Too Beautiful to be Fake: Attractive Faces are Less Likely to be Judged as Artificially Generated

Dominique Makowski[1], An Shu Te[1], Stephanie Kirk[1], Ngoi Zi Liang[1], Panagiotis Mavros[2], & S.H. Annabel Chen[1, 3, 4, 5]

[1] School of Social Sciences, Nanyang Technological University, Singapore

[2] Singapore-ETH Centre, Future Cities Laboratory, Singapore

[3] LKC Medicine, Nanyang Technological University, Singapore

[4] National Institute of Education, Singapore

[5] Centre for Research and Development in Learning, Nanyang Technological University, Singapore

Author Note

Correspondence concerning this article should be addressed to Dominique Makowski, HSS 04-18, 48 Nanyang Avenue, Singapore (dom.makowski@gmail.com).

The authors made the following contributions. Dominique Makowski: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft; An Shu Te: Data curation, Project administration, Resources, Investigation, Writing – original draft; Stephanie Kirk: Project administration, Resources, Writing – original draft; Ngoi Zi Liang: Project administration, Resources, Writing – review & editing; Panagiotis Mavros: Supervision, Writing – review & editing; S.H. Annabel Chen: Project administration, Supervision, Writing – review & editing.

Correspondence concerning this article should be addressed to Dominique Makowski, HSS 04-18, 48 Nanyang Avenue, Singapore. E-mail: dom.makowski@gmail.com

Abstract

Technological advances render the distinction between artificial (e.g., computer-generated faces) and real stimuli increasingly difficult, yet the factors driving our beliefs regarding the nature of ambiguous stimuli remain largely unknown. In this study, 150 participants rated 109 pictures of faces on 4 characteristics (attractiveness, beauty, trustworthiness, familiarity). The stimuli were then presented again with the new information that some of them were AI-generated, and participants had to rate each image according to whether they believed them to be real or fake. Strikingly, despite all images being pictures of real faces from the same database, most participants rated a large portion of them as "fake". Moreover, our results suggest a gender-dependent role of attractiveness on reality judgements, with faces rated as more attractive being classified as more real. We also report links between reality beliefs tendencies and dispositional traits such as narcissism and paranoid ideation.

Keywords: attractiveness, AI-generated images, fiction, fake news, sense of reality

Word count: 5223

Too Beautiful to be Fake: Attractive Faces are Less Likely to be Judged as Artificially Generated

For the first time in human history, technology has enabled the creation of near-perfect simulations indistinguishable from reality. These artificial, yet realistic constructs permeate all areas of life through immersive works of fiction, deep fakes (real-like images and videos generated by deep learning algorithms), virtual and augmented reality (VR and AR), artificial beings (artificial intelligence "bots" with or without a physical form), fake news and skewed narratives, of which ground truth is often hard to access (Nightingale & Farid, 2022). Such developments not only carry important consequences for the technological and entertainment sectors, but also for security and politics - for instance if used for propaganda and disinformation, recruitment into malevolent organizations, or religious indoctrination (Pantserev, 2020). This issue is central to what has been coined the "post-truth era" (Lewandowsky et al., 2017), in which the distinction (and lack thereof) between authentic and simulated objects will play a critical role.

While not all simulations have achieved perfect realism (Corvi et al., 2022; e.g., Computer Generated Images - CGI in movies or via recent algorithms such as GANs or diffusion models often include distortions or lack certain key details that makes them visually distinct from real images, McDonnell & Breidt, 2010), it is fair to assume that these technical limitations will become negligible in the near future. This is particularly true in the field of face generation, where face-generation algorithms are already able to create stimuli that are virtually indistinguishable from real photos (Moshel et al., 2022; Nightingale & Farid, 2022; Tucciarelli et al., 2020). Such technological feat, however, leads to a new question: if real and fake stimuli cannot be differentiated based on their objective "physical" characteristics, how can we form judgements regarding their nature?

Literature shows that the context surrounding a stimulus often plays an important role in the assessment of its reality (a process henceforth referred to as *simulation* *monitoring*, Makowski, 2018; Makowski, Sperduti, et al., 2019). With the extensive search and processing of cues within ambiguous stimuli being an increasingly complex and cognitively effortful strategy (Michael & Sanson, 2021; Susmann et al., 2021), people tend to draw on peripheral contextual cues (**Figure 1**), such as the source of the stimulus (e.g., which journal was the information published in), and its credibility, authority and expertise, to help facilitate their evaluation (Michael & Sanson, 2021; Petty & Cacioppo, 1986; Susmann et al., 2021). However, the atomization and decontextualization of information allowed by online social media (where text snippets or video excerpts are mass-shared with little context) makes this task increasingly difficult (Berghel, 2018; Y. Chen et al., 2015). Thus, in the absence of clear contextual information, what drives our beliefs of reality?

Evidence suggests that inter-individual characteristics play a crucial role in simulation monitoring, with factors such as cognitive style, prior beliefs, and personality traits (Bryanov & Vziatysheva, 2021; Ecker et al., 2022; Sindermann et al., 2020). For instance, individuals with stronger analytical reasoning have been found to better discriminate real from fake stimuli (Pehlivanoglu et al., 2021; Pennycook & Rand, 2019), and prior knowledge or beliefs about the stimulus influences one's perception of it by biasing the attention deployment towards information that is in line with one's expectations (Britt et al., 2019). Furthermore, dispositional traits, such as high levels of narcissism and low levels of openness and conscientiousness, have been associated with greater susceptibility to fake news (Piksa et al., 2022; Sindermann et al., 2020).

Beyond stimulus- and individual-related characteristics, evidence suggests that the interaction between the two (i.e., the subjective reaction associated with the experience of a given stimulus), contributes to simulation monitoring decisions. For instance, the intensity of experienced emotions have been shown to increase one's sense of presence - the extent to which one feels like "being there", as if the object of experience was real - when engaged in a fictional movie or a VR environment (Makowski et al., 2017; Sanchez-Vives &

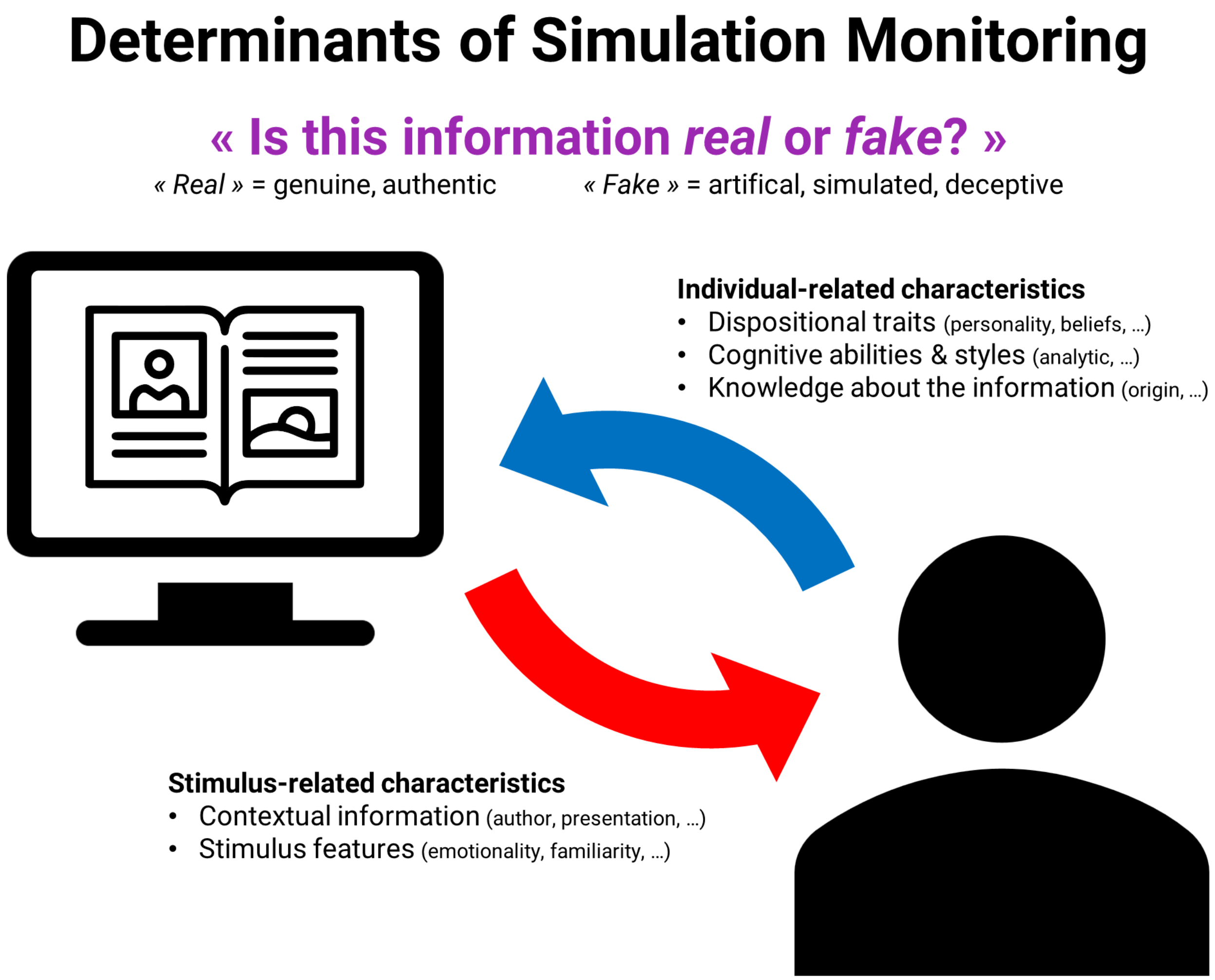

Figure 1. The decision to believe that an ambiguous stimulus (of any form, e.g., images, text, videos, environments, ...) is real or fake depends of individual characteristics (e.g., personality and cognitive styles), stimulus-related features (context, emotionality), and their interaction, which can manifest for instance in our bodily reaction.

Slater, 2005). Indeed, participants' self-reported emotional arousal were found to significantly predict the probability that they would perceive images as real (Azevedo et al., 2020). Conversely, beliefs that emotional stimuli were fake (e.g., that emotional scenes were not authentic but instead involved actors and movie makeup) were found to result in emotion down-regulation (Makowski, Sperduti, et al., 2019; Sperduti et al., 2017). In line with these findings, studies on susceptibility to fake news have also found heightened stimulus emotionality to be associated with greater belief (Bago et al., 2022; Martel et al., 2020), and higher neurophysiological arousal was predictive of judging realistic images as real (Azevedo et al., 2020). Additionally, other factors, such as the stimuli's perceived self-relevance (Goldstein, 2009; Sperduti et al., 2016), as well as familiarity (Begg et al., 1992), could also play a role in guiding our appraisal of a stimulus.

Due to their popularity as a target of CGI technology and the prospect offered with facial features that can be experimentally manipulated, AI-generated images of faces are increasingly used to study face processing (Dawel et al., 2021), in particular in relationship with saliency or emotions, as well as to other important components of face evaluation, such as trustworthiness or attractiveness (Balas & Pacella, 2017; Calbi et al., 2017; Sobieraj & Krämer, 2014; Tsikandilakis et al., 2019). Interestingly, artificially created faces rated as more attractive (by an independent group of raters) were perceived as less real (Tucciarelli et al., 2020). Conversely, Liefooghe et al. (2022) reports that attractiveness ratings were significantly lower when participants who were told that the faces were AI-generated were compared to those who had no prior knowledge. Whereas this line of evidence suggests that reality beliefs have an effect on face attractiveness ratings, the opposite question - whether attractiveness contributes to the formation of reality beliefs - has received little attention to date.

AI-generated content, in particular realistic images, is becoming commonplace and carries important risks for misinformation and black-mailing (Viola & Voto, 2023), emphasizing the need to understand the different components that come into play in the formation of reality beliefs. This exploratory study primarily aims at investigating the effect of facial attractiveness on simulation monitoring, i.e., on the beliefs that an image is real or artificially generated. Based on the *affective reality theory* (**makowski2023affectivefor?**), which suggests that salient and emotional stimuli are perceived to be more real (up to a point of reversal after which beliefs of fiction becomes used an emotion regulation strategy), we hypothesize a quadratic relationship between perceived realness and attractiveness: faces rated as highly attractive or unattractive will more likely be believed to be real. We expect a similar relationship with trustworthiness ratings given its well-established link with attractiveness (Bartosik et al., 2021; Garrido & Prada, 2017; Liefooghe et al., 2022; Little et al., 2011), and a positive relationship with familiarity (as more familiar faces would appear as more salient, self-relevant and anchored in reality). Additionally, we will further explore the link shared by dispositional traits, such as personality and attitude towards AI, with simulation monitoring tendencies. This study aims beyond the investigation of the discriminative accuracy between "true" photos and "true" artificially-generated images, focusing on the beliefs that a stimulus is real or fake, independently of its true nature.

Methods

All the material (preregistration, experiment demo, experiment code, raw data, analysis script with complementary figures and analyses, etc.) is available at **https://github.com/RealityBending/FakeFace**.

Ethics Statement. This study was approved by the NTU Institutional Review Board (NTU IRB-2022-187) and all procedures performed were in accordance with the ethical standards of the institutional board and with the 1964 Helsinki Declaration. All participants provided their informed consent prior to participation and were incentivized after completing the study.

Procedure. In the first part of the study, participants answered a series of personality questionnaires presented in the order below. These include the *Mini-IPIP6* (24 items, Sibley et al., 2011) measuring 6 personality traits, the *SIAS-6* and the *SPS-6* (6 items each, Peters et al., 2012) assessing social anxiety levels, 5 items we devised pertaining to expectations about AI-generated image technology ("I think current Artificial Intelligence algorithms can generate very realistic images"), of which we mixed with 5 items from the general attitudes towards AI scale to lower the former's saliency and the possibility of it priming the subjects about the task, (*GAAIS*, Schepman & Rodway, 2020) the *FFNI-BF* (30 items, Jauk et al., 2022) measuring 9 facets of narcissism; the *R-GPTS* (18 items, Freeman et al., 2021) measuring 2 dimensions related to paranoid thinking; and the *IUS-12* (12 items, Carleton et al., 2007) measuring intolerance to uncertainty. Self-rated attractiveness was also assessed using 2 items - one measuring general attractiveness ("How attractive would you say you are?" Marcinkowska et al., 2021) and the other measuring physical attractiveness ("How would you rate your own physical attractiveness relative to the average," Spielmann et al., 2020). 3 attention check questions were also embedded in the surveys.

In the second part of this study, 109 images of neutral-expression faces from the validated American Multiracial Face Database (AMFD, J. M. Chen et al., 2021) were presented to the participants for 500ms each, in a randomized order, following a fixation cross display (750 ms). After each stimulus presentation, ratings of *Trustworthiness* ("I find this person trustworthy") and *Familiarity* ("This person reminds me of someone I know") were collected using visual analog scales. Notably, as facial attractiveness is a multidimensional construct, encompassing evolutionary, sociocultural, biological as well as cognitive aspects (Han et al., 2018; Rhodes et al., 2006), we assessed attractiveness using 2 visual analog scales, measuring general *Attractiveness* ("I find this person attractive") and physical *Beauty* ("This face is good-looking").

In the last part of the study, participants were informed that "about half" of the images previously seen were AI-generated (the instructions used a cover story explaining that the aim of the research was to validate a new face generation algorithm). The same set of stimuli was displayed again for 500 ms in a new randomized order. This time, after each display, participants were asked to express their belief regarding the nature of the stimulus using a visual analog scale (with *Fake* and *Real* as the two extremes). The study was implemented using *jsPsych* (De Leeuw, 2015), and the exact instructions are available in the experiment code.

Participants. Although the main part of the study relied on within-subject design (with 109 trials per participant), we also planned to do between-participants analyses, thus aiming at collecting a larger sample than traditionally used in experimental psychology (with budget availability as the main constraint). One hundred and fifty participants were recruited via *Prolific*, a crowd-sourcing platform recognized for providing high quality data [Peer et al. (2022); douglas2023data]. The only inclusion criterion was a fluent proficiency in English to ensure that the experiment instructions would be well-understood. Participants were incentivised with a reward of about £7.5 for completing the study, which took about 45 minutes to finish. Demographic variables (age, gender, sexual orientation, education and ethnicity) were self-reported on a voluntary basis.

We excluded 5 participants that either failed 2 (>= 66.6%) or more attention check questions, took an implausibly short time to finish the questionnaires or had incomplete responses. The final sample included 145 participants (Mean age = 28.3, SD = 9.0, range: [19, 66]; Sex: 48.3% females, 51.0% males, 0.7% others).

Data Analysis. The real-fake ratings (measured originally with a [-1, 1] analog scale) were converted into two scores, corresponding to two conceptually distinct mechanisms: the dichotomous *belief* (real or fake, based on the sign of the rating) and the *confidence* (the rating's absolute value) associated with that belief. The former was analyzed using logistic mixed models, which modelled the probability of assigning a face to the real (>= 0) as opposed to fake (< 0). The latter, as well as the other face ratings (attractiveness, beauty, trustworthiness and familiarity), was modelled using mixed beta regressions (suited for outcome variables expressed in percentages). The models included the participants and stimuli as random factors.

We started by investigating the effect of the procedure and instructions to check whether the stimuli (which were all images of real faces) were judged as fake in sufficient proportion to warrant their analysis. Additionally, we assessed the effect of the re-exposure delay, i.e., the time between the first presentation of the image (corresponding to the face ratings) and the second presentation (for the real-fake rating), as well as that of the presentation order to check whether for habituation or learning effects.

The determinants of reality beliefs were modelled separately for attractiveness, beauty, trustworthiness, and familiarity, using second order raw polynomials coefficients to allow for possible quadratic relationships (**Figure 2**). Aside from attractiveness (conceptualized as a general construct), models for beauty, trustworthiness and familiarity were adjusted for the the two remaining variables *mutatis mutandis*. We took into account the gender of participants and stimuli by retaining the stimuli that were aligned with the participants' sexual preference (e.g., female faces for homosexual females, male faces for heterosexual females, and both for bisexual participants), and modeling the interaction with the participants' gender. For the attractiveness and beauty models, we then added the interaction with the reported self-attractiveness (the average of the two questions pertaining to it) to investigate its potential modulatory effect. Finally, we investigated the inter-individual correlates of simulation monitoring with similar models (but this time, for all items regardless of the participant's gender or sexual orientation) for each questionnaire, with all of the subscales as orthogonal predictors.

The analysis was carried out using *R 4.2* (R Core Team, 2022), the *tidyverse* (Wickham et al., 2019), and the *easystats* collection of packages (Lüdecke et al., 2021, 2019, 2020; Makowski et al., 2020; Makowski, Ben-Shachar, et al., 2019). As all the details, scripts and complimentary analyses are open-access, we will focus in the manuscript on findings that are highly statistically significant ($p < .01$).

Results

On average, across participants, 44% of images (95%~CI [0.12, 0.64]) were judged as fake and 56% of images (95%~CI [0.36, 0.88]) as real. An intercept-only model with the

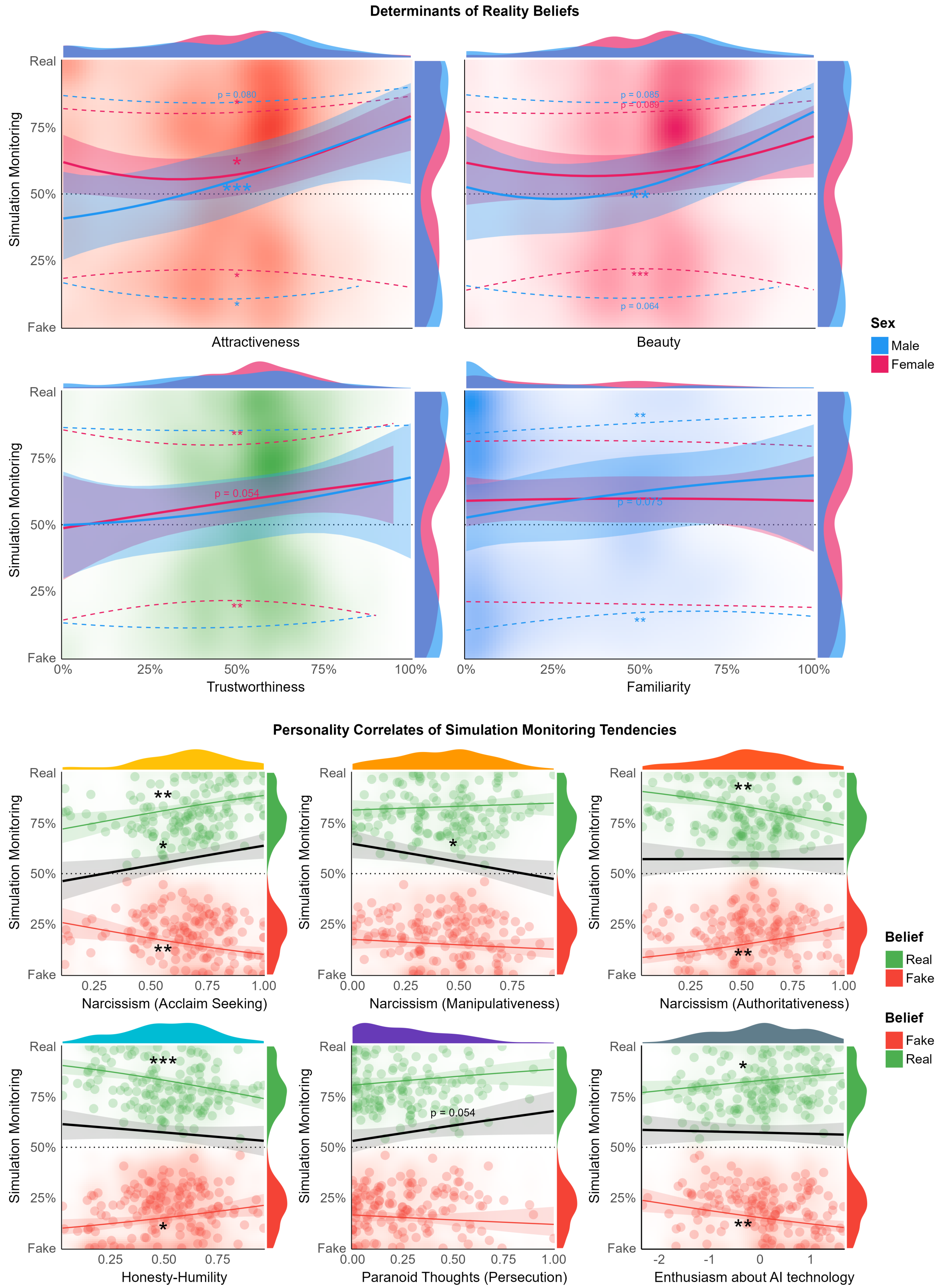

Figure 2. Top part shows the efffect of face ratings on 1) the probability of judging a face as real vs. fake (solid line) and 2) on the confidence associated with that judgement (dashed lines) depending on the sex. Bottom part shows the effect of personality traits on the belief (black line) and the confidence associated with it (colored lines). The points are the average per participant confidence for both types of judgements. Stars indicate significance (p < .001***, p < .01**, p < .05*).

participants and images as random factors showed that the Intraclass Correlation Coefficient (ICC), which can be interpreted as the proportion of variance explained by the random factors, was of 9.0% for the participants and 9.6% for the stimuli.

While the delay of stimulus re-exposure stimulus did not have a significant effect on participants' beliefs of reality ($OR = 1.00$, 95% $CI = [0.99, 1.00]$), judgement confidence was found to be negatively associated with re-exposure delay when the faces were judged as real ($\beta = -0.006$, 95% $CI = [-0.1, 0.002]$, $p = .004$). The presentation order also did not have have an effect on the belief ($OR = 1.00$, 95% $CI = [1.00, 1.00]$) but was related to a decrease of confidence ($\beta_{real} = -0.003$, 95% $CI = [-0.004, -0.002]$, $p < .001$; $\beta_{fake} = -0.002$, 95% $CI = [-0.004, -0.0003]$, $p = .021$): items presented at the end of the session were judged with a similar bias but a decreased overall confidence.

Determinants of Simulation Monitoring. Attractiveness had a significant positive and linear relationship ($R^2_{marginal} = 2.0\%$) with the belief that a stimulus was real ($\beta_{poly1} = 16.57$, 95% $CI = [7.33, 25.82]$, $z = 3.51$, $p < .001$) for males, and a quadratic relationship for females ($\beta_{poly2} = 7.82$, 95% $CI = [1.81, 13.84]$, $z = 2.55$, $p = .011$), with both non-attractive and attractive faces being judged as more real. Attractiveness was also found to have a significant positive and quadratic relationship with confidence in judging faces both as real ($\beta_{poly2} = 4.30$, 95% $CI = [0.97, 7.64]$, $z = 2.53$, $p = .011$) and as fake ($\beta_{poly2} = 5.23$, 95% $CI = [0.86, 9.60]$, $z = 2.35$, $p = .019$) for females. For males, however, a significant negative and quadratic relationship was found between attractiveness ratings and belief confidence only for faces judged as fake ($\beta_{poly2} = -9.92$, 95% $CI = [-18.99, -0.86]$, $z = -2.15$, $p = .032$). There was no interaction with reported self-attractiveness.

Beauty, adjusted for trustworthiness and familiarity, had a significant positive and linear relationship ($R^2_{marginal} = 2.0\%$) with the belief that a stimulus was real ($\beta_{poly1} = 11.82$, 95% $CI = [4.28, 20.21]$, $z = 2.76$, $p = .006$) for males only. No effect on confidence was found, aside from a quadratic relationship in females for faces judged as fake, suggesting that non-beautiful and highly beautiful faces were rated as fake with more confidence than average faces ($\beta_{poly2} = 7.84$, 95% $CI = [3.39, 12.29]$, $z = 3.46$, $p < .001$). There was no interaction with reported self-attractiveness.

Trustworthiness, adjusted for beauty and familiarity, had a predominantly positive and linear relationship ($R^2_{marginal} = 2.0\%$) with the belief that a stimulus was real ($\beta_{poly1} = 6.44$, 95% $CI = [-0.11, 13.00]$, $z = 1.93$, $p = .0054$) for females only. No effect on confidence was found for males, whereas a quadratic relationship was found for females for both faces judged as real ($\beta_{poly2} = 6.14$, 95% $CI = [2.13, 10.14]$, $z = 3.00$, $p = .003$) as well as fake ($\beta_{poly2} = 6.12$, 95% $CI = [1.49, 10.75]$, $z = 2.59$, $p = .001$), suggesting that non-trustworthy and highly trustworthy faces were rated with more confidence than average faces.

We did not find any significant relationships for familiarity adjusted for beauty and trustworthiness ($R^2_{marginal} = 2.0\%$). However, a significant positive and linear relationship was found between familiarity and the confidence judgements of rating faces as real ($\beta_{poly1} = 9.98$, 95% $CI = [3.83, 16.13]$, $z = 3.18$, $p = .001$) whereas a negative linear relationship was found with those judged as fake ($\beta_{poly1} = -12.41$, 95% $CI = [-20.27, -4.54]$, $z = -3.09$, $p = .002$) for males only. This hence suggests that males more confidently judge faces as real with when they are familiar, and as fake when they are unfamiliar.

Note that we also tested as predictors the normative attractiveness and trustworthiness scores (i.e., the average values from the stimuli database validation), which showed a significant positive linear relationship between beliefs of reality and attractiveness, as well as trustworthiness, only for males (see Supplementary Analysis for details).

Inter-Individual Correlates of Simulation Monitoring. The models including the personality traits suggested that *Honesty-Humility* had a significant negative relationship with the confidence associated with real as well as fake judgements ($\beta_{real} = -1.62$, 95% $CI = [-2.55, -0.70]$, $z = -3.43$, $p < .001$; $\beta_{fake} = -1.16$, 95% $CI = [-2.09, -0.23]$, $z = -2.45$, $p = 0.014$).

Significant positive associations were found between the probability of judging faces as real and dimensions of narcissism such as *Acclaim Seeking* ($\beta = 2.24$, 95% $CI = [1.17, 4.27]$, $z = 2.44$, $p = .015$), and *Manipulativeness* ($\beta = 0.47$, 95% $CI = [0.25, 0.87]$, $z = -2.4$, $p = 0.017$). Confidence judgements also shared significant links with narcissism through various facets, such as a positive relationship between the confidence for both real and fake judgements with *Acclaim Seeking* ($\beta_{real} = 1.65$, 95% $CI = [0.59, 2.70]$, $z = 3.07$, $p = .002$; $\beta_{fake} = 1.62$, 95% $CI = [0.56, 2.68]$, $z = 3.00$, $p = .003$), and a negative relationship with *Authoritativeness* ($\beta_{real} = -1.57$, 95% $CI = [-2.58, -0.57]$, $z = -3.08$, $p = .002$; $\beta_{fake} = -1.49$, 95% $CI = [-2.50, -0.48]$, $z = -2.89$, $p = .004$).

A positive trend was found in the relationship between the *Persecutory Ideation* dimension of paranoid thinking and the belief that the faces were real ($\beta = 1.87$, 95% $CI = [0.99, 3.54]$, $z = 1.93$, $p = .054$).

The *Prospective Anxiety* aspect of intolerance to uncertainty shared a negative trend in its association with confidence ratings ($\beta_{real} = 1.43$, 95% $CI = [0.10, 2.76]$, $z = 2.10$, $p = .036$; $\beta_{fake} = -0.91$, 95% $CI = [-1.93, 0.11]$, $z = -1.75$, $p = .081$). No significant effect was found for social anxiety.

Questions pertaining to the attitude towards AI were reduced to 3 dimensions through factor analysis, labelled AI-Enthusiasm (loaded by items expressing interest and excitement in AI development and applications), AI-Realness (loaded by items expressing positive opinions on the ability of AI to create realistic material), and AI-Danger (loaded by items expressing concerns on the unethical misuse of AI technology). However, only

AI-Enthusiasm displayed a significant positive relationship with the confidence in both real and fake judgements ($\beta_{real} = 0.21$, 95% $CI = [0.02, 0.40]$, $z = 2.20$, $p = .028$; $\beta_{fake} = 0.31$, 95% $CI = [0.12, 0.50]$, $z = -8.90$, $p < 0.001$).

Discussion

This study aimed at investigating the effect of facial ratings (attractiveness, beauty, trustworthiness and familiarity) on simulation monitoring, i.e., on the belief that a stimulus was artificially generated. Most strikingly, despite all the stimuli being real faces from the same database, all participants believed (to high degrees of confidence) that a significant proportion of them were fake. This finding not only attests to the effectiveness of our instructions, but highlights the current levels of expectation regarding CGI technology. The strong impact of prior expectations and information on reality beliefs demonstrated here underlines the volatility of our sense of reality. In fact, stimuli-related and participant-related characteristics accounted together for less than 20% of the beliefs variance, suggesting a large contribution of other subjective processes.

Although attractiveness did not seem to be the primary drive underlying simulation monitoring of face images, we do nonetheless report significant associations, with different patterns observed depending on the participant's gender. The quadratic relationship found for female participants is aligned with our hypothesis that salient faces (i.e., rated as very attractive or very unattractive) are judged to be more real. The fact that this effect did not reach significance for beauty underlines that attractiveness judgement, and its role in simulation monitoring, is a multidimensional construct that cannot be reduced to physical facial attractiveness, in particular for women (Buunk et al., 2002; Qi & Ying, 2022). In fact, female participants were more confident in judging faces as fake only when they were rated very high or low on beauty, suggesting that physical beauty and attractiveness are not analogous in their effects on simulation monitoring decisions.

Interestingly, we found a significant positive linear relationship in male participants for both attractiveness and beauty on simulation monitoring that we could interpret under an evolutionary lens. Specifically, males purportedly place more emphasis on facial attractiveness as a sign of reproductive potential, as compared with females, who tend to value characteristics signaling resource acquisition capabilities (Buunk et al., 2002; Fink et al., 2006; Qi & Ying, 2022). It is thus possible that the evolutionary weight associated with attractiveness skewed the perceived saliency of attractive faces for men, rendering them significantly more salient than unattractive faces, and in turn distorting the relationship with simulation monitoring. However, future studies should test this saliency-based hypothesis by measuring constructs closer to salience and its effects, for instance using neuroimaging (Indovina & Macaluso, 2007; Lou et al., 2015) or physiological markers (e.g., heart rate deceleration, Skora et al., 2022).

Our results found a positive linear trend between trustworthiness and simulation monitoring for females only. Given prior evidence that faces presented as computer-generated were rated less trustworthy (Balas & Pacella, 2017; Hoogers, 2021; Liefooghe et al., 2022), we expected such a linear association to be more clearly present for both genders. One of the underlying mechanisms that possibly contributed to this dimorphism could be the increased risk-taking aversion reported in females (explained evolutionarily as a compromise to their reproductive potential, Van Den Akker et al., 2020), to which perceived facial trustworthiness relates (Hou & Liu, 2019). However, if that was the case, faces judged as highly untrustworthy should have appeared as even more salient (representing an evolutionary threat), and hence be judged as more real, leading to a quadratic relationship between trustworthiness and simulation monitoring instead. Further studies are needed to investigate the causes of the increased simulation monitoring sensitivity to trustworthiness in females.

Contrary to our hypothesis, we did not find familiarity to be significantly related to simulation monitoring decisions. Interestingly, there were significant linear relationships between familiarity and confidence judgements for males only, where familiarity increased the confidence of reality beliefs. Although the familiarity measure was not a "recognition" measure, evidence from studies pertaining to the latter could be linked, reporting better face memory for females (Lewin & Herlitz, 2002; Mishra et al., 2019; Sommer et al., 2013), as well as an overconfidence in face recall for males (Bailey, 2021; Herbst, 2020). However, it should be noted that the present study's distribution of familiarity ratings was strongly skewed, and only a low number of pictures was rated as highly familiar. As such, future studies should clarify this point by experimentally manipulating familiarity, for instance by modulating the amount of exposure to items before querying the simulation monitoring judgements.

Regarding the role of inter-individual characteristics in simulation monitoring tendencies, we found higher scores of honesty-humility - a trait related to an increased risk perception and aversion (Levidi et al., 2022; Weller & Thulin, 2012) - to be related to a lower confidence in simulation monitoring judgements. Notably, greater narcissistic tendencies in dimensions such as acclaim seeking and manipulativeness were associated with a higher number of faces judged as real. This is in line with recent research which found people with higher narcissism scores less likely to engage in analytical reasoning strategies such as reflective thinking (Ahadzadeh et al., 2021; Littrell et al., 2020), and to be more vigilant and attentive to external stimuli (Carolan, 2017; Eddy, 2021; Grapsas et al., 2020).

Moreover, putting the significant positive links between narcissistic acclaim seeking and confidence judgements in perspective with the negative correlation between honesty-humility and narcissism (Hodson et al., 2018), we confirm previous evidence regarding the relationship between narcissistic grandiosity and over-confidence in decision-making (Brunell & Buelow, 2017; Campbell et al., 2004; Chatterjee & Pollock, 2017; O'Reilly & Hall, 2021). Although an inverse effect was found for the narcissistic facet of authoritativeness, we interpret this relationship as related to a higher response assertiveness. Taken together, these results suggest that participants with low humility and high recognition desires are more confident in their judgement regarding the real or fake nature of ambiguous stimuli. Alternatively, participants with opposite traits might perceive a higher risk in the decision-making process and its potential consequences (e.g., being seen as bad at the task at hand), resulting in more conservative confidence ratings.

Our findings suggest - though with weak significance - a positive link between paranoid ideation and the tendency to believe that the stimuli were real. Given previous reports that people with higher levels of paranoia are more sensitive to cues of social threat (Fornells-Ambrojo et al., 2015; Freeman et al., 2003; King & Dudley, 2017), it is plausible that paranoid traits confer greater saliency and emotionality to observed faces, hence increasing perceptions of its realness. This hypothesis, if confirmed by future studies, would be in line with previous findings that persecutory delusions are predicted by a greater sense of presence in VR environments populated with virtual characters (Freeman et al., 2005).

Despite the ubiquity of AI, the literature pertaining to the influence of people's AI attitudes on simulation monitoring is scarce. Contrary to our expectations, we did not find evidence for the role of participants' expectations regarding the capabilities of AI technology (in terms of the realism of its productions). Instead, we found only one's enthusiasm about AI technology to be related to an increased confidence in simulation monitoring ratings. This could potentially be because participants with a highly positive attitude towards AI perceive themselves as having greater knowledge about AI and its capabilities (Said et al., 2022), hence permitting themselves to be more confident in their simulation monitoring decisions. In fact, this result is in line with reports that AI attitudes interacts with people's perceived self-knowledge to influence their perception of the opportunities and risks accorded by AI applications (Said et al., 2022).

On a methodological level, although the order of presentation of the facial images was randomized to reduce effects of adaptation, participants were more confident in their judgements for faces perceived as real following a shorter re-exposure delay. Such shorter durations could be associated with the faces being better remembered and appearing more familiar, thereby triggering self-referential and autobiographical memory processing during the repeated display (Abraham & Von Cramon, 2009; Gobbini et al., 2013; Taylor et al., 2009). Indeed, this finding is consistent with studies in which fictional stimuli that were associated with familiarity up-regulated emotions, biasing its salience and perceived realness (Makowski et al., 2017; Sperduti et al., 2016). However, if that was the case, we would expect shorter re-exposure delays to impact the decision bias as well towards reality, rather than simply the confidence. Future studies should further investigate the modulatory effects of types and degrees of familiarity on perceived realness judgements.

Several limitations have to be noted. The current experimental paradigm required participants to judge the realness of faces they had prior exposure to (which was done to prevent reality judgements from influencing the other ratings). Although the effect of re-exposure delay was negligible, the potential bias induced by face familiarity (as compared to judging completely new items) cannot be discarded. Future studies could examine that by incorporating novel face images or increasing the duration of the re-exposure delay.

Another issue is the impact on reality judgements of the prior explicit instruction that “about half of the faces were AI-generated and the other half real photos”. Given this prior information given to participants, it migjt seems like our enthusiasm pertaining to the finding that most people did indeed believe a high number of stimuli to be fake might be unwarranted: all it shows is that participants followed the instructions. First of all, even if that was the case, the fact that our beliefs of reality can be so easily re-programmed with basic instructions ad lead to high-confidence answers remains an interesting phenomenon. Moreover, it is to note that the paradigm did not instruct participants to balance their answers according to a certain distribution (e.g., 50-50), merely providing a description of the dataset. The fact that no presentation order effect was found on reality beliefs gives credit to the assumption that participants did not try to actively distribute their responses to match the instructions, in which case we would have expected a different pattern, for instance the first items judged as real (the initial "true" belief of the participants), and later a hard bias towards responding "fake" (as participants realize that all stimuli are of similar nature and that they have to "make up" for the prevalence of their "real" answers to fulfil the instructions).

That said, the potential confounding effect of the instructions still exists, and a control condition without the cover story with AI-generated images would in-principle be welcome. However, the distinction real/fake is hard to operationalize and introduce to participants in a vacuum (asking them to discriminate real from fake without introducing the context and meaning of "fake" seems hardly feasible). That said, future studies should study the impact of these higher-order expectations on ratings (for instance, Tucciarelli et al., 2020 found that merely mentioning that some faces were AI generated decreased on average the trustworthiness ratings for all faces) as well as on the simulation monitoring process itself (i.e., the "criterion": would people form and distribute judgements differently). This can be studied by modulating this expectation in a controlled fashion (e.g., "most of the images *but a few* are real" vs. "most of the images *but a few* are fake") or inventing some implicit way of measuring reality belief that would not require the explicit introduction of the concept of fake vs. real to participants.

Finally, it is important to note that although consistent in their directions across models and variables, the magnitude of the effects found in the study was relatively small, suggesting that the facial features measured in the study were not the key determinants of simulation monitoring. Hence, beyond exploring new potential mechanisms, future studies should include a more thorough debriefing to try to capture what conscious strategies (if any) the participants used (e.g., focusing on some features of the stimulus - like hair or eyes in the case of faces) to guide their reality beliefs.

In summary, the aim of the present study was to examine whether a subset of specific characteristics, in particular face attractiveness, significantly influences our simulation monitoring decisions. Notably, we found faces rated as attractive to be perceived as more real, with a possible sexual dimorphism affecting the shape of the relationship. We also found that inter-individual traits, such as narcissistic acclaim-seeking and manipulativeness, as well as persecutory ideation, were related to a systematic bias towards beliefs that the stimuli were real or fake. We believe that these findings provide the foundations to help us understand what drives reality beliefs in an increasingly reality-ambiguous world.

Data Availability

The datasets generated and/or analysed during the current study are available in the GitHub repository https://github.com/RealityBending/FakeFace

Funding

This work was supported by the Presidential Postdoctoral Fellowship Grant (NTU-PPF-2020-10014) from Nanyang Technological University (awarded to DM) and the Intra-CREATE Seed Collaboration Grant (NRF2021-ITS008-0010) from the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) programme (awarded to DM and PM).

Acknowledgments

We would like to thank Taong Ren Qing Malcolm for his contribution to the selection of the materials.

References

Abraham, A., & Von Cramon, D. Y. (2009). Reality= relevance? Insights from spontaneous modulations of the brain's default network when telling apart reality from fiction. *PloS One*, *4*(3), e4741.

Ahadzadeh, A. S., Ong, F. S., & Wu, S. L. (2021). Social media skepticism and belief in conspiracy theories about COVID-19: The moderating role of the dark triad. *Current Psychology*, 1–13.

Azevedo, R., Tucciarelli, R., De Beuklaer, S., Ambroziak, K., Jones, I., & Tsakiris, M. (2020). *A body of evidence:'feeling in seeing'predicts realness judgments for photojournalistic images.*

Bago, B., Rosenzweig, L. R., Berinsky, A. J., & Rand, D. G. (2022). Emotion may predict susceptibility to fake news but emotion regulation does not seem to help. *Cognition and Emotion*, 1–15.

Bailey, A. (2021). *A gender in-group effect on facial recall* [PhD thesis]. University of Tasmania.

Balas, B., & Pacella, J. (2017). Trustworthiness perception is disrupted in artificial faces. *Computers in Human Behavior*, *77*. https://doi.org/10.1016/j.chb.2017.08.045

Bartosik, B., Wojcik, G. M., Brzezicka, A., & Kawiak, A. (2021). Are you able to trust me? Analysis of the relationships between personality traits and the assessment of attractiveness and trust. *Frontiers in Human Neuroscience*, *15*, 685530.

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(4), 446.

Berghel, H. (2018). Weaponizing twitter litter: Abuse-forming networks and social media. *Computer*, *51*(4), 70–73.

Britt, M. A., Rouet, J.-F., Blaum, D., & Millis, K. (2019). A reasoned approach to dealing with fake news. *Policy Insights from the Behavioral and Brain Sciences*, *6*(1), 94–101.

Brunell, A. B., & Buelow, M. T. (2017). Narcissism and performance on behavioral decision-making tasks. *Journal of Behavioral Decision Making*, *30*(1), 3–14.

Bryanov, K., & Vziatysheva, V. (2021). Determinants of individuals' belief in fake news: A scoping review determinants of belief in fake news. *PLoS One*, *16*(6), e0253717.

Buunk, B. P., Dijkstra, P., Fetchenhauer, D., & Kenrick, D. T. (2002). Age and gender differences in mate selection criteria for various involvement levels. *Personal Relationships*, *9*(3), 271–278.

Calbi, M., Heimann, K., Barratt, D., Siri, F., Umiltà, M. A., & Gallese, V. (2017). How context influences our perception of emotional faces: A behavioral study on the kuleshov effect. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01684

Campbell, W. K., Goodie, A. S., & Foster, J. D. (2004). Narcissism, confidence, and risk attitude. *Journal of Behavioral Decision Making*, *17*(4), 297–311.

Carleton, R. N., Norton, M. P. J., & Asmundson, G. J. (2007). Fearing the unknown: A short version of the intolerance of uncertainty scale. *Journal of Anxiety Disorders*, *21*(1), 105–117.

Carolan, P. L. (2017). *Searching "inaffectively": A behavioral, psychometric, and electroencephalographic investigation of psychopathic personality and visual-spatial attention* [PhD thesis]. Arts & Social Sciences: Department of Psychology.

Chatterjee, A., & Pollock, T. G. (2017). Master of puppets: How narcissistic CEOs construct their professional worlds. *Academy of Management Review*, *42*(4), 703–725.

Chen, J. M., Norman, J. B., & Nam, Y. (2021). Broadening the stimulus set: Introducing the american multiracial faces database. *Behavior Research Methods*, *53*(1), 371–389.

Chen, Y., Conroy, N. K., & Rubin, V. L. (2015). News in an online world: The need for an "automatic crap detector." *Proceedings of the Association for Information Science and Technology*, *52*(1), 1–4.

Corvi, R., Cozzolino, D., Zingarini, G., Poggi, G., Nagano, K., & Verdoliva, L. (2022). On the detection of synthetic images generated by diffusion models. *arXiv Preprint arXiv:2211.00680*.

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.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29.

Eddy, C. M. (2021). Self-serving social strategies: A systematic review of social cognition in narcissism. *Current Psychology*, 1–19.

Fink, B., Neave, N., Manning, J. T., & Grammer, K. (2006). Facial symmetry and judgements of attractiveness, health and personality. *Personality and Individual Differences*, *41*(3), 491–499.

Fornells-Ambrojo, M., Freeman, D., Slater, M., Swapp, D., Antley, A., & Barker, C. (2015). How do people with persecutory delusions evaluate threat in a controlled social environment? A qualitative study using virtual reality. *Behavioural and Cognitive Psychotherapy*, *43*(1), 89–107.

Freeman, D., Garety, P. A., Bebbington, P., Slater, M., Kuipers, E., Fowler, D., Green, C., Jordan, J., Ray, K., & Dunn, G. (2005). The psychology of persecutory ideation II: A virtual reality experimental study. *The Journal of Nervous and Mental Disease*, *193*(5), 309–315.

Freeman, D., Loe, B. S., Kingdon, D., Startup, H., Molodynski, A., Rosebrock, L., Brown, P., Sheaves, B., Waite, F., & Bird, J. C. (2021). The revised green et al., Paranoid thoughts scale (r-GPTS): Psychometric properties, severity ranges, and clinical cut-offs. *Psychological Medicine*, *51*(2), 244–253.

Freeman, D., Slater, M., Bebbington, P. E., Garety, P. A., Kuipers, E., Fowler, D., Met, A., Read, C. M., Jordan, J., & Vinayagamoorthy, V. (2003). Can virtual reality be used to investigate persecutory ideation? *The Journal of Nervous and Mental Disease*, *191*(8), 509–514.

Garrido, M. V., & Prada, M. (2017). KDEF-PT: Valence, emotional intensity, familiarity and attractiveness ratings of angry, neutral, and happy faces. *Frontiers in Psychology*, *8*, 2181.

Gobbini, M. I., Gors, J. D., Halchenko, Y. O., Rogers, C., Guntupalli, J. S., Hughes, H., & Cipolli, C. (2013). Prioritized detection of personally familiar faces. *PloS One*, *8*(6), e66620.

Goldstein, T. R. (2009). The pleasure of unadulterated sadness: Experiencing sorrow in fiction, nonfiction, and" in person.". *Psychology of Aesthetics, Creativity, and the Arts*, *3*(4), 232.

Grapsas, S., Brummelman, E., Back, M. D., & Denissen, J. J. (2020). The “why” and “how” of narcissism: A process model of narcissistic status pursuit. *Perspectives on Psychological Science*, *15*(1), 150–172.

Han, S., Li, Y., Liu, S., Xu, Q., Tan, Q., & Zhang, L. (2018). Beauty is in the eye of the beholder: The halo effect and generalization effect in the facial attractiveness evaluation. *Acta Psychologica Sinica*, *50*(4), 363.

Herbst, T. H. (2020). Gender differences in self-perception accuracy: The confidence gap and women leaders’ underrepresentation in academia. *SA Journal of Industrial Psychology*, *46*(1), 1–8.

Hodson, G., Book, A., Visser, B. A., Volk, A. A., Ashton, M. C., & Lee, K. (2018). Is the dark triad common factor distinct from low honesty-humility? *Journal of Research in Personality*, *73*, 123–129.

Hoogers, E. (2021). *The effect of attitude towards computer generated faces on face perception* [{B.S.} thesis].

Hou, C., & Liu, Z. (2019). The survival processing advantage of face: The memorization of the (un) trustworthy face contributes more to survival adaptation. *Evolutionary Psychology*, *17*(2), 1474704919839726.

Indovina, I., & Macaluso, E. (2007). Dissociation of stimulus relevance and saliency factors during shifts of visuospatial attention. *Cerebral Cortex*, *17*(7), 1701–1711.

Jauk, E., Olaru, G., Schürch, E., Back, M. D., & Morf, C. C. (2022). Validation of the german five-factor narcissism inventory and construction of a brief form using ant colony optimization. *Assessment*, 10731911221075761.

King, A., & Dudley, R. (2017). Paranoia, worry, cognitive avoidance and intolerance of uncertainty in a student population. *Journal of Applied Psychology and Social Science*, *3*(2), 70–89.

Levidi, M. D. C., McGrath, A., Kyriakoulis, P., & Sulikowski, D. (2022). Understanding criminal decision-making: Links between honesty-humility, perceived risk and negative affect: Psychology, crime & law. *Psychology, Crime and Law*, 1–29.

Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition—women's faces make the difference. *Brain and Cognition*, *50*(1), 121–128.

Liefooghe, B., Oliveira, M., Leisten, L. M., Hoogers, E., Aarts, H., & Hortensius, R. (2022). *Faces merely labelled as artificial are trusted less.*

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: Evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1638–1659.

Littrell, S., Fugelsang, J., & Risko, E. F. (2020). Overconfidently underthinking: Narcissism negatively predicts cognitive reflection. *Thinking & Reasoning*, *26*(3), 352–380.

Lou, B., Hsu, W.-Y., & Sajda, P. (2015). Perceptual salience and reward both influence feedback-related neural activity arising from choice. *Journal of Neuroscience*, *35*(38), 13064–13075.

Lüdecke, D., Ben-Shachar, M., Patil, I., & Makowski, D. (2020). Extracting, computing and exploring the parameters of statistical models using R. *Journal of Open Source Software*, *5*(53), 2445. https://doi.org/10.21105/joss.02445

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Lüdecke, D., Waggoner, P., & Makowski, D. (2019). Insight: A unified interface to access information from model objects in R. *Journal of Open Source Software*, *4*(38), 1412. https://doi.org/10.21105/joss.01412

Makowski, D. (2018). *Cognitive neuropsychology of implicit emotion regulation through fictional reappraisal* [PhD thesis]. Sorbonne Paris Cité.

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Makowski, D., Ben-Shachar, M., Patil, I., & Lüdecke, D. (2020). Methods and algorithms for correlation analysis in R. *Journal of Open Source Software*, *5*(51), 2306. https://doi.org/10.21105/joss.02306

Makowski, D., Sperduti, M., Nicolas, S., & Piolino, P. (2017). "Being there" and remembering it: Presence improves memory encoding. *Consciousness and Cognition*, *53*, 194–202.

Makowski, D., Sperduti, M., Pelletier, J., Blondé, P., La Corte, V., Arcangeli, M., Zalla, T., Lemaire, S., Dokic, J., Nicolas, S., et al. (2019). Phenomenal, bodily and brain correlates of fictional reappraisal as an implicit emotion regulation strategy. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(4), 877–897.

Marcinkowska, U. M., Jones, B. C., & Lee, A. J. (2021). Self-rated attractiveness predicts preferences for sexually dimorphic facial characteristics in a culturally diverse sample. *Scientific Reports*, *11*(1), 1–8.

Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*, *5*(1), 1–20.

McDonnell, R., & Breidt, M. (2010). Face reality: Investigating the uncanny valley for virtual faces. In *ACM SIGGRAPH ASIA 2010 sketches* (pp. 1–2).

Michael, R. B., & Sanson, M. (2021). Source information affects interpretations of the news across multiple age groups in the united states. *Societies*, *11*(4), 119.

Mishra, M. V., Likitlersuang, J., B Wilmer, J., Cohan, S., Germine, L., & DeGutis, J. M. (2019). Gender differences in familiar face recognition and the influence of sociocultural gender inequality. *Scientific Reports*, *9*(1), 1–12.

Moshel, M. L., Robinson, A. K., Carlson, T. A., & Grootswagers, T. (2022). Are you for real? Decoding realistic AI-generated faces from neural activity. *Vision Research*, *199*, 108079. https://doi.org/10.1016/j.visres.2022.108079

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), e2120481119. https://doi.org/10.1073/pnas.2120481119

O'Reilly, C. A., & Hall, N. (2021). Grandiose narcissists and decision making: Impulsive, overconfident, and skeptical of experts–but seldom in doubt. *Personality and Individual Differences*, *168*, 110280.

Pantserev, K. (2020). *The malicious use of AI-based deepfake technology as the new threat to psychological security and political stability* (pp. 37–55). https://doi.org/10.1007/978-3-030-35746-7_3

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Pehlivanoglu, D., Lin, T., Deceus, F., Heemskerk, A., Ebner, N. C., & Cahill, B. S. (2021). The role of analytical reasoning and source credibility on the evaluation of real and fake full-length news articles. *Cognitive Research: Principles and Implications*, *6*(1), 1–12.

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50.

Peters, L., Sunderland, M., Andrews, G., Rapee, R. M., & Mattick, R. P. (2012). Development of a short form social interaction anxiety (SIAS) and social phobia scale (SPS) using nonparametric item response theory: The SIAS-6 and the SPS-6. *Psychological Assessment*, *24*(1), 66.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and persuasion* (pp. 1–24). Springer.

Piksa, M., Noworyta, K., Piasecki, J., Gwiazdzinski, P., Gundersen, A. B., Kunst, J., & Rygula, R. (2022). Cognitive processes and personality traits underlying four phenotypes of susceptibility to (mis) information. *Frontiers in Psychiatry*, 1142.

Qi, Y., & Ying, J. (2022). Gender biases in the accuracy of facial judgments: Facial attractiveness and perceived socioeconomic status. *Frontiers in Psychology*, *13*.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rhodes, G. et al. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*, 199.

Said, N., Potinteu, A.-E., Brich, I., Buder, J., Schumm, H., & Huff, M. (2022). *An artificial intelligence perspective: How knowledge and confidence shape risk and opportunity perception.*

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, *6*(4), 332–339.

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence scale. *Computers in Human Behavior Reports*, *1*, 100014.

Sibley, C., Luyten, N., Wolfman, M., Mobberley, A., Wootton, L. W., Hammond, M., Sengupta, N., Perry, R., West-Newman, T., Wilson, M., McLellan, L., Hoverd, W. J., & Robertson, A. (2011). The mini-IPIP6: Validation and extension of a short measure of the big-six factors of personality in new zealand. *New Zealand Journal of Psychology*, *40*, 142–159.

Sindermann, C., Cooper, A., & Montag, C. (2020). A short review on susceptibility to falling for fake political news. *Current Opinion in Psychology*, *36*, 44–48.

Skora, L., Livermore, J., & Roelofs, K. (2022). The functional role of cardiac activity in perception and action. *Neuroscience & Biobehavioral Reviews*, 104655.

Sobieraj, S., & Krämer, N. C. (2014). What is beautiful in cyberspace? Communication with attractive avatars. *International Conference on Social Computing and Social Media*, 125–136.

Sommer, W., Hildebrandt, A., Kunina-Habenicht, O., Schacht, A., & Wilhelm, O. (2013). Sex differences in face cognition. *Acta Psychologica*, *142*(1), 62–73.

Sperduti, M., Arcangeli, M., Makowski, D., Wantzen, P., Zalla, T., Lemaire, S., Dokic, J., Pelletier, J., & Piolino, P. (2016). The paradox of fiction: Emotional response toward fiction and the modulatory role of self-relevance. *Acta Psychologica*, *165*, 53–59.

Sperduti, M., Makowski, D., Arcangeli, M., Wantzen, P., Zalla, T., Lemaire, S., Dokic, J., Pelletier, J., & Piolino, P. (2017). The distinctive role of executive functions in implicit emotion regulation. *Acta Psychologica*, *173*, 13–20.

Spielmann, S. S., Maxwell, J. A., MacDonald, G., Peragine, D., & Impett, E. A. (2020). The predictive effects of fear of being single on physical attractiveness and less selective partner selection strategies. *Journal of Social and Personal Relationships*, *37*(1), 100–123.

Susmann, M. W., Xu, M., Clark, J. K., Wallace, L. E., Blankenship, K. L., Philipp-Muller, A. Z., Luttrell, A., Wegener, D. T., & Petty, R. E. (2021). Persuasion amidst a pandemic: Insights from the elaboration likelihood model. *European Review of Social Psychology*, 1–37.

Taylor, M. J., Arsalidou, M., Bayless, S. J., Morris, D., Evans, J. W., & Barbeau, E. J. (2009). Neural correlates of personally familiar faces: Parents, partner and own faces. *Human Brain Mapping*, *30*(7), 2008–2020.

Tsikandilakis, M., Bali, P., & Chapman, P. (2019). Beauty is in the eye of the beholder: The appraisal of facial attractiveness and its relation to conscious awareness. *Perception*, *48*(1), 72–92.

Tucciarelli, R., Vehar, N., & Tsakiris, M. (2020). *On the realness of people who do not exist: the social processing of artificial faces*. https://doi.org/10.31234/osf.io/dnk9x

Van Den Akker, O. R., Assen, M. A. van, Van Vugt, M., & Wicherts, J. M. (2020). Sex differences in trust and trustworthiness: A meta-analysis of the trust game and the gift-exchange game. *Journal of Economic Psychology*, *81*, 102329.

Viola, M., & Voto, C. (2023). Designed to abuse? Deepfakes and the non-consensual diffusion of intimate images. *Synthese*, *201*(1), 1–20.

Weller, J. A., & Thulin, E. W. (2012). Do honest people take fewer risks? Personality correlates of risk-taking to achieve gains and avoid losses in HEXACO space. *Personality and Individual Differences*, *53*(7), 923–926.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686