CG-high, and 20 were CG-low faces. Within each block, pictures were randomly presented. Each participant was presented with only one block and therefore was presented with 60 faces (20 real faces, 20 CG-high faces and 20 CG-low faces) in random order. For each face, participants were then asked to answer six questions that were presented in random order. Participants were asked to evaluate faces using 7-point semantic differential scales measuring human-likeness (artificial-human; original Italian anchors: “artificiale-umano”), eeriness (eerie-reassuring; original Italian anchors: “inquietante-rassicurante”), likability (unpleasant-pleasant; original Italian anchors: “sgradevole-piacevole”), attractiveness (unattractive-attractive; original Italian anchors: “non attraente-attraente”), and trustworthiness (untrustworthy-trustworthy; original Italian anchors: “inaffidabile-affidabile”). Furthermore, we asked participants to rate CG faces perceived gender to identify negligent respondents. This variable was measured by asking participants “How feminine/masculine do you think the face in the photo is?” using a 7-points semantic differential scale (feminine-masculine; original Italian anchors: “femminile-mascolino”). The rationale behind this approach was that a consistent failure to appropriately capture the gender of the faces, or random responses across this scale, would indicate a lack of attention or engagement with the task, thus signalling a negligent respondent. To operationalize this, we focused on responses where the “extreme scores” for the gender that was clearly represented by the CG face were not selected. For instance, if a CG face was evidently male, any rating other than 6 or 7 (the extreme scores on the masculine end of the scale) was flagged for further review. This method was based on the assumption that attentive and engaged respondents would likely choose these extreme scores for a clearly gendered face. We then looked for patterns where the same respondents consistently failed to select the extreme scores for the correct gender. A threshold was set at 20% - if a respondent failed to choose the extreme scores in more than 20% of such cases, they were considered to potentially exhibit a negligent pattern. We did not identify any respondents who demonstrated patterns of responses indicative of negligence.

2.2.4. Statistical analyses

A series of linear mixed models (LMMs) were estimated to assess the effects of *face type* (*real faces, CG-high faces, CG-low faces*) and *self-reported exposure* (*low, high*) on face appraisals (*human-likeness, eeriness, likeability, attractiveness, and trustworthiness*) using the lmer test in R (Kuznetsova, Brockhoff, & Christensen, 2015).

For each model with a significant main effect or interaction, planned contrasts were implemented using the *emmeans* package (Lenth & Lenth, 2018). Regression coefficients, standard errors, t values, and corresponding p values are reported for these planned contrasts. Specifically, where a main effect of type was detected, planned contrasts between real faces’ appraisals and CG-high and CG-low faces were performed to test H1. Furthermore, planned contrasts between CG-high and CG-low faces were performed to test H2. Finally, where a significant interaction was identified, planned contrasts between LEPs and HEPs at each “face type” level was performed to test H3. The database and the script are available at Open Science Framework (https://osf.io/nz9j3/?view_only=ee446255e5504650b6e0312354a523c9, DB_Study_1).

2.3. Results

Table 1 shows the estimated mean marginals and standard errors for

Table 1
Estimated mean marginals (study 1).

	Exposure	Face type		
		Real	CG-high	CG-low
Human-likeness	LEPs	5.99 (0.09)	2.48 (0.09)	1.88 (0.09)
	HEPs	6.01 (0.10)	3.09 (0.09)	2.31 (0.10)
Eeriness	LEPs	4.03 (0.09)	4.51 (0.09)	4.81 (0.09)
	HEPs	3.80 (0.10)	4.05 (0.10	4.30 (0.10)
Likeability	LEPs	4.12 (0.09)	3.50 (0.08)	3.30 (0.08)
	HEPs	4.16 (0.10)	3.80 (0.10)	3.46 (0.10)
Attractiveness	LEPs	4.12 (0.08)	3.50 (0.08)	3.30 (0.08)
	HEPs	4.16 (0.10)	3.80 (0.10)	3.46 (0.10)
Trustworthiness	LEPs	4.22 (0.08)	3.76 (0.08)	3.58 (0.08)
	HEPs	4.35 (0.10)	4.00 (0.10)	3.80 (0.10)

the considered dimensions, including human-likeness, eeriness, likeability, attractiveness, and trustworthiness, across different conditions. The conditions include evaluations by LEPs and HEPs, as well as evaluations of real faces and two variations of CG faces: high realism (CG-high) and low realism (CG-low). The mean values and their standard errors are provided for each condition, allowing for a comparison of participants' ratings across different factors. The figures representing the results is provided in the Supplementary Materials.

2.3.1. Human-likeness

The LMM performed on "human-likeness" (artificial-human) evaluations explained about the 67% of the variance (conditional $R^2 = 0.67$). Within the model we found a significant main effect of "face type" [F (2,5719) = 5094.66, $p < 0.001$]. Planned contrasts showed that "human-likeness" evaluations differed significantly between real faces and both CG-low faces and CG-high faces, confirming H1. Planned contrasts between CG-low faces and CG-high faces revealed an effect of realism, revealing a significant difference in "human-likeness" evaluations between CG-low and CG-high faces, confirming H2. Planned contrasts are shown in Table 2. The analysis further revealed a significant main effect of "self-reported exposure" [F(1,95) = 7.41, $p < 0.01$] and a significant interaction between "face type" and "self-reported exposure" [F(2, 5719) = 26.07, $p < 0.001$]. Planned contrasts revealed a significant difference in "human-likeness" evaluations between HEPs and LEPs in both CG-low and CG-high faces. As expected, the effect of self-reported exposure was not significant on real faces' human-likeness evaluations. These results confirmed H3. Planned contrasts are shown in Table 3.

2.3.2. Eeriness

The LMM performed on "eeriness" (reassuring-eerie) evaluations explained about the 30% of the variance (conditional $R^2 = 0.30$). Within the model we found a significant main effect of "face type" [F(2,5719) = 108.97, $p < 0.001$]. Planned contrasts revealed that "eeriness" evaluations differed significantly between real faces and both CG-low faces and CG-high faces, confirming hypothesis H1. Planned contrasts further revealed an effect of CG faces' realism, showing a significant difference in "eeriness" evaluations between CG-low and CG-high faces, confirming H2. Planned contrasts are shown in Table 2. The analysis further revealed a significant main effect of "self-reported exposure" [F(1,95) = 9.81, $p < 0.01$] and a significant interaction between "face type" and "self-reported exposure" [F(2, 5719) = 5.63, $p < 0.01$]. Planned contrasts showed a significant difference in "eeriness" evaluations between LEPs and HEPs in CG-low and CG-high faces. As expected, we did not find a significant effect of self-reported exposure on real faces' "eeriness" evaluations. These results confirmed H3. Planned contrasts are shown in Table 3.

2.3.3. Likeability

The LMM performed on "likeability" (unpleasant-pleasant) evaluations explained about the 18% of the variance (conditional $R^2 = 0.18$). Within the model we found a significant main effect of "face type" [F (2,5719) = 157.19, $p < 0.001$]. Planned contrasts revealed that "likeability" evaluations differed significantly between real faces and both CG-low faces and CG-high faces, confirming hypothesis H1. Planned contrasts further revealed an effect of CG faces' realism, showing a significant difference in "likeability" evaluations between CG-low and CG-high faces, confirming H2. Planned contrasts are shown in Table 2. The analysis further revealed a significant interaction between "face type" and "self-reported exposure" [F(2, 5719) = 4.61, $p < 0.05$]. Planned contrasts showed a significant difference in "likeability" evaluations between LEPs and HEPs only in CG-high faces, but not on CG-low faces. The effect of self-reported exposure was not significant on real faces' "likeability" evaluations. These results partially confirmed H3. Planned contrasts are shown in Table 3.

2.3.4. Attractiveness

The LMM performed on "attractiveness" (unattractive-attractive) evaluations explained about the 16% of the variance (conditional $R^2 = 0.16$). Within the model we found a significant main effect of "face type" [F(2,5719) = 72.35, $p < 0.001$]. Planned contrasts revealed that "attractiveness" evaluations differed significantly between real faces and both CG-low faces and CG-high ì faces, confirming hypothesis H1. Planned contrasts further revealed an effect of CG faces' realism, showing a significant difference in "attractiveness" evaluations between CG-low and CG-high faces, confirming H2. Planned contrasts are shown in Table 2. The analysis further revealed a significant main effect of "self-reported exposure" [F(1,95) = 17.13, $p < 0.001$] and a significant interaction between "face type" and "self-reported exposure" [F(2, 5719) = 4.21, $p < 0.05$]. Planned contrasts revealed a significant difference in "attractiveness" evaluations between LEPs and HEPs on both CG-low faces and CG-high faces. However, we further found a significant effect of self-reported exposure on real faces' "attractiveness" evaluations. These results partially confirmed H3. Planned contrasts are shown in Table 3.

2.3.5. Trustworthiness

The LMM performed on "trustworthiness" (untrustworthy-trustworthy) evaluations explained about the 18% of the variance (conditional $R^2 = 0.18$). Within the model we found a significant main effect of

Table 2
Planned contrasts of "face type" (Study 1).

	Hp	Level 1	Level 2	β	95% CI		SE	df	t
					CI_low	CI_high			
HL	1	Real	CG-low	3.91	3.81	4.01	0.04	5719	94.57***
		Real	CG-high	3.22	3.12	3.32	0.04	5719	77.85***
	2	CG-low	CG-high	−0.69	−0.79	−0.59	0.04	5719	−16.72***
EE	1	Real	CG-low	−0.64	−0.75	−0.54	0.04	5719	−14.72***
		Real	CG-high	−0.37	−0.47	−0.26	0.04	5719	−8.36***
	2	CG-low	CG-high	0.28	0.17	0.38	0.04	5719	6.35***
LK	1	Real	CG-low	0.76	0.66	0.87	0.04	5719	17.49***
		Real	CG-high	0.49	0.39	0.60	0.04	5719	11.27***
	2	CG-low	CG-high	−0.27	−0.38	−0.17	0.04	5719	−6.22***
AT	1	Real	CG-low	0.57	0.45	0.68	0.05	5719	12.00***
		Real	CG-high	0.25	0.14	0.36	0.05	5719	5.26***
	2	CG-low	CG-high	−0.32	−0.43	−0.20	0.05	5719	−6.74***
TR	1	Real	CG-low	0.60	0.49	0.70	0.04	5719	14.05***
		Real	CG-high	0.41	0.31	0.51	0.04	5719	9.64***
	2	CG-low	CG-high	−0.19	−0.29	−0.09	0.04	5719	−4.41***

Note that *** < 0.001; ** < 0.01; * < 0.05.

Table 3
Planned contrasts of "self-reported exposure" at each level of "face type" (Study 1).

	Hp	Level 1	Level 2	Type	β	95% CI		SE	df	t
						CI_low	CI_high			
HL	3	HEPs	LEPs	real	0.02	−0.25	0.30	0.14	122.48	0.17
		HEPs	LEPs	CG-high	0.61	0.33	0.88	0.14	122.48	4.38***
		HEPs	LEPs	CG-low	0.43	0.16	0.70	0.14	122.48	3.11**
EE	3	HEPs	LEPs	real	−0.24	−0.51	0.04	0.14	126.48	−1.70
		HEPs	LEPs	CG-high	−0.47	−0.74	−0.19	0.14	126.48	−3.37***
		HEPs	LEPs	CG-low	−0.51	−0.78	−0.23	0.14	126.48	−3.67***
LK	3	HEPs	LEPs	real	0.03	−0.22	0.29	0.13	132.16	0.25
		HEPs	LEPs	CG-high	0.30	0.04	0.55	0.13	132.16	2.30*
		HEPs	LEPs	CG-low	0.16	−0.10	0.42	0.13	132.16	1.24
AT	3	HEPs	LEPs	real	0.34	0.08	0.60	0.13	122.48	2.57*
		HEPs	LEPs	CG-high	0.58	0.32	0.84	0.13	122.48	4.40***
		HEPs	LEPs	CG-low	0.57	0.31	0.83	0.13	138.26	4.33***
TR	3	HEPs	LEPs	real	0.21	−0.04	0.47	0.13	129.56	1.65
		HEPs	LEPs	CG-high	0.24	−0.01	0.50	0.13	129.56	1.89
		HEPs	LEPs	CG-low	0.13	−0.13	0.39	0.13	129.56	1.01

Note that *** < 0.001; ** < 0.01; * < 0.05.

"face type" [$F(2,5719) = 103.3$, $p < 0.001$]. Planned contrasts revealed that "trustworthiness" evaluations differed significantly between real faces and both CG-low faces and CG-high faces, confirming hypothesis H1. Planned contrasts further revealed an effect of CG faces' realism, showing a significant difference in "trustworthiness" evaluations between CG-low and CG-high faces, confirming H2. Planned contrasts are shown in Table 2. The analysis did not reveal any other significant main effect or interaction.

2.4. Discussion

In this study, we compared real and CG faces' evaluations by further analyzing the specific effect of CG faces' realism and observers' self-reported exposure to CG characters.

We first examined the effect of "face type" by comparing participants' perceived human-likeness of real faces and CG-high and CG-low faces. The analysis revealed a significant effect of the realism manipulation on perceived human-likeness, showing that perceived human-likeness increased with increasing realism. Planned contrasts revealed that both CG faces were judged much less human-like than real faces. This is in line with previous literature showing that CG and real faces are perceived categorically rather than continuously (Cheetham et al., 2011; Looser & Wheatley, 2010). According to this view, CG faces are judged artificial until they become sufficiently realistic to cross the category boundary to move into the human category. Our results show that even highly realistic CG faces are still judged significantly less human-like than real faces and may fall into the virtual category side of the continuum. Similar findings were found on the evaluations of the facial traits considered in this study (i.e., eeriness, likeability, attractiveness, and trustworthiness), with CG faces receiving, overall, more negative evaluations than real faces and CG-low faces receiving more negative evaluations than CG-high faces. These findings confirmed our first hypothesis (H1) that real faces were rated more human-like, likeable, attractive, trustworthy, and less eerie than both CG-high and CG-low faces. This supports the notion that people tend to perceive real faces more positively than CG faces, as shown in previous literature (Balas & Pacella, 2017; Schindler et al., 2017; Zell et al., 2015). Taken together these findings suggest that researchers should be aware of the limits that CG faces, even when highly realistic, might have before they can be considered proxies for human faces.

Furthermore, it has been hypothesized that the degree of realism with which CG faces are created is a crucial determinant of how observers evaluate these stimuli. Whether increasing realism is beneficial or not for CG faces' evaluation remains unclear. However, most studies have found a positive correlation between face evaluations and realism, with more realism leading to more favourable ratings (Cheetham, Suter, & Jancke, 2014; Kätsyri et al., 2019). Our results confirmed this prediction, supporting our second hypothesis (H2), as individuals tended to perceive more realistic CG faces as more human-like, leading to a greater affinity and more favourable ratings over less realistic CG faces. Specifically, CG-high faces were rated more human-like, likeable, attractive, trustworthy, and less eerie than CG-low faces. This result suggests that increasing CG faces' realism may reduce the discrepancies between CG and real faces. This is in line with the uncanny slope hypothesis (Kätsyri et al., 2019), which suggests that increased realism in CG characters can lead to a more positive response from viewers. Our findings indicate that the realism of CG faces plays a critical role in their evaluations, providing valuable insights for developing more realistic and socially acceptable virtual characters.

Our third hypothesis (H3) was also confirmed, showing that self-reported exposure affected CG faces' evaluations. Our results showed that individuals more frequently exposed to digital characters evaluate CG faces more positively than people reporting being less exposed. Specifically, these individuals rated CG-high and CG-low faces as more human-like, likeable, attractive, trustworthy, and less eerie than LEPs.

Contrary to our expectations, self-reported exposure also influenced real faces' attractiveness evaluations. This could be due to HEPs having a more refined perception of attractiveness based on their exposure to a wider variety of facial features, including those found in CG faces. This result could be explained by the expertise hypothesis (Diamond & Carey, 1986). According to this hypothesis, face experts see attractiveness in all faces as an indication of a broader appreciation of faces. Being more exposed to CG faces widens the range and the number of faces individuals are exposed to, and therefore they become more familiar with faces in general which in turn affects faces perceptions (Carr, Brady, & Winkielman, 2017; Claypool, Hugenberg, Housley, & Mackie, 2007; Yan, Young, & Andrews, 2017). Further research is needed to clarify this point.

3. Study 2: CG faces identity recognition

3.1. Introduction

Faces represent a special kind of visual stimuli (Farah, Wilson, Drain, & Tanaka, 1998). In particular, it has been argued that faces are processed differently from other stimuli, for example, objects. According to the holistic hypothesis (Maurer, Le Grand, & Mondloch, 2002), individual facial structures are simultaneously encoded and integrated into a single global perception. One of the most important pieces of evidence in support of the holistic hypothesis, together with the composite (Young, Hellawell, & Hay, 1987) and the part-whole (Tanaka & Farah, 1993) effects, is the inversion effect. According to the inversion effect, face recognition is slower and less accurate when the face is presented upside down (Yin, 1969). This effect is traditionally interpreted as indicative of holistic processing (but see Gerlach, Kühn, Mathiassen, Kristensen, & Starrfelt, 2023; Murphy, Gray, & Cook, 2020). While this effect has been extensively observed in studies with real faces (e.g., Taubert, Apthorp, Aagten-Murphy, & Alais, 2011; Turati, Sangrigoli, Ruely, & de Schonen, 2004; Valentine, 1988; Yovel & Kanwisher, 2005), little is known about whether it applies also to CG faces. To the best of our knowledge, only three studies have analyzed the inversion effect for both real and CG faces in recognition tasks (Balas & Pacella, 2015; Kätsyri, 2018; Matheson & McMullen, 2011). All these results revealed that CG faces are susceptible to the inversion effect and, therefore, suggest that CG faces are processed as face stimuli, in a holistic manner.

However, further results revealed that, although CG faces are processed as face stimuli, they might be processed less efficiently than real faces. In particular, different results were obtained when CG faces' recognition accuracy was measured, specifically in the context of the other-race effect (ORE). According to the ORE, we perceive more quickly and accurately the faces of our ethnic group than others (O'toole, Deffenbacher, Valentin, & Abdi, 1994). One possible explanation of such effect is that we have an innate predisposition to processing stimuli belonging to conspecifics. However, the most supported explanation for this effect is that people are more accurate in recognizing the faces of their ethnic group since they are more exposed to that type of stimuli throughout their lives (Tanaka, Kiefer, & Bukach, 2004). The effect of expertise has indeed been shown also with other categories of stimuli, including faces of different ages (Kuefner et al., 2008). Crookes et al. (2015) investigated whether the expertise acquired with processing real faces is applied to processing CG faces. The authors asked participants to complete discrimination and face recognition tasks. In both experiments, participants (Caucasian and Asian) were presented with real and CG faces of both ethnicities (Caucasian and Asian). The authors found that accuracy in recognizing faces of one's ethnicity was drastically reduced for CG faces compared to real faces and that the ORE was substantially attenuated for CG faces. According to the authors, taken together these results demonstrate that our perceptual expertise gained with real faces do not entirely transfer to CG faces that, in turn, are less accurately recognized. Other studies have hypothesized that artificial faces themselves generate the ORE (Balas & Pacella, 2015; Kätsyri, 2018). Balas and Pacella (2015) analyzed participants' performance in a recognition and a face-matching task performed with real faces and their CG counterparts, presented in upright or inverted orientation. The authors observed that performances in both tasks were poorer for CG faces than real faces, meaning CG faces were harder to remember. Similarly, Kätsyri (2018) presented participants with real and CG faces in both upright and inverted orientations and asked them to complete a recognition task. Results showed a greater tendency to respond "seen before" incorrectly (false alarms) for CG faces than for real faces. This result is consistent with the ORE and the author concluded that individuals demonstrate a greater difficulty of processing CG faces than real faces.

These latter studies suggest that although CG faces are processed in a face-like manner, as indicated by the inversion effect, they are processed less efficiently. For this reason, the first hypothesis of this study was:

H1. CG faces are more difficult to recognize than real faces.

Crookes et al. (2015) suggested three possible explanations for this effect.

First, according to the authors (Crookes et al., 2015) identity recognition tasks performed with CG might be more difficult as they contain less perceptual information, such as texture details, that are normally used to discriminate a face. Based on this premises we formulated the following hypothesis:

H2. CG faces with a lower texture resolution (CG-low) are harder to remember than CG faces with a higher texture resolution (CG-high).

Second, the result of their study (Crookes et al., 2015) showed that CG faces were less efficiently processed than real faces, and all participants reported having minimal exposure to CG faces. They therefore postulated that our perceptual system may not be tuned to CG faces as to real ones, given that individuals are less exposed to CG faces than real faces in everyday life. Following this premises, we formulated an additional hypothesis:

H3. *Participants that are less exposed to CG characters, find CG faces harder to remember than participants with more experience.*

Lastly, Crookes et al. (2015) proposed that the lower recognition of computer-generated (CG) faces could be attributed to the UV effect, according to which highly realistic human-like entities evoke discomfort and aversion responses. This unfavourable reaction would lead observers to categorize CG faces as an out-group, such as "non-human". This effect could lead observers to perceive all CG faces as similar and, therefore, harder to differentiate and recognize individually. As a result, CG faces may be more likely to be categorized as a single out-group rather than as a set of diverse individuals, further contributing to their lower recognition rates. While this study did not directly investigate this hypothesis, it is noteworthy that we utilized the same stimuli as those employed in Study 1, which included measurements of the UV variables. Therefore, we here also discussed the implications of this hypothesis.

3.2. Methods

3.2.1. Materials

Face stimuli were the same as those used in Study 1. Similarly, faces were randomly divided into six blocks of 60 faces each (20 real faces, 20 CG-high faces, 20 CG-low faces).

3.2.2. Participants

As an approximation of the present analysis using linear mixed-effects models, the sample size of this study was calculated a-priori with G*Power and it was found that 28 participants were needed to find a medium effect size ($f = 0.25$) with 80% power in a repeated measures within/between design (number of groups 2; number of measurements: 3; correlation between repeated measurements: 0.5). Considering possible issues related to missing data or incomplete experimental sessions, we invited more participants than the minimum required to ensure enough data to maintain statistical power even if some participants did not complete the study or if technical problems lead to incomplete data collection. As a result, a total of 39 participants were invited.

All participants were informed about the characteristics of the study and before the experiments they were required to provide informed consent. The study was approved by the local ethical committee of the University of Milano-Bicocca (Protocol RM-2021-474). Each participant completed a short questionnaire about their experience with artificial faces in different contexts (e.g., video games and digital media), as in Study 1. Out of 39 participants, 24 were LEPs and 9 were HEPs. Analyses were then performed on a total of 33 participants, a sample size consistent with similar studies in the field (e.g., Crookes et al., 2015).

3.2.3. Procedure

Participants completed three face recognition tasks, separately for real faces, CG-high faces and CG-low faces. The three tasks were separated by a short break and administered randomly across subjects. The procedure was administered following Crookes and colleagues' paradigm (Crookes et al., 2015). Specifically, each task consisted of a learning phase and a testing phase. During the learning phase, 10 faces (5 females, 5 males) were presented for 3000 ms in random order. Each face was preceded by a fixation cross in the center of the screen for 250 ms. After the face had disappeared, participants were asked to identify the gender of the face. Once participants had responded, a blank screen appeared for 500 ms and a second face stimulus appeared. The same 10 faces were then presented a second time in a different random order to complete the learning phase. In the testing phase, participants were presented with 20 faces in random order. Ten were faces previously presented during the learning phase, while ten were new faces. Participants were asked to indicate whether they had already seen the face in the learning phase as quickly and accurately as possible using the keyboard. Stimuli remained on the screen until response, or up to a maximum of 5000 ms. The procedure is shown in Fig. 2. All faces were presented in color and cropped in the experiment.

3.2.4. Statistical analyses

To measure accuracy, the d' index was calculated according to the standard formula of signal detection theory d' = z(hits)-z(false alarms) and used to measure accuracy. Hits are defined as the correct response to the presentation of the previously seen stimulus and false alarms as the "already seen" response to faces not previously presented in the learning phase. To compare recognition accuracy between LEPs and HEPs with CG-low, CG-high and real faces, we run a mixed model with realism (CG-low faces, CG-high faces and real faces) and self-reported exposure (low, high) as predictors of accuracy (d'). We also included participants' intercept as a random effect. As in Study 1, if the model exhibited a significant main effects or interactions, planned contrasts were conducted using the emmeans package (Lenth & Lenth, 2018). Specifically, if a main effect related to the type of face was observed, planned contrasts were performed between evaluations of real faces and CG-high or CG-low faces to test H1. Moreover, planned contrasts were conducted between CG-high and CG-low faces to examine H2. Lastly, in cases where a significant interaction was detected, planned contrasts were carried out between LEPs and HEPs at each level of "face type," serving to test H3. The database and the script are available at Open Science Framework (https://osf.io/nz9j3/?view_only=ee446255e5504650b6e0312354a523c9, DB_Study_2).

3.3. Results

Table 4 shows the estimated mean marginals and standard errors of accuracy (d') as a function of realism and self-reported exposure.

The LMM performed on "accuracy" (d') explained about the 62% of the variance (conditional $R^2 = 0.62$). Within the model we found a significant main effect of "face type" [$F(2,62) = 18.56$, $p < 0.001$]. The figure representing the results is provided in the Supplementary Materials. Planned contrasts revealed that "accuracy" (d') differed significantly between real faces and both CG-low faces and CG-high faces, confirming hypothesis H1. Planned contrasts further revealed an effect of CG faces' realism, showing a significant difference in "accuracy" scores between CG-low and CG-high faces, confirming H2. Planned contrasts are shown in Table 5. The figure representing the result is shown in the Supplementary Materials section.

The analysis further revealed a significant interaction between "face type" and "self-reported exposure" [$F(2, 62) = 3.54$, $p < 0.05$]. Planned contrasts revealed a significant difference in "accuracy" scores between LEPs and HEPs on CG-low faces. These results partially confirmed H3. Planned contrasts are shown in Table 6.

Table 4
Estimated mean marginals (study 2).

Exposure	Face type		
	Real	CG-high	CG-low
LEPs	2.31 (0.15)	1.53 (0.15)	1.04 (0.15)
HEPs	2.17 (0.24)	1.83 (0.24)	1.69 (0.24)

3.4. Discussion

The present study investigated the relationship between individuals' self-reported exposure to CG faces and recognition accuracy on a face memory task. To do that we measured the frequency of exposure to or interaction with these stimuli in everyday contexts (e.g., video games, social media, movies, professional experience, etc.). The main objective was to establish how self-reported exposure impacts participants' identity recognition abilities with CG faces. We also investigated what effect different degrees of realism of CG faces might have had on recognition accuracy. We, therefore, manipulated CG faces' realism by varying their texture resolution (Burleigh et al., 2013) on two levels. We generated our CG faces using Character Creator 3 (*Character Creator*, 2022). This choice was justified because this program allowed us to easily manipulate texture resolution on the same face.

The first result of this study revealed that, in general, recognition accuracy is reduced for CG faces compared to real faces. We indeed found a significant main effect of face type (CG-low faces, CG-high faces, real faces) accuracy scores. These results confirm the first two hypotheses, suggesting that CG faces are indeed harder to remember than real faces (H1) and are in line with previous research investigating identity recognition skills with CG and real faces (Balas & Pacella, 2015; Crookes et al., 2015; Kätsyri, 2018).

Unlike previous studies, we further investigated the effect of the realism of CG faces and of participants' self-reported exposure CG characters on recognition accuracy. Our results are discussed following the explanations proposed by Crookes et al. (2015).

First, the authors (Crookes et al., 2015) suggested that lower accuracy scores for CG faces may be because, unlike real faces, CG faces contain less perceptual information that can be used to discriminate a face. They suggested that CG faces have less detailed texture than real faces and thus lack a number of features and details that are used in face recognition. Our results found support for this hypothesis. Specifically, we found a significant difference between real faces, CG-high and

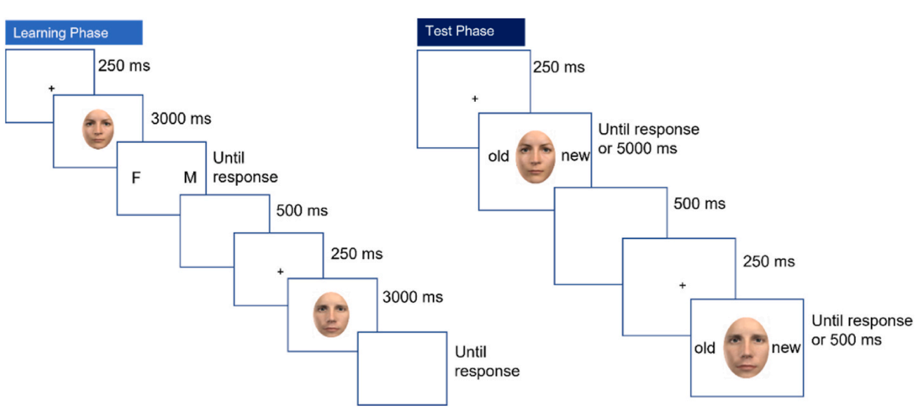

Fig. 2. The procedure for the face recognition task.
Participants first completed a learning phase (left) and then a test phase (right). In both phases, stimuli were presented in the center of the screen. Viewing distance was approximately 50 cm and stimuli subtended a visual angle of approximately 6.3° × 8.57°.

Table 5
Planned contrasts of "face type" (Study 2).

Hp	Level 1	Level 2	β	95% CI		SE	df	t
				CI_low	CI_high			
1	Real	CG-low	−0.31	−0.67	0.04	0.14	62	−2.16***
	Real	CG-high	−0.87	−1.23	−0.51	0.14	62	−5.97***
2	CG-low	CG-high	−0.55	−0.91	−0.19	0.14	62	−3.81*

Note that *** < 0.001; ** < 0.01; * < 0.05.

Table 6
Planned contrasts of "self-reported exposure" at each level of "face type" (Study 2).

Hp	Level 1	Level 2	Type	β	95% CI		SE	df	t
					CI_low	CI_high			
3	HEPs	LEPs	real	−0.13	−0.69	0.42	0.28	65.16	−0.47
	HEPs	LEPs	CG-high	0.30	−0.26	0.86	0.28	65.16	1.07
	HEPs	LEPs	CG-low	0.65	0.08	1.21	0.28	65.16	2.30*

Note that *** < 0.001; ** < 0.01; * < 0.05.

GC-low faces, and between CG-high and CG-low faces on recognition accuracy. Among different manipulation techniques previously used, we manipulated realism specifically in terms of texture resolution (Burleigh et al., 2013; Diel et al., 2021). In particular, we generated two CG variants from each selected real photograph by only modifying their texture realism. We chose this technique as it was one of the suggested manipulations that did not introduce other confounding variables, including non-human or abnormal features (Burleigh et al., 2013), which could have affected our results. Our results revealed that CG faces with the least realistic texture (CG-low faces), were less accurately remembered than CG-high and real faces. CG-low faces were the stimuli containing less surface information than the others. Under these conditions, it is possible that lower realism indeed decreases the accuracy of face recognition.

Second, Crookes et al. (2015) hypothesized that when it comes to analyzing CG faces, our face processing system is not as efficient as it is with real faces. In their study, all participants reported having little exposure to CG faces and found poorer recognition performances with CG faces. Therefore, they suggested that individuals with little or no experience with CG faces may be less efficient in recognizing these stimuli. Our results supported this hypothesis as well. Interestingly, our results revealed a significant interaction between face type and participant's self-reported exposure. Specifically, LEPs remembered CG-low faces less accurately than HEPs did. This finding partially supports the hypothesis that participants less exposed to CG characters may find CG faces harder to remember than participants with more experience (H3). However, the effect was limited to CG-low faces, and no significant difference was found between LEPs and HEPs in remembering CG-high faces. The difference in self-reported exposure between the two groups may affect how they process and attend to visual information, resulting in different memory performances for low-resolution CG faces. For example, HEPs may rely more on top-down cognitive processing and prior knowledge to identify identities in low-resolution CG faces. In contrast, LEPs may rely more on bottom-up sensory processing and feature detection.

Finally, Crookes et al. (2015) suggested that CG faces are less well recognized as a result of the UV effect (Mori, 1970). According to this theory, realistic anthropomorphic entities elicit an adverse reaction denoted by feelings of unease and unpleasantness. As a result of this adverse reaction, observers would classify artificial faces as an out-group (e.g., "non-human"). Our results cannot directly support or refute this hypothesis. However, results from our previous study described in Study 1 might help in this case. We have indeed discussed the possibility that CG stimuli elicit an aversive reaction that possibly leads to avoidance behaviours. Given that, in the current study, we used the same stimuli from Study 1, this allows us to compare results between the two studies. In our former experiment, we found CG faces to be rated eerier than real faces, with CG-low ì faces rated as the uncanniest entity. This would seem to support Crookes and colleagues' hypothesis (2015). However, we have no data directly correlating eeriness with recognition accuracy, as participants in studies 1 and 2 were not the same.

Therefore, it is possible that one or more of these factors contributed to the observed differences in the accuracy of artificial and real face recognition accuracy. Future studies should investigate each aspect in more detail to better understand how each modulates identity recognition mechanisms. First, to clarify the role of realism, it would be recommended that future studies consider different realism manipulation techniques to investigate what aspects of the aesthetic appearance of artificial faces might make them harder to process. Texture realism is only one of many techniques that are used in digital graphics to create virtual characters with different degrees of realism (Diel et al., 2021; MacDorman, Green, Ho, & Koch, 2009). First, our result may be restricted to this type of manipulation, and other techniques produce different results. Second, to further explore the role of experience, it would be desirable to conduct training experiments. If exposure is indeed the main predictor of CG faces' recognition accuracy, then we would expect that training with these stimuli, either by being expose to or by interacting with virtual characters, would be beneficial for recognition accuracy scores, reducing the variability in individual's performances (Gauthier & Tarr, 1997).

Furthermore, to define the role of uncanniness in face recognition, it would be interesting for future studies to directly investigate the relationship between CG faces' recognition accuracy and eeriness evaluations. To our knowledge, only one study has attempted to explore this relationship further, though not with CG faces. Geiger and Balas (Geiger & Balas, 2021) conducted a study comparing recognition accuracy with eeriness evaluations of a robot entity. The authors found that higher recall accuracy was linked to an increased perception of uncanniness. The authors speculated that these results could be explained by how eeriness judgments affect a face's distinctiveness. The process of face recognition implies a comparison between the face to be recognized and the set faces observed in our lives: the more it deviates from a perceptual norm, or prototype, the more distinctive it is and the easier it is to recognize it compared to typical faces (Goldstein & Chance, 1980; Valentine, 1991). Based on this assumption, the authors suggested that an uncanny face should be perceived as more atypical than others and they drew on this hypothesis to explain their results. This result is interesting as it seems to contradict the prediction made by Crookes et al. (2015) who, in contrast, hypothesized that CG faces are less efficiently remembered because, since they are evaluated eerier and categorized as out-group, they are less efficiently processed. Nevertheless, results from the experiment in Study 1, seem to support this second hypothesis, as CG faces, especially less realistic ones, were considered eerier and less memorable than real faces.

In conclusion, the present study highlights the challenges associated with recognizing CG faces, compared to real faces. Our findings suggest that CG faces are generally harder to remember and recognize than real faces, with CG faces with lower texture realism being the most difficult. Furthermore, participants with little experience with CG faces were found to have lower recognition accuracy than those with more experience, particularly with CG faces with lower texture resolution. While the UV effect may contribute to the lower recognition of CG faces, more research is needed to investigate the relationship between eeriness, memorability, typicality, and recognition accuracy of CG faces. Overall, this study provides insights into the factors that affect the recognition of CG faces, which could have again implications for the design of CG characters in various contexts.

4. General discussion

In psychological research, using CG faces to replace traditional photographs in face perception studies is becoming increasingly common (Dawel et al., 2021). Although CG faces represent a versatile and flexible alternative to conventional pictures, preliminary research suggested that CG faces' appraisals and processing differ from real faces. This work aimed to analyze CG faces' evaluations (Study 1) and identity recognition (Study 2) by further providing evidence of the effects of CG faces' realism and individuals' self-reported exposure to virtual characters.

4.1. Face appraisal: affinity and first impressions

Initial research has suggested that CG faces tend to receive more unfavourable perceptual and social evaluations than real faces (Balas et al., 2018; Kätsyri, Förger, Mäkäräinen, & Takala, 2015). Results obtained in Study 1 confirmed that CG faces are judged more negatively than real faces. In particular, individuals rated CG faces as less human-like, likable, attractive, trustworthy and eerier than real faces. We further hypothesized that the degree of realism with which CG faces are created is crucial to how observers evaluate these stimuli. Most studies have found a positive correlation between face evaluations and realism, with more realism leading to more favourable ratings (Cheetham et al., 2014; Kätsyri et al., 2019). Our results confirmed this prediction as individuals tended to perceive more realistic CG faces as more human-like, leading to more favourable ratings over less realistic CG faces. This result suggests that increasing CG faces' realism may reduce the discrepancies between CG and real faces. However, researchers should be aware that currently available programs are limited in terms of the level of realism they can achieve. Inconsistencies in the levels of realism of different parts of the CG faces seem unavoidable with the current available graphical programs. Therefore, we suggest that researchers aiming to use CG faces in their experimental paradigms should pre-test whether the realism achieved by their stimuli does not evoke negative feelings and interfere with individuals' evaluations.

Study 1 also showed that individuals more frequently exposed to CG characters evaluate CG faces more positively than people with report less exposure. This distinction supports the notion that repeated exposure to highly realistic CG characters, potentially develop a form of expertise with these stimuli. HEPs' increased acceptance of CG characters as human-like, alongside their more positive impressions, could be interpreted as a manifestation of this expertise. Such observation aligns with previous research within the UV literature, indicating that a lack of familiarity with virtual characters can influence observers' appraisals of virtual faces (Chattopadhyay & MacDorman, 2016; Kätsyri, 2018; Kätsyri et al., 2019) and that regular exposure to an artificial entity diminishes negative feelings and leads to a more positive perception of the character (Burleigh & Schoenherr, 2015; Cheetham, Suter & Jäncke, 2011, 2014; Yamada et al., 2013). The expertise hypothesis (Diamond & Carey, 1986) provides a theoretical framework suggesting that extensive exposure to a category of stimuli, such as faces, enhances one's capability to discern and appreciate stimuli within that category. In the context of CG character evaluation, this hypothesis suggests that HEPs, through their prolonged interaction with CG faces, become more familiar with and skilled at processing CG characters. This increased familiarity and expertise could lead to a greater propensity to perceive these characters as human-like, thereby diminishing the UV effect and fostering more positive evaluations.

Furthermore, CG characters are typically used to replace a human being in a virtual context (e.g., video games). Therefore, HEPs might be more likely to accept a CG character as a human, feel more familiar with it, and form more positive impressions of it. However, this observation doesn't conclusively prove a direct cause-and-effect relationship between higher exposure and positive perceptions. An alternative explanation could be that individuals who find CG faces more aesthetically pleasant or less eerie are more inclined to interact with or view such faces in their daily life.

4.2. Face processing: identity recognition

Preliminary findings indicated that individuals' accuracy in discriminating CG faces is poorer than their accuracy in recognizing real faces. The ORE has been proposed to explain this result. Specifically, since individuals are less exposed to CG faces than real faces in daily contexts, their perceptual systems would be less tuned to these stimuli. The present work results confirmed that CG faces are less efficiently recognized than real faces. Study 2 revealed that individuals' accuracy in recognizing CG faces is lower than real faces. Specifically, recognition accuracy scores on a face's memory task were lower for CG faces than real faces. This result could be interpreted by looking at the specific role of realism and self-reported exposure.

Concerning face recognition, it has been suggested that increasing realism may be beneficial for face identity recognition (Crookes et al., 2015). Our results showed that less realistic CG faces were more challenging to recall than realistic CG faces. There are two possible explanations for this. First, the current digital graphics techniques generate stimuli with limited texture and surface information typically used to discriminate face identities. Consequently, less realistic CG faces with fewer texture details are harder to remember as they lack discriminating information essential for face identity recognition. Given that we manipulated realism in terms of texture resolution, our results support this hypothesis. A second explanation could be that, because highly realistic CG faces resemble humans more accurately, their processing may rely on perceptual expertise gained with real faces. Therefore, their recognition would be facilitated by the individuals' already acquired expertise with real faces.

Furthermore, it has been suggested that expertise may play a key role in determining CG faces' recognition. Research indicates that since individuals are exposed to many different human faces throughout their lives, they develop significant expertise in recognizing and discriminating these stimuli. In particular, because individuals are more frequently exposed to own-ethnicity faces, their accuracy in identifying them is greater than in recognizing other-race faces (Rhodes et al., 1989). Undoubtedly, individuals are less exposed to CG faces than real faces in their daily lives. Consequently, researchers suggested that CG faces may represent a different category of stimuli than real faces, as other-race faces, and therefore be less efficiently recognized. Nonetheless, previous studies have not directly investigated the relationship between individuals' expertise with digital characters and CG faces' perception. Our results showed that individuals frequently exposed to realistic digital characters recognized CG faces more accurately than people with little or no experience with these entities. Therefore, individuals can acquire expertise by being exposed to or interacting with digital characters in everyday life (e.g., playing video games, using digital representations on social media, or using CG faces in professional contexts). Developing expertise with CG faces may improve individuals' capacity to discriminate these stimuli and make their processing less vulnerable to ORE.

5. Limitations

The present studies had some limitations. First, we compared face appraisal and processing using only three levels of realism (CG-low, CG-high and real), making it hard to generalize our results. However, we decided to create only two CG variants to isolate the specific effect of our manipulation technique, namely texture resolution. If we had, for example, created morphs or modified components of the faces, we would have indeed introduced other possible noise variables. Nevertheless, we may have obtained different results using other manipulation techniques, as different changes in CG faces' appearance may produce different responses. To expand external validity, future studies should consider other techniques to understand the specific contribution of each modification on CG character evaluations. Additionally, the design of our study might have influenced the categorical perception observed between CG and real faces. Employing a broader, more graded spectrum of realism in CG faces could potentially yield a more continuous perception of human-likeness. Future research with a finer gradation of realism may provide further insights into how subtle variations in CG realism influence perception.

Similarly, we generated CG faces using Character Creator 3 (*Character Creator,* 2022). This choice was motivated by the fact that it allowed us to manipulate texture resolution directly. However, our results could be limited to the software used, and therefore future comparison between different graphical programs is encouraged.

Finally, another key limitation is that we relied on self-reported exposure to CG characters as an indicator of expertise, without directly manipulating exposure levels. This may not fully capture the intricacies of expertise in recognizing and evaluating CG faces. Future research should consider more direct methods of assessing and manipulating exposure to CG characters for more robust conclusions. For instance, researchers could conduct experiments in which participants are introduced to CG characters in a controlled environment, with adjustments made to how long and how often they are exposed. Additionally, longitudinal studies could be implemented, tracking individual exposure to CG characters over time to assess how exposure influences evaluations and perceptions.

6. Conclusions and future directions

Altogether, these results add new insights into the current literature by showing that realism and self-reported exposure are implicated in CG faces' perception. Specifically, these studies suggest that, to date, CG faces present several limitations that might discourage their use as research stimuli in experimental studies where face perception mechanisms are examined. However, the results on realism and self-reported exposure suggest that the advances in graphics technologies and our increased exposure to realistic digital characters in our life may soon change how we perceive CG faces.

These results extend beyond experimental psychology, with particular relevance for developing CG characters in domains like video games, movies, and virtual reality environments. Such insights hold significant value in contexts where favourable social evaluations can profoundly enrich user experiences. Furthermore, these findings hold substantial implications for computer-mediated communication and the utilization of avatars. Avatars, which symbolize individuals in virtual settings or online platforms (Nowak & Fox, 2018), are crucial in shaping user'' interactions and perceptions, impacting communication quality and source credibility (Di Natale et al., 2021).

In conclusion, our studies contribute to understand how real and CG faces are evaluated differently, the role of realism in these evaluations, and the impact of exposure. Our results emphasize the importance of realism and self-reported exposure in evaluating CG faces and provide valuable insights for developing more socially acceptable CG characters. Researchers have recently considered another kind of CG faces, created by generative adversarial networks, or GANs (Goodfellow et al., 2020; Karras, Laine, & Aila, 2019, pp. 4401–4410). These faces are produced by artificial intelligence (AI) and machine learning algorithms. The machine is trained on photographs of real faces to generate fake faces that realistically look like real people. Generated Photos (Generated Photos, 2022) is an example of such AI system. So far, no evidence exists of their use in psychological research contexts, but the results are encouraging (Nightingale & Farid, 2022). Future research should continue to explore the factors influencing the perception and evaluation of CG faces and their implications for various applications and user experiences.

Ethics approval

All the studies presented in the paper have been approved by the Local Ethical Committee of the Department of Psychology of the University of Milano-Bicocca (Study 1: Protocol RM-2021-407; Study 2: Protocol RM-2021-474). The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Availability of data and materials

The data of both studies are available at Open Science Framework (https://osf.io/nz9j3/?view_only=ee446255e5504650b6e0312354a523c9).

All stimuli used in the studies are accessible from the CFD website, under the additional resources page (https://www.chicagofaces.org/resources/, "Computer-generated CFD faces").

Code availability

The code used for analyses are available at Open Science Framework (https://osf.io/nz9j3/?view_only=ee446255e5504650b6e0312354a523c9).

Funding

No funds, grants, or other support was received.

Preregistration information

This study was not preregistered.

CRediT authorship contribution statement

Anna Flavia Di Natale: Writing – review & editing, Writing – original draft, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stefania La Rocca:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Matilde Ellen Simonetti:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Emanulea Bricolo:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data are shared on a provided OSF link

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbr.2024.100397.

References

Balas, B., & Pacella, J. (2015). Artificial faces are harder to remember. *Computers in Human Behavior, 52*, 331–337.
Balas, B., & Pacella, J. (2017). Trustworthiness perception is disrupted in artificial faces. *Computers in Human Behavior, 77*, 240–248.
Balas, B., Tupa, L., & Pacella, J. (2018). Measuring social variables in real and artificial faces. *Computers in Human Behavior, 88*, 236–243.
Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269.
Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior, 29*(3), 759–771.
Carr, E. W., Brady, T. F., & Winkielman, P. (2017). Are you smiling, or have I seen you before? Familiarity makes faces look happier. *Psychological Science, 28*(8), 1087–1102.
[Software]. *Character creator (Versione 3)* (2022). Reallusion Inc https://www.reallusion.com/character-creator/.
Chattopadhyay, D., & MacDorman, K. F. (2016). Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. *Journal of Vision, 16*(11), 7–7.
Cheetham, M., Suter, P., & Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": Behavioral and functional MRI findings. *Frontiers in Human Neuroscience, 5*, 126.
Cheetham, M., Suter, P., & Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: More like a "Happy Valley". *Frontiers in Psychology, 5*, 1219.
Claypool, H. M., Hugenberg, K., Housley, M. K., & Mackie, D. M. (2007). Familiar eyes are smiling: On the role of familiarity in the perception of facial affect. *European Journal of Social Psychology, 37*(5), 856–866.
Crookes, K., Ewing, L., Gildenhuys, J., Kloth, N., Hayward, W. G., Oxner, M., et al. (2015). How well do computer-generated faces tap face expertise? *PLoS One, 10*(11), Article e0141353.
Dal Ben, R. (2021). *SHINE_color: Controlling low-level properties of colorful images*.
Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2021). A systematic survey of face stimuli used in psychological research 2000–2020. *Behavior Research Methods, 1–13*.
Di Natale, A. F., La Rocca, S., Simonetti, M. E., & Bricolo, E. (2023b). *Using computer-generated faces in experimental psychology: The role of realism and perceptual expertise*. https://osf.io/nz9j3/.
Di Natale, A. F., Simonetti, M. E., La Rocca, S., & Bricolo, E. (2023a). Uncanny valley effect: A qualitative synthesis of empirical research to assess the suitability of using virtual faces in psychological research. *Computers in Human Behavior Reports*, Article 100288.
Di Natale, A. F., Triberti, S., Sibilla, F., Imperato, C., Villani, D., Mancini, T., et al. (2021). Behind a digital mask: Users' subjective experience of animated characters and its effect on source credibility. *Interacting with Computers, 33*(5), 499–510.
Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107.
Diel, A., Weigelt, S., & MacDorman, K. F. (2021). A meta-analysis of the uncanny valley's independent and dependent variables. *ACM Transactions on Human-Robot Interaction (THRI), 11*(1), 1–33.
Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is" special" about face perception? *Psychological Review, 105*(3), 482.
Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673–1682.
Geiger, A. R., & Balas, B. (2021). Robot faces elicit responses intermediate to human faces and objects at face-sensitive ERP components. *Scientific Reports, 11*(1), Article 17890.
Gerlach, C., Kühn, C. D., Mathiassen, A. B., Kristensen, C. L., & Starrfelt, R. (2023). The face inversion effect or the face upright effect? *Cognition, 232*, Article 105335.
Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. *Journal of Psychology, 105*(1), 47–59.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2020). Generative adversarial networks. *Communications of the ACM, 63*(11), 139–144.
Green, R. D., MacDorman, K. F., Ho, C.-C., & Vasudevan, S. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior, 24*(5), 2456–2474.
Grewe, C. M., Liu, T., Kahl, C., Hildebrandt, A., & Zachow, S. (2021). Statistical learning of facial expressions improves realism of animated avatar faces. *Frontiers in Virtual Reality, 2*, Article 619811.
Gross, R. (2005). *Face databases. Handbook of face recognition* (pp. 301–327).
Karras, T., Laine, S., & Aila, T. (2019). *A style-based generator architecture for generative adversarial networks* (pp. 4401–4410).
Kätsyri, J. (2018). Those virtual people all look the same to me: Computer-rendered faces elicit a higher false alarm rate than real human faces in a recognition memory task. *Frontiers in Psychology, 9*, 1362.
Kätsyri, J., de Gelder, B., & Takala, T. (2019). Virtual faces evoke only a weak uncanny valley effect: An empirical investigation with controlled virtual face images. *Perception, 48*(10), 968–991.
Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology, 6*, 390.
Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science, 18*(12), 1084–1089.
Kuefner, D., Macchi Cassia, V., Picozzi, M., & Bricolo, E. (2008). Do all kids look alike? Evidence for an other-age effect in adults. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 811.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest'. *R package version, 2*(0), 734.
La Rocca, S., Gobbo, S., Tosi, G., Fiora, E., & Daini, R. (2023). Look at me now! Enfacement illusion over computer-generated faces. *Frontiers in Human Neuroscience, 17*, Article 1026196.
Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition & Emotion, 24*(8), 1377–1388.
Lenth, R., & Lenth, M. R. (2018). Package 'lsmeans'. *The American Statistician, 34*(4), 216–221.
Looser, C. E., & Wheatley, T. (2010). The tipping point of animacy: How, when, and where we perceive life in a face. *Psychological Science, 21*(12), 1854–1862.
Lu, Z. L., Hua, T., Huang, C. B., Zhou, Y., & Dosher, B. A. (2011). Visual perceptual learning. *Neurobiology of Learning and Memory, 95*(2), 145–151.
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*, 1122–1135.
MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25*(3), 695–710.
Marzi, T., Righi, S., Ottonello, S., Cincotta, M., & Viggiano, M. P. (2014). Trust at first sight: Evidence from ERPs. *Social Cognitive and Affective Neuroscience, 9*(1), 63–72.
Matheson, H. E., & McMullen, P. A. (2011). A computer-generated face database with ratings on realism, masculinity, race, and stereotypy. *Behavior Research Methods, 43*, 224–228.
Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255–260.
McKone, E., & Robbins, R. (2011). *Are faces special* (pp. 149–176). Oxford handbook of face perception.
Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: A meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience, 8*(3), 285–299.
Miller, E. J., Foo, Y. Z., Mewton, P., & Dawel, A. (2023). How do people respond to computer-generated versus human faces? A systematic review and meta-analyses. *Computers in Human Behavior Reports*, Article 100283.
Modeller, F. G. (2021) [Software]. Singular Inversions Inc http://www.facegen.com/.
Mori, M. (1970). The uncanny valley: The original essay by Masahiro Mori. *IEEE Spectrum*.
Murphy, J., Gray, K. L., & Cook, R. (2020). Inverted faces benefit from whole-face processing. *Cognition, 194*, Article 104105.
Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences, 119*(8), Article e2120481119.
Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations. *Review of Communication Research, 6*, 30–53.
Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion, 5*(4), 498.
Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092.
O'toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition, 22*, 208–224.
Photos, G. (2022) [Software]. Generated Media Inc https://generated.photos/.
Poser. (2021) [Software]. Bondware Inc https://www.posersoftware.com/.
Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology, 80*(3), 313–331.
Schindler, S., Zell, E., Botsch, M., & Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific Reports, 7*(1), Article 45003.
Scott, L. S., & Monesson, A. (2009). The origin of biases in face perception. *Psychological Science, 20*(6), 676–680.
Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 46*(2), 225–245.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93*(1), B1–B9.

Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception: Evidence from the face inversion effect. *Vision Research, 51*(11), 1273–1278.

Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology, 23*(3), 373–380.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology, 87*(4), 482.

Turati, C., Sangrigoli, S., Ruely, J., & de Schonen, S. (2004). Evidence of the face inversion effect in 4-month-old infants. *Infancy, 6*(2), 275–297.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*(4), 471–491.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology, 43*(2), 161–204.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598.

Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japanese Psychological Research, 55*(1), 20–32.

Yan, X., Young, A. W., & Andrews, T. J. (2017). The automaticity of face perception is influenced by familiarity. *Attention, Perception, & Psychophysics, 79*, 2202–2211.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747–759.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology, 15*(24), 2256–2262.

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science, 26*(3), 237–242.

Zell, E., Aliaga, C., Jarabo, A., Zibrek, K., Gutierrez, D., McDonnell, R., et al. (2015). To stylize or not to stylize? The effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics, 34*(6), 1–12.