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] LKC Medicine, Nanyang Technological University, Singapore

[3] National Institute of Education, Singapore

[4] 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: 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

Abstract abstract abstract.

Keywords: attractiveness, simulation monitoring, fiction, deep fakes, sense of reality

Word count: 5114

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 carries 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 (e.g., Computer Generated Images - CGI in movies often 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, in particular in the field of face generation and replacement (Moshel et al., 2022; Nightingale & Farid, 2022; Tucciarelli et al., 2020). This fact, however, leads to a new issue: if real and fake stimuli cannot be distinguished based on their objective characteristics, how can we make judgments regarding their nature?

Literature shows that the context surrounding a stimulus often plays an important role in the assessment of its reality (Makowski, 2018; a process henceforth referred to as *simulation monitoring*, 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, 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) can render this task difficult (Berghel, 2018; Y. Chen et al., 2015). In the absence of contextual information, what drives our beliefs of reality?

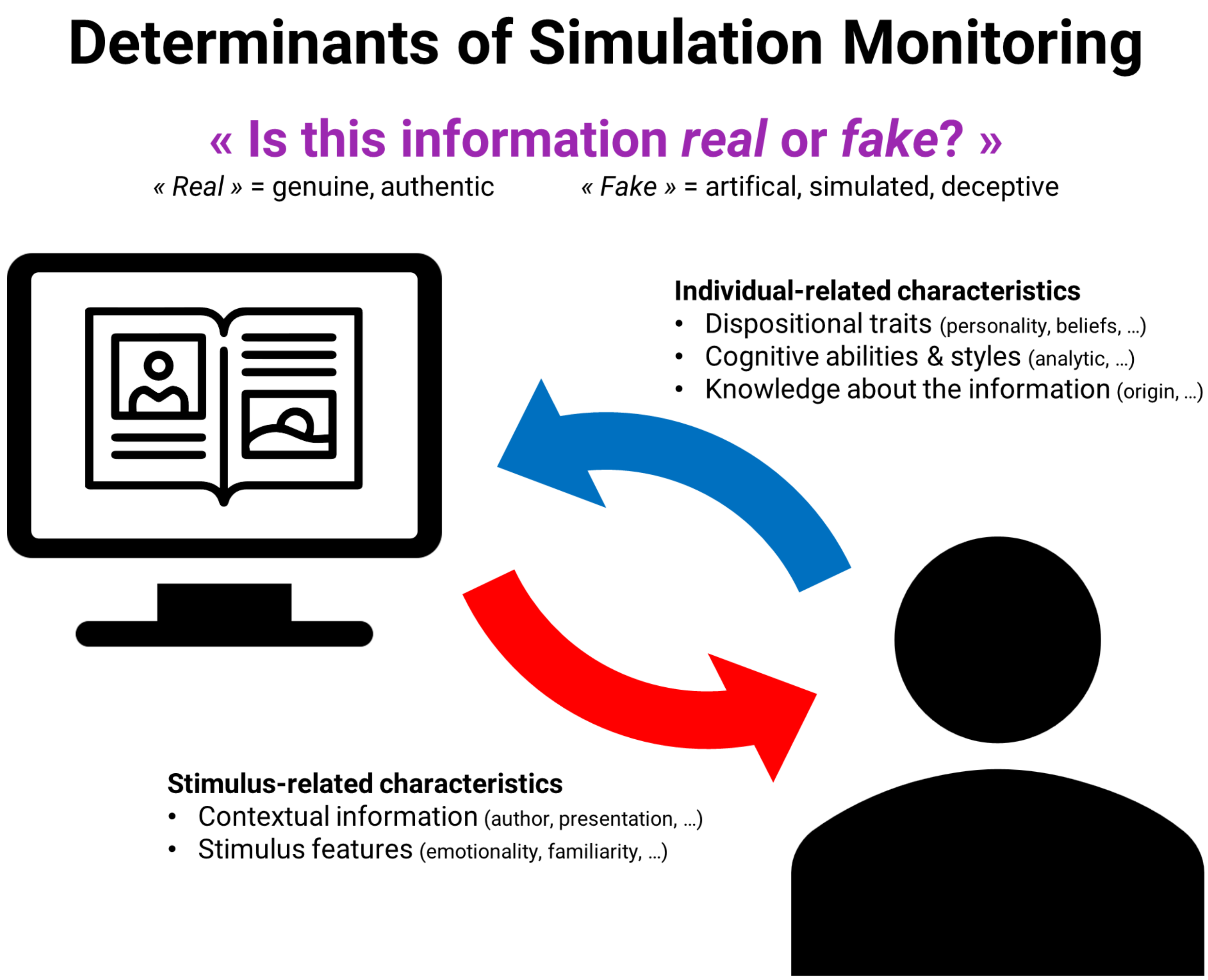

Figure 1. The decision to believe that an ambiguous stimulus (of any form, e.g., images, text, videos, environments, ...) is real or fake depends of individual characteristics (e.g., personality and cognitive styles), stimulus-related features (context, emotionality), and their interaction, which can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

Moreover, evidence from research indicates that inter-individual characteristics also play a crucial role in the formation of beliefs of reality, with factors such as cognitive style, prior beliefs, and personality traits significantly impacting simulation monitoring (Bryanov & Vziatysheva, 2021; Ecker et al., 2022; Sindermann et al., 2020). For instance, individuals with higher levels of analytical reasoning have been found to better discriminate real from fake stimuli (Pehlivanoglu et al., 2021; Pennycook & Rand, 2019). 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 our 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 & Slater, 2005). 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). Additionally, other factors, such as the stimuli's perceived self-relevance (T. R. Goldstein, 2009; Sperduti et al., 2016), as well as familiarity (Begg et al., 1992), could also play a role in our processing and reaction to real as opposed to non-real material.

AI-generated images of faces, due to their popularity as a target of CGI technology and the possibility of experimentally manipulating facial features, are increasingly used to study face processing as related to 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, some studies report that when the nature of the faces was ambiguous, artificially created faces that were previously rated as more attractive were judged by subjects to be less real (Tucciarelli et al., 2020). However, as the attractiveness ratings were given by independent raters instead of the participants, the direction of the relationship between perceived realness and attractiveness cannot be concluded. To this end, Liefooghe et al. (2022) reports that attractiveness ratings differed significantly between participants who were told that the faces were AI-generated from 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 could drive simulation monitoring, has received little attention to date.

This study primarily aims at exploring the effect of facial attractiveness on simulation monitoring, i.e., on the beliefs that an image is real or artificially generated. Based on the embodied reality theory (outlined in Makowski, 2018; Makowski, Sperduti, et al., 2019), which suggests that salient and emotional stimuli are perceived to be more real, 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 of dispositional traits, such as personality and attitude towards AI, with inter-individual simulation monitoring tendencies. Note that the discriminative accuracy between “true” photos and “true” artificially-generated images is not relevant for this study, which focuses on the beliefs that a stimulus is real or fake, independently of its true nature.

Methods

In line with open-science standards, all the material (stimuli generation code, experiment code, raw data, analysis script with complementary figures and analyses, preregistration, etc.) is available at **https://github.com/RealityBending/FakeFace**.

Procedure. In the first part of the study, participants answered a series of personality questionnaires, including 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, 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. Finally, we devised 5 items pertaining to expectations about AI-generated images technology ("I think current Artificial Intelligence algorithms can generate very realistic images"). To lower their saliency and the possibility of it priming the subjects about the task, we mixed these items with 5 items from the general attitudes towards AI scale (*GAAIS*, Schepman & Rodway, 2020). This scale was presented after the social anxiety questionnaires, and 3 attention check questions were embedded in the surveys.

In the second part of this study, 109 photos of neutral-expression faces of real individuals from the validated American Multiracial Face Database (AMFD, (J. M. Chen et al., 2021)) were presented to the participants for 50 0ms 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 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 face images previously seen were AI-generated (the instructions used a cover story mentioning that the research was aimed at validating 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. One hundred and three participants were recruited via *Prolific*, a crowd-sourcing platform recognized for providing high quality data (Peer et al., 2022). 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 3 participants that failed 2 (>= 66%) or more attention check questions. The final sample included 100 participants (Mean age = 27.9, SD = 8.5, range: [19, 66]; Sex: 48% females, 52% males).

Data Analysis. The real-fake ratings (measured originally on a [-1, 1] analog scale) were converted into two scores, corresponding to two conceptually distinct mechanisms: the dichotomous *belief* (real or fake, derived 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 (with the participants and images entered as random factors), which models 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 modeled using beta regression models (suited for outcome variables expressed in percentages).

We started by investigating the effect of the procedure and instructions to check whether the stimuli (which were real pictures of faces) were indeed judged as fake in a 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).

The determinants of reality beliefs were modeled 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 pictures 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, the manuscript will focus on

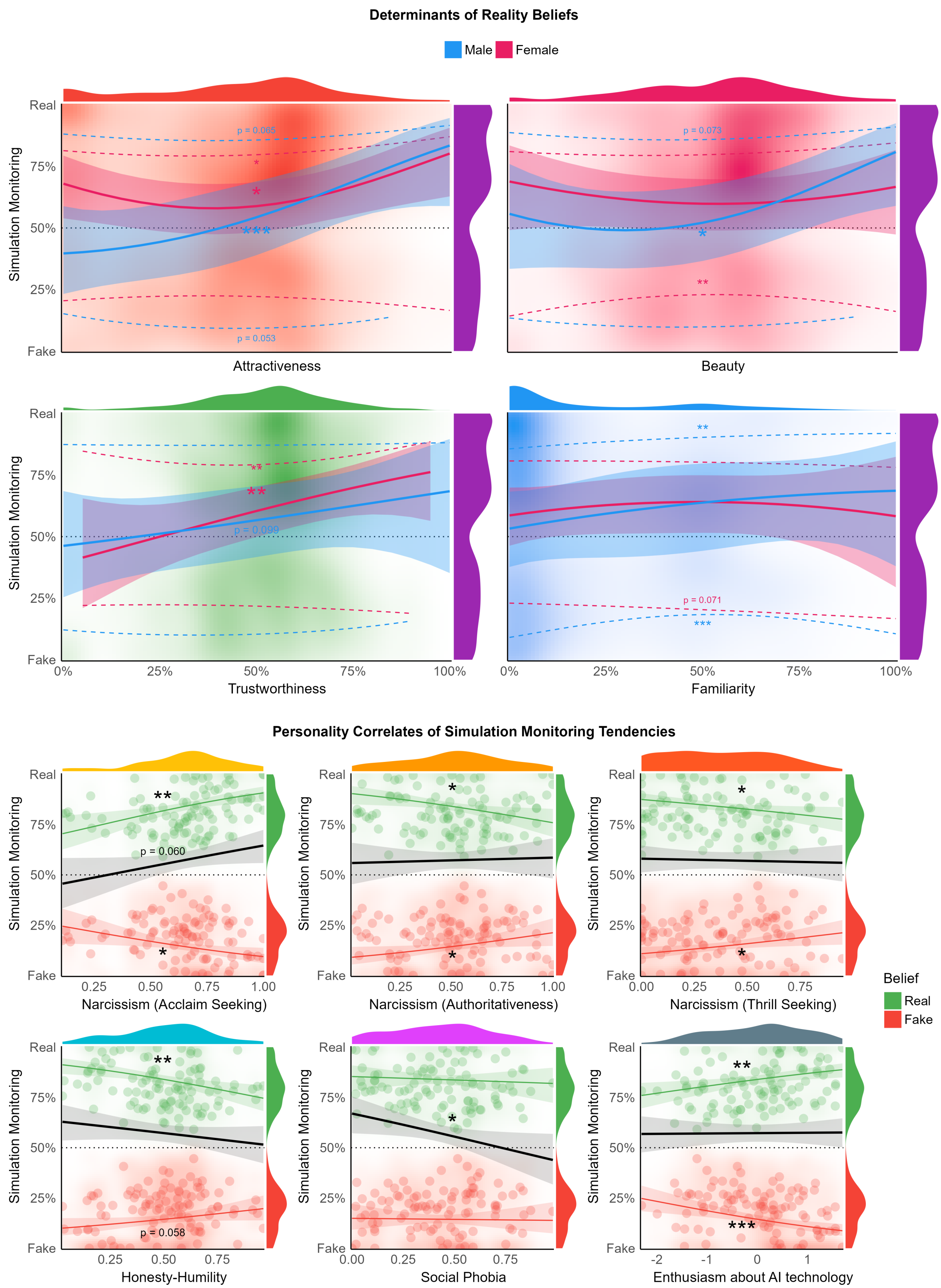

Figure 2. Top part shows the efffect of face ratings on 1) the probability of judging a face as real vs. fake (solid line) and 2) on the confidence associated with that judgment (dashed lines) depending on the sex. Bottom part shows the effect of personality traits on the belief (black line) and the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

significant results.

Results

Manipulation Check. Only one image file yielded a strong simulation monitoring bias (> 85%), being classified as fake in 87.4% of trials. This image was removed from further analysis, leaving 108 trials per participant. On average, across participants, 44% of images (95% CI [0.11, 0.64]) were judged as fake and 56% of images (95% CI [0.36, 0.89]) as real. An intercept-only model with the 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 10.5% for the participants and 8.7% for the pictures.

There was a significant negative effect of the delay of re-exposure (with 95% of values between 1.58 and 30.31 min), suggesting that shorter delays were associated with a slight bias towards the belief of reality (60% at a theoretical delay of 0), which decreased to 50% at a theoretical delay of 60 min ($OR = 0.99$, $95\%CI = [0.99, 1.00]$, $z = -2.27$, $p = .023$). There was also a significant negative effect on judgment confidence, but only in the real condition ($\beta = -0.005$, $95\%CI = [-0.1, 0.0]$, $p = .023$).

Determinants of Simulation Monitoring. Attractiveness had a significant positive and linear relationship ($R^2_{marginal} = 2.8\%$) with the belief that a stimulus was real ($\beta_{poly1} = 16.37$, $95\%CI = [7.76, 24.98]$, $z = 3.73$, $p < .001$) for males, and a quadratic relationship for females ($\beta_{poly2} = 7.77$, $95\%CI = [1.41, 14.13]$, $z = 2.40$, $p = .017$), with both non-attractive and attractive faces being judged as more real. No significant relationship was found between attractiveness ratings and belief confidence, aside of a similar trend for females only, for faces judged as real ($\beta_{poly2} = 4.38$, $95\%CI = [0.96, 7.79]$, $z = 2.51$, $p = .012$). There was no interaction with reported self-attractiveness.

Beauty, adjusted for trustworthiness and familiarity, had a significant positive and linear relationship ($R^2_{marginal} = 3.5\%$) with the belief that a stimulus was real ($\beta_{poly1} = 9.54$, $95\% CI = [1.43, 17.65]$, $z = 2.31$, $p = .021$) for males only. No effect on confidence was found, aside from a quadratic relationship for females for faces judged as fake, suggesting that non-beautiful and highly beautiful faces were rated as fake with more confidence than average faces ($\beta_{poly2} = 6.61$, $95\% CI = [1.98, 11.24]$, $z = 2.80$, $p = .005$). There was no interaction with reported self-attractiveness.

Trustworthiness, adjusted for beauty and familiarity, had a significant positive and linear relationship ($R^2_{marginal} = 3.0\%$) with the belief that a stimulus was real ($\beta_{poly1} = 11.60$, $95\% CI = [4.15, 19.06]$, $z = 3.05$, $p = .002$) for females only. No effect on confidence was found, aside from a quadratic relationship for females for faces judged as real, suggesting that non-trustworthy and highly trustworthy faces were rated as real with more confidence than average faces ($\beta_{poly2} = 6.47$, $95\% CI = [1.73, 11.21]$, $z = 2.68$, $p = .007$).

We did not find any significant relationships for familiarity adjusted for beauty and trustworthiness ($R^2_{marginal} = 3.0\%$). However, a significant positive and linear relationship was found with the confidence in faces judged as real ($\beta_{poly1} = 9.31$, $95\% CI = [3.45, 15.17]$, $z = -3.11$, $p = .002$), and a quadratic relationship for faces judged as fake ($\beta_{poly1} = -12.67$, $95\% CI = [-19.87, -5.47]$, $z = -3.45$, $p < .001$; $\beta_{poly2} = 8.14$, $95\% CI = [0.01, 16.28]$, $z = 1.96$, $p = .05$), for males only, suggesting that faces are judged as real with more confidence when they are familiar, and judged as fake with less confidence when they are of not familiar or highly familiar.

Inter-Individual Correlates of Simulation Monitoring. The models including the personality traits suggested that *Honesty-Humility* had a significant negative relationship predominantly with the confidence associated with faces judged as real ($\beta_{real} = -1.69$, $95\% CI = [-2.74, -0.63]$, $z = -3.14$, $p < .01$; $\beta_{fake} = -1.02$, $95\% CI = [-2.07, 0.03]$, $z = -1.90$, $p = 0.058$). Narcissism shared significant links through various dimensions, such as positive relationship between the confidence for both real and fake judgments and *Acclaim Seeking* ($\beta_{real} = 2.08$, $95\%CI = [0.79, 3.37]$, $z = 3.16$, $p < .01$; $\beta_{fake} = 1.60$, $95\%CI = [0.30, 2.90]$, $z = 2.42$, $p < .05$), and a negative relationship with *Authoritativeness* ($\beta_{real} = -1.38$, $95\%CI = [-2.50, -0.27]$, $z = -2.43$, $p < .05$; $\beta_{fake} = -1.23$, $95\%CI = [-2.35, -0.11]$, $z = -2.15$, $p < .05$) and *Thrill Seeking* ($\beta_{real} = -0.93$, $95\%CI = [-1.75, -0.10]$, $z = -2.20$, $p < .05$; $\beta_{fake} = -1.03$, $95\%CI = [-1.87, -0.20]$, $z = -2.44$, $p < .05$). Finally, the social phobia dimension was significantly associated with a decreased probability of judging faces as fake ($\beta = -0.97$, $95\%CI = [-1.86, -0.08]$, $z = -2.14$, $p < .05$), but no effect on the confidence judgments was found. No significant relationships was found for intolerance to uncertainty or paranoid beliefs.

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). Only AI-Enthusiasm displayed a significant positive relationship with the confidence in both real and fake judgments ($\beta_{real} = 0.29$, $95\%CI = [0.08, 0.50]$, $z = 2.74$, $p < .01$; $\beta_{fake} = 0.38$, $95\%CI = [0.18, 0.59]$, $z = 3.60$, $p < 0.001$).

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 the stimulus is artificially generated. The most striking result, in our opinion, is that despite all the stimuli being real faces from the same database, all participants, when given the information, believed (to high degrees of confidence) that a significant proportion of them were fake. This finding is a testimony to both the current expectations regarding CGI technology in the population, as well as to the volatility of our sense of reality. It underlines the strong impact of prior expectations and information on reality beliefs. In fact, stimuli-related and participant-related characteristics accounted for less than 20% of the beliefs variance, suggesting that a large part of it is associated with 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 a different pattern 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 judgments, and its role in simulation monitoring, are a multidimensional construct that cannot be reduced to physical facial attractiveness, in particular for women [**REF?**].

Indeed, 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 (Fink et al., 2006; Qi & Ying, 2022). However, while it is possible that the evolutionary weight associated with attractiveness skewed the perceived saliency towards attractive faces, rendering them significantly more salient than unattractive faces, 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 (in particular, heart rate deceleration, A. Goldstein et al., 2002).

Our results suggested that trustworthiness ratings linearly increased beliefs in realness, for females only. Given the evidence suggesting that faces presented as real were rated as more trustworthy (Balas & Pacella, 2017; Hoogers, 2021; Liefooghe et al., 2022), we expected such link to be present for both genders. One of the contributing mechanism to this dimorphism could have been 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. As such, further studies are needed to investigate the causes of the increased simulation monitoring sensitivity to trustworthiness in females.

Contrary to our hypothesis, familiarity was not found to be significantly related to simulation monitoring decisions. However, the distribution of familiarity ratings was strongly skewed, and only a low number of pictures was rated as highly familiar. Future studies could re-examine this aspect by experimentally manipulating familiarity, for instance by modulating the amount of exposure to items before the simulation monitoring judgments.

Regarding the role of inter-individual characteristics in simulation monitoring tendencies, we found that participants with stronger narcissism (in particular the acclaim seeking and grandiose fantasies dimensions) displayed a stronger confidence in their simulation monitoring judgments. This is in line with the literature

While early researchers had proposed narcissism to be a unitary construct (Raskin & Terry, 1988; Sedikides et al., 2004), recent studies cast narcissism as encompassing 2 unique dimensions - namely narcissistic vulnerability, which is characterized by low self-esteem and a need for admiration, and narcissistic grandiosity, which encompasses high self-esteem, arrogance and aggressive self-assertiveness (Pincus & Roche, 2011). Whereas narcissism has long been linked to over-confidence in decision-making (Campbell et al., 2004), a growing body of research indicates that this relationship is likely mostly driven by facets of narcissistic grandiosity (such as Acclaim Seeking), which drives individuals to see themselves as more superior than others and to externalize blame for poor decision-making consequences (O'Reilly & Hall, 2021). Indeed, the evidence alludes that attributes specific to narcissistic grandiosity predict individuals' overconfidence in cognitive decision-making (Brunell & Buelow, 2017; Chatterjee & Pollock, 2017; O'Reilly & Hall, 2021).

Conversely, the literature on the correlation between honesty-humility and perceptual confidence judgements remain comparatively scarce, albeit some existing research links higher levels of honesty-humility to increased perceptions of risk and greater risk aversion (Levidi et al., 2022; Weller & Thulin, 2012). As such, it could be argued that individuals with higher levels of honesty-humility could inherently possess a greater simulation monitoring sensitivity, perceiving a higher risk involved in making decisions on realness and therefore be less confident of their beliefs of reality. This is further supported by convergent evidence finding a negative link between honesty-humility and narcissism (Hodson et al., 2018). Therefore, further research examining the relationship between simulation monitoring decisions, personality traits, particularly that related to the dark triad and honesty-humility, and the perceived risks involved in making such evaluations, is warranted.

Interestingly, despite the ubiquity of AI, there is a lack of research pertaining to the influence of people's AI attitudes on their confidence in their discriminative ability to distinguish between output that is "true" (i.e., images of real faces) from those that is artificially-generated. One possible explanation for the positive relationship observed between AI-Enthusiasm and confidence in reality judgements could be that individuals with highly positive AI attitudes may perceive themselves as having greater knowledge about AI and its capabilities, hence permitting them to be more confident in their simulation monitoring decisions. This is in line with previous research findings, which indicate that greater enthusiasm about AI is related to having a higher confidence in one's knowledge regarding AI competencies (Said et al., 2022). Moreover, Said et al. (2022) further reports that AI attitudes interacts with people's perceived self-knowledge to influence their perception of the opportunities and risks accorded by AI applications. Specifically, people with AI attitudes that were highly positive and that were more confident in their AI knowledge were more likely to over-estimate the advantages of AI and under-estimate the risks involved in their application. As such, it is plausible that individuals' perceived self-knowledge about AI could have modulated the relationship between their attitudes towards AI and their confidence in their reality judgements.

Although the order of presentation of the facial images was randomized to reduce effects of adaptation, re-exposure delay was found to have a significant negative effect on simulation monitoring decisions, i.e., shorter delays were associated with faces being rated as more real. This may be because shorter re-exposure delays could have led to the faces being better remembered and appearing more familiar than faces that were displayed after a longer delay, thereby triggering self-referential and autobiographical memory processing during the repeated display (Abraham & Von Cramon, 2009; Gobbini et al., 2013; Taylor et al., 2009). Although the distribution variance in re-exposure delay is small, recent research using event-related brain potentials report representations of faces to be qualitatively similar regardless of the degree and type of familiarity (e.g., personal or media-based), only distinguishing between familiar and unfamiliar faces (Wiese et al., 2021). Moreover, fictional stimuli has been shown to up-regulate emotions when associated with familiarity (Makowski et al., 2017; Sperduti et al., 2016), thus biasing its salience and perceived realness. Future studies could hence further investigate the modulatory effects of types and degrees of familiarity on perceived realness judgements.

In conclusion, research on our ability to discriminate between what is real from what is not is increasingly relevant given the current expansion on technologies related to virtual and augmented realities. The aim of the present study was to examine whether facial attractiveness can significantly influence our simulation monitoring decisions. In line with our hypothesis, we found faces rated as attractive and unattractive to be perceived as more real for females as a result of an increased salience. However, for males, the observation of a positive linear relationship between attractiveness and beauty with simulation monitoring decisions, adjusted for evolutionarily advantageous features such as trustworthiness, suggest a more complex relationship than hypothesized. As such, future studies could further investigate the role of attractiveness in simulation monitoring, taking greater consideration of the former's multidimensionality.

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.

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.

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.

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.

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.

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.

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.

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.

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, A., Spencer, K. M., & Donchin, E. (2002). The influence of stimulus deviance and novelty on the P300 and novelty P3. *Psychophysiology*, *39*(6), 781–790.

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.

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.

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.

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.

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.

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.

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.

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.

Pincus, A. L., & Roche, M. J. (2011). Narcissistic grandiosity and narcissistic vulnerability. *The Handbook of Narcissism and Narcissistic Personality Disorder: Theoretical Approaches, Empirical Findings, and Treatments*, 31–40.

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/

Raskin, R., & Terry, H. (1988). A principal-components analysis of the narcissistic personality inventory and further evidence of its construct validity. *Journal of*

Personality and Social Psychology, *54*(5), 890.

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.

Sedikides, C., Rudich, E. A., Gregg, A. P., Kumashiro, M., & Rusbult, C. (2004). Are normal narcissists psychologically healthy?: Self-esteem matters. *Journal of Personality and Social Psychology*, *87*(3), 400.

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.

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.

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.

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.

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

Wiese, H., Hobden, G., Siilbek, E., Martignac, V., Flack, T. R., Ritchie, K. L., Young, A. W., & Burton, A. M. (2021). Familiarity is familiarity is familiarity: Event-related brain potentials reveal qualitatively similar representations of personally familiar and famous faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.