Computers in Human Behavior Reports 17 (2025) 100563

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

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

Synthetic and natural face identity processing share common mechanisms

Kim Uittenhove [a,b], Hatef Otroshi Shahreza [c,d], Sébastien Marcel [c,e], Meike Ramon [b,f,*]

[a] Center for Learning Science, EPFL, Lausanne, Switzerland
[b] Applied Face Cognition Lab, Institute of Psychology, University of Lausanne, Switzerland
[c] Idiap Research Institute, Martigny, Switzerland
[d] School of Engineering, EPFL, Lausanne, Switzerland
[e] School of Criminal Justice, University of Lausanne, Switzerland
[f] AIR – Association for Independent Research, Zurich, Switzerland

ARTICLE INFO

Keywords:
Face identity processing
Natural and synthetic faces
Deepfakes
Stimulus similarity
Human ability
Individual differences
Face recognition
Face discrimination

ABSTRACT

Recent developments in generative AI offer the means to create synthetic identities, or deepfakes, at scale. As deepfake faces and voices become indistinguishable from real ones, they are considered as promising alternatives for research and development to enhance fairness and protect humans' rights to privacy. Notwithstanding these efforts and intentions, a basic question remains unanswered: Are natural faces and facial deepfakes perceived and remembered in the same way? Using images created via professional photography on the one hand, and a state-of-the-art generative model on the other, we investigated the most studied process of face cognition: perceptual matching and discrimination of facial identity. Our results demonstrate that identity discrimination of natural *and* synthetic faces is governed by the same underlying perceptual mechanisms: objective stimulus similarity and observers' ability level. These findings provide empirical support both for the societal risks associated with deepfakes, while also underscoring the utility of synthetic identities for research and development.

1. Introduction

Traditionally, vision research involved the use of analogue images to study level low-, mid-, or high-level visual processes (Benton et al., 1983; Snodgrass & Vanderwart, 1980; Weber, Ross, & Murray, 2018). Over time, novel tools to create, manipulate, and display digital stimuli emerged, which were readily adopted, particularly in the field of face processing. In the early 2000s, commercial photo editing software provided means to achieve (even unnoticeable) facial manipulations (Ramon et al., 2016). Early generative software enabled creation of images, however these could easily be detected as artificial as opposed to natural images of real people (FaceGen Modeller, 2009; Frowd, Hancock, & Carson, 2004; Frowd et al., 2005).

1.1. Synthetic faces: Status Quo

Recent advances in generative AI have fundamentally changed this situation, as well as society at large. In the visual domain, Generative Adversarial Networks (GANs) and Diffusion Models (DMs) enable rapid creation of synthetic facial identities, or deepfakes, at scale (Farid, 2022; Groh et al., 2021). Most importantly, however, the *quality* of deepfakes has increased dramatically. Contrary to early approaches, state-of-the-art (SOTA) models can generate extremely realistic, i.e. natural-looking synthetic identities, and for a given synthetic identity produce images involving viewpoint and age-related changes (Kammoun et al., 2022).

Alongside advances in deepfake *generation*, a growing body of research is focused on machine-based solutions for automatic deepfake *detection* (Wang et al., 2022). *Human* deepfake processing on the other hand remains understudied. The limited number of empirical investigations collectively demonstrate that humans cannot reliably distinguish natural from synthetic faces (Farid, 2022; Groh et al., 2021; Lago et al., 2021; Ramon, Vowels, & Groh, 2024). Law enforcement professionals, who could be tasked with analyzing digital content to verify its authenticity, are challenged by deepfakes – even those with superior face processing skills (Ramon et al., 2024). Moreover, interventions and training regimes for deepfake detection do not lead to significant improvements (Bray, Johnson, & Kleinberg, 2022; Nightingale & Farid, 2022).

* Corresponding author. Applied Face Cognition Lab, Institute of Psychology, University of Lausanne, Switzerland.
E-mail address: meike.ramon@gmail.com (M. Ramon).

https://doi.org/10.1016/j.chbr.2024.100563
Received 3 September 2024; Received in revised form 4 November 2024; Accepted 7 December 2024
Available online 17 December 2024

1.2. Challenges and opportunities

Our inability to distinguish natural from synthetic identities harbors both substantial negative and positive potential across different areas of society. On the one hand, deepfakes can be used to deceive and exploit aiding crimes such as fraud and revenge porn(ography) (Damiani, 2019; Delfino, 2019). At the societal level, digital proliferation of AI-generated disinformation campaigns can e.g. interfere with political processes and decrease trust in governments and institutions (Chesney & Citron, 2018; Vaccari & Chadwick, 2020). International law enforcement agencies regard synthetic information and deepfakes as a major challenge and a threat to society (DoHSP-Pae, 2021; Lab, 2022). This situation is compounded by the difficulty in prosecuting the creation and dissemination of deepfakes due to inadequate legal frameworks (Chesney & Citron, 2018). A positive effect of the ability to create indistinguishable-from-natural content is the support of creative avenues for artists and media professionals (Epstein et al., 2023). For example, synthetic avatars are commonplace in the gaming industry and various types of deepfakes are used in the film industry (Li, 2024).

In the scientific community, considerable efforts are currently directed towards the creation of synthetic face datasets (Boutros et al., 2023; Colbois, de Freitas Pereira, & Marcel, 2021; Geissbühler, Shahreza, & Marcel, 2024; Kim et al., 2023; Melzi et al., 2023; Otroshi-Shahreza et al., 2024; Shahreza & Marcel, 2024), which offer at least two main advantages. First, synthetic datasets are devoid of the legal, ethical, and privacy concerns associated with datasets of real identities, which are often scraped from internet without individuals' consent, and are used to train face recognition models. Note that several of such web-scraped face datasets, such as VGGFace2 (Cao et al., 2018) and MS-Celeb-1M (Guo et al., 2016) have been retracted by their owners due to privacy concerns (The rise and fall, 2022). Second, synthetic datasets could also save resources, e.g., those associated with the creation of shareable, fair natural databases. Recent advances in SOTA generative models, e.g. (Chan et al., 2022; Karras et al., 2021; Rombach et al., 2022), can provide tools to generate complex synthetic datasets with sufficient real-life variations. For instance, generative models based on Neural Radiance Fields (NeRF) (Mildenhall et al., 2021), such as EG3D (Chan et al., 2022), are capable of 3D identity representations. This allows to generate not only distinct facial identities, but also create any desired viewpoint of a given synthetic identity. In the scientific community, considerable efforts are currently directed towards the creation of synthetic face datasets (Boutros et al., 2023; Colbois et al., 2021; Geissbühler et al., 2024; Kim et al., 2023; Melzi et al., 2023; Otroshi-Shahreza et al., 2024; Shahreza & Marcel, 2024), which offer at least two main advantages. First, synthetic datasets are devoid of the legal, ethical, and privacy concerns associated with datasets of real identities, which are often scraped from internet without individuals' consent, and are used to train face recognition models. Note that several of such web-scraped face datasets, such as VGGFace2 (Cao et al., 2018) and MS-Celeb-1M (Guo et al., 2016) have been retracted by their owners due to privacy concerns (The rise and fall, 2022). Second, synthetic datasets could also save resources, e.g., those associated with the creation of shareable, fair natural databases. Recent advances in SOTA generative models, e.g. (Chan et al., 2022; Karras et al., 2021; Rombach et al., 2022), can provide tools to generate complex synthetic datasets with sufficient real-life variations. For instance, generative models based on Neural Radiance Fields (NeRF) (Mildenhall et al., 2021), such as EG3D (Chan et al., 2022), are capable of 3D identity representations. This allows to generate not only distinct facial identities, but also create any desired viewpoint of a given synthetic identity. In the scientific community, considerable efforts are currently directed towards the creation of synthetic face datasets (Boutros et al., 2023; Colbois et al., 2021; Geissbühler et al., 2024; Kim et al., 2023; Melzi et al., 2023; Otroshi-Shahreza et al., 2024; Shahreza & Marcel, 2024), which offer at least two main advantages. First, synthetic datasets are devoid of the legal, ethical, and privacy concerns associated with datasets of real identities, which are often scraped from internet without individuals' consent, and are used to train face recognition models. Note that several of such web-scraped face datasets, such as VGGFace2 (Cao et al., 2018) and MS-Celeb-1M (Guo et al., 2016) have been retracted by their owners due to privacy concerns (The rise and fall, 2022). Second, synthetic datasets could also save resources, e.g., those associated with the creation of shareable, fair natural databases. Recent advances in SOTA generative models, e.g. (Chan et al., 2022; Karras et al., 2021; Rombach et al., 2022), can provide tools to generate complex synthetic datasets with sufficient real-life variations. For instance, generative models based on Neural Radiance Fields (NeRF) (Mildenhall et al., 2021), such as EG3D (Chan et al., 2022), are capable of 3D identity representations. This allows to generate not only distinct facial identities, but also create any desired viewpoint of a given synthetic identity.

In practice, however, creating synthetic datasets that mirror the variability of natural faces remains a major challenge. Put simply, natural and synthetic faces may differ in terms of their respective feature dimensions, and in turn their theoretical multidimensional face spaces (Valentine, 2001). Realistically, a formal comparison of the *full* range of human and synthetic variability is, of course, not feasible. However, a thorough and fair comparison of subsets of natural and synthetic faces *can* be achieved via in-depth investigation of human performance, under careful consideration of its known determinants.

1.3. Determinants of natural face discrimination

Arguably, the most challenging computations performed by the human brain are those required to discriminate unfamiliar faces (Wichmann & Geirhos, 2023). Deciding whether two images show the same person, or two different people is an evolutionarily recent task (i.e., emerged with analogue photography). Therefore, unsurprisingly, but often overlooked is the fact that this task is extremely error-prone (Fysh & Bindemann, 2018; Tummon, Allen, & Bindemann, 2019) —even for highly trained professionals (Papesh, 2018). Observers' face identity matching performance is determined on the one hand by objective external factors, e.g. stimulus similarity and viewing conditions such as exposure duration and resolution (Fysh & Ramon, 2021; Ramon et al., 2015; Ramon & van Belle, 2016). Moreover, studies demonstrate a high degree of variation in face identity processing (FIP) ability among individual observers (Bobak et al., 2023; Fysh, Stacchi, & Ramon, 2020; Ramon, 2021; Ramon, Bobak, & White, 2019; Stacchi et al., 2020), which can only be partially attributed to stable, inter-individual differences (Bobak et al., 2023; Fysh et al., 2020; Ramon, 2021). Overall, the same external conditions (i.e., experience, stimulus quality) affect individuals differently, depending on their respective location on the FIP ability spectrum (Ramon & Rjosk, 2022).

1.4. Synthetic and natural face discrimination: shared cognitive mechanisms?

In this study we sought to answer the currently open question of whether synthetic identity processing is governed by the same mechanisms that determine natural face processing. To this end, we investigated the impact of stimulus similarity and differences in observers' FIP ability in two experiments probing rapid natural and synthetic face discrimination. Uniquely, we use two different types of face stimuli, derived from best-case real-world and generative scenarios. First, natural identities were represented via professional photographs derived from an artistic project that captured thousands of identities under strictly controlled and thus comparable conditions. Second, synthetic identities were created using a SOTA generative model that can achieve 3D viewpoint-invariant identity representations (Chan et al., 2022). We aimed to investigate, for the first time to our knowledge, whether behavioral facial identity matching assessed using stimuli created with SOTA generative models mirrors that observed for natural faces.

2. Methods

All procedures and protocols were approved by the Ethics Committee of the University of Fribourg (approval number 473), the University of Lausanne, and conducted in accordance with the guidelines set forth in the Declaration of Helsinki. All volunteering participants provided informed consent, and were not financially compensated for their participation.

2.1. Participants

Data collection was distributed via students enrolled in a seminar at the University of Lausanne, who recruited up to five participants each. Participants (N = 104) completed two face discrimination experiments, in which natural or synthetic stimuli were used. The majority (n = 97) then completed an independent assessment of face identity processing ability (see below). Of these 97 complete data sets, seven were excluded due to exceedingly elevated reaction times across both face discrimination experiments (>3rd quartile + 1.5* interquartile range, n = 5), or low accuracy (<1st quartile - 1.5* interquartile range, n = 2). The final sample subject to analyses comprised 90 participants (41 and 59 identifying as fe-/male, mean age: 32, SD = 14, range: 18 to 62), 80 of which were right-handed (7 left-handed, 3 ambidextrous).

2.2. Synthetic and natural face discrimination

All participants completed two face discrimination experiments, which involved the same design, but presented either natural or synthetic face images as stimulus material. These experiments were delivered via Testable (Rezlescu et al., 2020); order of completion was counterbalanced across participants. Participants performed 2-alternative forced-choice (2AFC) identity matching for pairs of stimuli presented in rapid succession (see Fig. 1), indicating their responses via (bi-manual) button-press. Target and probe stimuli were presented for 500ms each, with a 500ms inter-stimulus interval; observers had unlimited time to indicate their decision. Targets' and probes' location was centrally offset to prevent the use of local matching strategy (−/+25 pixels vertically and horizontally from the center in four quasi-random locations). Each experiment comprised 420 trials (210 *same/different*), across which different combinations of 21 White female identities were presented. *Same* trials were generated by pairing five different images of the same identity with one another. *Different* trials were generated by pairing each of the 21 identities with every other identity. Each face discrimination experiment began with four practice trials, which displayed identities not included in the experiment, and which were excluded from analyses.

2.2.1. Natural face stimuli

Natural identities were selected from the FACITY image database created and kindly provided by the professional photographer H. Caspar. This database was created as an artistic project, through which hundreds of professional photographers worldwide took images of thousands of people following the same criteria (i.e. under the same conditions, to ensure comparable image quality and style; see Fig. 1a). From this large database, 21 target identities were selected, all of which were devoid of obvious features (e.g., piercings, moles, scars, etc.) that could aid rapid matching based on specific image features, as opposed to identity matching. Fig. 1a shows the set of natural identities used in the

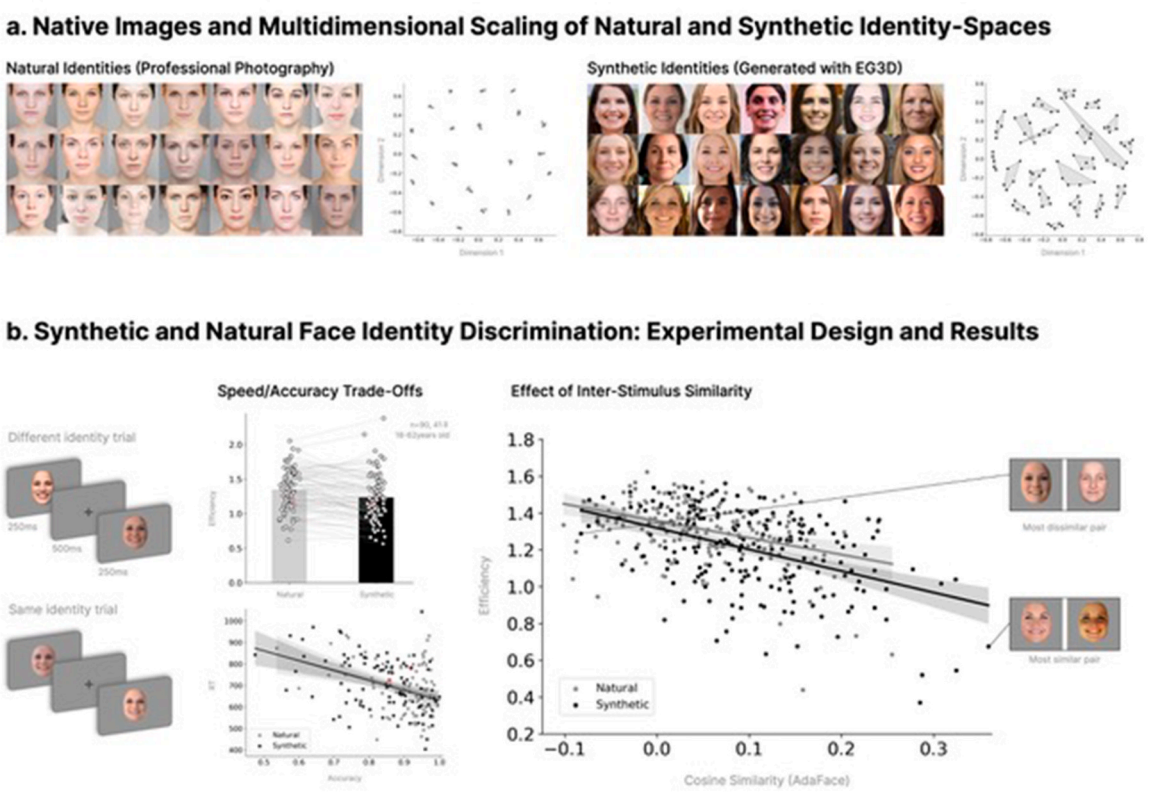

Fig. 1. Natural and Synthetic Face Discrimination. a. *Examples of native stimuli and multi-dimensional scaling of face space*. Images of identities were either created via professional photography (left), or generated using EG3D (Chan et al., 2022), a model based on Generative Adversarial Network and Neural Radiance Fields structure (right). **b.** *Schematic of face discrimination experiments and observed relationship between performance measures across experiments.* Examples of different and same identity trials (left), and individual observers' overall efficiency and RT as a function of accuracy across experiments. Red points indicate performance of the only observer who met the proposed criteria for Super-Recognizer identification (Ramon, 2021). Effect of inter-stimulus similarity computed using AdaFace on face discrimination efficiency (right) with most dis-/similar image pairs.

natural face discrimination experiment. The selected original images were then processed to create the final stimuli for the natural face discrimination experiment. The midpoint of each face was vertically offset from the center by approximately 10% of the image height. To ensure that natural and synthetic stimuli occupied the same image space in both dimensions, we applied padding to the top of the image (20% of the image height) and to either side of the image (10% of the image width). The resulting images were centered on the midpoint of the face, which (consistent with synthetic images) occupied ca. 55% of the horizontal space. These images were resized to their original size (512×512 pixels, 96 dpi). For each identity we created four modified versions by in-/decreasing the contrast and luminosity of the original image. To prevent the use of external features and contour information, faces were cropped and edges were blurred using the Pyfacer library, with pre-trained models for face detection (Deng et al., 2020) and face parsing (Zheng et al., 2022). We created a binary mask to isolate the facial region (removing a section of the top to resolve artifacts around the top of the natural images, due to the nature of their creation in the artistic database). To extract the edges of the facial region mask, we checked each pixel of the mask (value = 1) and marked this pixel as part of the edge if any of the eight surrounding pixels was not part of the facial mask (value = 0). We defined an edge zone by dilating the extracted edge using a 15×15 kernel. For a natural blending effect, we first applied a 29×29 Gaussian kernel to the face region, creating a blurred version. We then recombined the blurred image with the original image: pixels within the edge zone were replaced with pixels from the blurred version, while pixels outside this zone retained their original values. Finally, we introduced further variability by scaling down the face size and applying a rotation to the four additional instances of the original image. Stimuli were presented on a full-screen grey background; a calibration prior to each experiment ensured comparable on-screen stimulus size across devices.

2.2.2. Synthetic face stimuli

Synthetic identities were generated using EG3D (Chan et al., 2022). EG3D is a face generator model based on Generative Adversarial Network (GAN) and Neural Radiance Fields (NeRF) structure. It starts from a random $\boldsymbol{z} \in \mathcal{N}(0, I)$ to generate an intermediate latent code $\boldsymbol{w} \in \mathscr{W}$ that is fed a long with camera parameters to generate a face image from desired point of view. Starting from different random noise, we generated several facial identities with frontal pose, retaining only those whose appearance was female White within the range of ca. 20–40 years. Per synthetic identity, multiple instances were generated by adding random noise to the intermediate latent code $\boldsymbol{w}$ of the reference instance. The resulting different images per identity could vary in their visual characteristics and facial expressions. Identity correspondence across these images was ensured using the pre-trained face recognition model ArcFace (Deng et al., 2019). To this end, similar to the typical operation of face verification systems (Jain et al., 2004, 2006), we extracted ArcFace feature vectors from each generated instance and the reference instance. Then, if the cosine similarity of the feature vectors for the generated and reference instance was greater than a predefined threshold, we considered the new instance as the same identity. Only images devoid of paraphernalia (e.g., glasses or hats) and artifacts (e.g, on the skin or eyes) were selected, and processed (cropped, contour-blurred) in the same manner as described above for the natural face images.

2.2.3. Multi-dimensional scaling of face stimuli

The calculation of pairwise similarity between *n* stimuli places each stimulus in an *n*-dimensional space where its position is determined by its dis-/similarity to all other stimuli. Multi-dimensional scaling (MDS) allows us to create a lower-dimensional embedding to visualize the dis-/similarity between stimuli in a simple way, as shown in Fig. 1a. First, we transformed our similarity matrix into a dissimilarity matrix by subtracting the similarity values from 1. Then, we created a two-dimensional embedding using the MDS function from the scikit learn manifold package (Pedregosa et al., 2011). Larger distance between points reflects more dissimilarity between the stimuli. Stimuli representing the same identity are closer in space (especially for natural stimuli; see above); for synthetic identities, stimuli representing the same identity are connected to form grey zones.

2.3. Assessment of participants' face identity processing ability

Participants were invited to participate in independent tests of face perception and memory ability. The tests included the long version of the Yearbook Test (YBT (Bruck, Cavanagh, & Ceci, 1991; Stacchi et al., 2020);), and the Cambridge Face Memory Test (CFMT+ (Russell, Duchaine, & Nakayama, 2009);). The YBT probes face perception through unfamiliar face identity matching across image variations and considerable age-related changes in appearance. Participants are presented with 8 (same-gender) panels displaying five (young adult) target identities and 10 (older) probes, half of which represent non-match distractor identities. The CFMT + assesses face memory by soliciting identity recognition for six experimentally learned male White identities in a 3-alternative forced-choice format under increasingly difficult visual conditions. These tests represent sensitive means to assess individual differences in observers' face identity processing ability, and are used to identify so-called Super-Recognizers, i.e. individuals with naturally superior ability for processing facial identity (Nador et al., 2021a, 2021b; Ramon, 2021; Ramon et al., 2024). The YBT was administered as a paper-based test (cf. (Bruck et al., 1991; Stacchi et al., 2020), while the CFMT+ was administered online via Testable (Rezlescu et al., 2020).

2.4. Analyses

Our analyses aimed to answer the question of whether perceptual discrimination of, i.e. telling apart synthetic identities follows the same pattern as that of natural ones. In a first step, we omitted trials associated with exceedingly short (<250ms) or long (> third quartile + 1.5* interquartile range) RTs (considered to reflect anticipatory responses or external disruption). This led to elimination of an average of 6.81% of RTs/trials per participant (SD = 2.43%, range = 2.02–18.93%). To investigate face *discrimination* in the natural (FD_{nat}) and synthetic (FD_{syn}) experiments, we analyzed *different* trials to determine the relationship between performance measures (accuracy, correct RTs, efficiency; see below) and objective inter-stimulus similarity, and individual observers' independently assessed FIP ability, respectively. Bayesian analyses were performed with JASP version 0.18.3, following the approach recommended by (Van Doorn et al., 2021) that allows quantifying evidence for the null and alternative hypotheses. Bayes Factors (BFs) were interpreted as follows: Values between 1 and 1/3 indicated inconclusive evidence favoring H*0*; values between 1/3 and 1/10 provided substantial evidence for H*0*; and values less than 1/10 indicated strong evidence for H*0*. Values between 1 and 3 indicated inconclusive evidence favoring H*a*; values between 3 and 10 provided substantial evidence for H*a*; and values greater than 10 indicated strong evidence for H*a*. To maintain neutrality, we used non-informative priors in all analyses, allowing the data to drive our inferences. Specifically, we conducted Bayesian linear regressions using a JZS prior with an r scale of .354 for the coefficients and a Beta Binomial model prior with parameters a = 1 and b = 1. Our aim was to quantify the evidence for the inclusion of each predictor across different models. For comparisons between conditions, we used a Bayesian Wilcoxon signed-rank test with a Cauchy prior with scale of .707.

2.4.1. Overall comparison of performance for natural and synthetic face matching: efficiency and speed/accuracy relationships

First, we determined whether both experiments were comparable in terms of overall difficulty. To this end, we computed individual

observers' efficiency scores (mean accuracy/mean correct RT *1000) and compared them across experiments. Additionally, we sought to determine the *relationship* between accuracy and speed across experiments and participants. The rationale was that individual observers could in principle prioritize one measure over the other, and do so differently across experiments. Ensuring comparable speed-accuracy trade-offs is necessary to be able to compare efficiency scores and investigate the effects of objective similarity and FIP ability (see below). To this end, we used a Bayesian linear regression on the log of RT as dependent variable, and with accuracy and experiment (natural vs. synthetic), as well as their interaction, as predictors.

2.4.2. Objective stimulus similarity and observers' independently assessed face identity processing ability

In a second step, we investigated the relationship between observers' performance for the FD_{nat} and FD_{syn} experiments in greater detail. We examined observers' performance measures as a function of the pairwise similarity between identities as determined using the pre-trained face recognition model AdaFace (Kim, Jain, & Liu, 2022). AdaFace computes the cosine similarity between the features extracted from the images of each trial (i.e, image pair), resulting in a similarity score ranging from −1 (maximum dissimilarity) to 1 (maximum similarity). This procedure has been used in face verification systems to match a pair of face images. We created similarity score bins of 0.05 and calculated observers' efficiency per bin. Additionally, we determined whether FD_{nat} and FD_{syn} identity discrimination performance systematically varied depending on observers' overall face identity processing (FIP) ability. This overall ability was computed based on their performance across two sensitive tests of face *matching* and *recognition*. Our sample obtained a mean score of 9.49 ($SD = 4.26$) for the YBT and 65.67 for the CFMT+ ($SD = 11.56$). We computed each observer's overall FIP ability scores by averaging their individual z-standardized YBT and CFMT + scores.

3. Results

3.1. Natural and synthetic face discrimination: overall performance

See Table 1 for an overview of mean accuracy, correct RT, and efficiency scores per experiment and trial type. Bayesian Wilcoxon signed-rank test of individual observers' aggregate scores of efficiency yielded strong evidence in favor of an effect of *experiment* (BF10 = 77, W = 3166, Rhat = 1.017). Bayesian linear regression with log *RT* as dependent variable and *accuracy* as independent variable, with *experiment* as moderator, indicated strong evidence for an effect of *accuracy* on *RT* (coefficient: −.256; 95% CI: −.351, −.139; BFinclusion = 11624), reflecting higher accuracy being associated with lower RTs. The analysis yielded evidence against an effect a main effect of *experiment* (coefficient: .0003; 95% CI: −.072, .036; BFinclusion = .278), and against a moderating effect of *experiment* (coefficient: −.002; 95% CI: −.117, .006; BFinclusion = .233). Overall, we find evidence against (differential) speed/accuracy trade-offs across experiments.

3.2. Effects of objective similarity and observers' face processing ability

Fig. 1b displays overall performance for trials showing different identities for each *experiment*, and as a function of stimulus *similarity*. Bayesian linear regression analysis of efficiency scores indicated evidence for an effect of *experiment* (coefficient: −.045; 95% CI: [-.089, .000]; BFinclusion = 3.270), stimulus *similarity* (coefficient: −.831; 95% CI: [−1.017, −.673]; BFinclusion = $6.788 \cdot 10^{23}$), and FIP *ability* (coefficient: .040; 95% CI: [$-5.164 \cdot 10^{-6}$, .067]; BFinclusion = 44.113). We found inconclusive evidence against an interaction between FIP and experiment (*ability* x *experiment*, BFinclusion = .383). Critically, we found evidence against (any) interactions involving the factor *similiarity* (*similarity* x *experiment*, BFinclusion = .165; *ability* x *similarity*, BFinclusion = .208; *ability* x *experiment* x *similarity*, BFinclusion = .034).

4. Discussion

We investigated whether natural and synthetic facial identity discrimination is governed by the same perceptual mechanisms. Uniquely, we used images derived from large-scale databases comprising thousands of identities derived from either professional photography of real people, and an open source SOTA generative model optimized for viewpoint invariance (Chan et al., 2022). For both natural and synthetic images, we quantified the similarity between identities and image pairs (Kim et al., 2022), and assessed observers' individual FIP abilities.

To our knowledge, the present study provides the first empirical evidence that FIP observed for natural facial identities extends to stimuli created with SOTA generative models. Filling this knowledge gap provides three crucial contributions. First, it contributes to the development of the lacking unified framework of natural information processing accounting for commonalities and differences between humans and machines (Botvinick, 2022; Wichmann & Geirhos, 2023). Second, it underscores the major societal challenges associated with SOTA facial deepfakes and their detection (Groh et al., 2021; Ramon et al., 2024). Third, and finally, it provides initial evidence to support the use of synthetic facial identities as an ecologically valid tool for research on face processing and its applications.

4.1. Comparable processing of natural and synthetic identities

In two experiments, we report strong evidence that both objective stimulus similarity, as well as individual differences in observers' FIP abilities affect natural and synthetic face discrimination in the same manner. Observers' face discrimination performance varied with their independently measured FIP ability, but their performance did not vary as a function of face type. Confronted with a pair of different identities – natural or synthetic – observers' exhibited comparable face discrimination efficiency. Moreover, across face types and observers, we found comparable effects of objective stimulus similarity. Performance decreased with increased inter-identity similarity, and observers with higher FIP ability showed higher proficiency. To our knowledge, these results provide the first, and moreover compelling evidence that synthetic faces are indeed processed in the same manner as natural ones.

Table 1
Performance metrics across Natural and Synthetic conditions for Same and Different trials.

	Natural				Synthetic			
	Same		Different		Same		Different	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Accuracy	92.72	5.76	89.01	9.44	78.13	14.84	81.84	13.06
Reaction Time (correct trials)	641	138	714	155	700	170	738	195
Efficiency	1.51	0.30	1.31	0.32	1.17	0.34	1.19	0.37

These findings emerged in the context of stimuli created using a generative model that enables viewpoint variations and dynamic content of a given synthetic facial identity (Chan et al., 2022). These features provide a range of advantages across research and applied settings, e.g. comparative assessment of human and machine performance under challenging, simulated surveillance footage conditions.

4.2. Outlook

Our findings suggest that synthetic identities can, in principle, be generated and used as alternatives where, traditionally, natural faces were used. In research settings, synthetic face stimuli could entirely replace the use of images of human likenesses. The advantages are clear: The absence of privacy concerns enable sharing and thus generally contribute to enhancing reproducibility and replicability of research. Generative models offer the means to enhance fairness, e.g. by creating synthetic datasets that are balanced in terms of demographic diversity.

The advantage of shareable synthetic faces created based on models trained with demographically diverse datasets extends to applied areas, particularly where comparable assessment of individuals is critical. On the one hand, they could support the development and adoption of general standards for training and assessing abilities for high-stakes professional roles, e.g. forensic expert training and continued performance evaluation. On the other hand, provided an established legal framework for the use of synthetic identities, law enforcement professionals dealing with eye witness testimonies could create and use synthetic image lineups that are demographically fair and objectively balanced in terms of image similarity. For example, rather than having to locate a specific identity with certainty, witnesses could select identities they perceive as most similar to the perpetrator. The selected probable synthetic matches could then be compared to criminal databases. This would enhance reproducibility of lineup procedures, and also decrease exposure to images (and therefore potential stigmatization) of individuals previously taken into custody. Naturally, further studies are required to determine the effectiveness of such synthetic forward approaches.

Notwithstanding their advantages, synthetic faces are not a simple, immediately ready-to-use alternative. After automated model-based stimulus generative, manual selection of appropriate images is always required. In our case, we generated numerous iterations of identity pools, from which target identities were selected. Indeed, a very large proportion of synthetic identities had to be excluded due to the presence of abnormal information (see Supplemental Material for examples of exclusions). Further work is required to determine whether the similarities we report for frontal natural and synthetic face matching — akin to mugshot comparisons — extends to other real-life scenarios. For instance, are model-based variations of synthetic identities (e.g., aging, viewpoint, illumination) perceived in the same way as for natural faces? Our study also provides a canvas to determine whether processing of natural/synthetic identity *voices* also involves shared mechanisms.

5. Conclusion

Synthetic identities are ubiquitous and will remain a fixture of digital societies. We believe that synthetic identities provide a useful tool for future research, and across a broad range of applications. They offer means to increase transparency and privacy protection alike. Outwith academia and industry, their advantages should be leveraged in the design of equitably accessible AI literacy programs, culturally tailored for its recipients.

CRediT authorship contribution statement

Kim Uittenhove: Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Hatef Otroshi Shahreza:** Writing – review & editing, Software, Resources, Methodology, Formal analysis. **Sébastien Marcel:** Writing – review & editing, Supervision, Resources, Methodology. **Meike Ramon:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgements

MR is grateful to H. Caspar for kindly sharing his art. KU and MR thank their previous students for support during data acquisition and all volunteering observers for their participation. MR is supported by a Swiss National Science Foundation PRIMA (Promoting Women in Academia) grant (PR00P1 179872). HOS is supported by the H2020 TReSPAsS-ETN Marie Skłodowska-Curie early training network (grant agreement 860813); SM is supported by the Idiap Research Institute.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbr.2024.100563.

Data availability

Anonymized research data reported subject to analysis and analysis code can be found on the accompanying OSF project.

References

Benton, A., et al. (1983). *Contribution to neuropsychological assessment.* NY: Oxford University Press.

Bobak, A., et al. (2023). Data-driven studies in face identity processing rely on the quality of the tests and data sets. *Cortex, 166*, 348–364. https://doi.org/10.1016/j.cortex.2023.05.018

Botvinick, M. M. (2022). Realizing the promise of AI: A new calling for cognitive science. *Trends in Cognitive Sciences, 26*, 1013–1014.

Boutros, F., et al. (2023). Synthetic data for face recognition: Current state and future prospects. *Image and Vision Computing, 135*, Article 104688.

Bray, S. D., Johnson, S. D., & Kleinberg, B. (2022). Testing human ability to detect deepfake images of human faces. *Journal of Cybersecurity, 9*.

Bruck, M., Cavanagh, P., & Ceci, S. (1991). Fortysomething: Recognizing faces at one's 25th reunion. *Memory & Cognition, 19*, 221–228. https://doi.org/10.3758/BF03211146

Cao, Q., et al. (2018). VGGFace2: A dataset for recognising faces across pose and age. In *Proceedings of the 13th IEEE international conference on automatic face and gesture recognition (FG)* (pp. 67–74). IEEE.

Chan, E. R., et al. (2022). Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16123–16133).

Chesney, R. M., & Citron, D. K. (2018). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*, 1753.

Colbois, L., de Freitas Pereira, T., & Marcel, S. (2021). On the use of automatically generated synthetic image datasets for benchmarking face recognition. In *2021 IEEE international joint conference on biometrics (IJCB)* (pp. 1–8).

Damiani, J. (2019). *A voice deepfake was used to scam a CEO out of 243,000.* Forbes. https://www.forbes.20com/sites/jessedamiani/2019/09/03/avoice-deepfake-was-used-to-scam-a-ceoout-of-243000/#3937cfdd2241.

Delfino, R. A. (2019). Pornographic deepfakes: The case for federal criminalization of revenge porn's next tragic act. In *Actual problems of economics and law*.

Deng, J., et al. (2019). Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4690–4699).

Deng, J., et al. (2020). Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5203–5212).

DoHSP-Pae, P. (2021). *Increasing threat of DeepFake identities*. https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf.

Epstein, Z., et al. (2023). Art and the science of generative AI. *Science, 380*, 1110–1111.

FaceGen modeller. [Software]. (2009). Singular Inversions Inc. http://www.facegen.com/.

Farid, H. (2022). *Creating, using, misusing, and detecting deep fakes*. Journal of Online Trust and Safety.

Frowd, C. D., et al. (2005). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology, 10*, 63–81.
Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perception, 1*, 19–39. url: https://api.semanticscholar.org/CorpusID:4505645.
Fysh, M. C., & Bindemann, M. (2018). The kent face matching test. *British Journal of Psychology, 109*, 219–231.
Fysh, M. C., & Ramon, M. (2021). Accurate but inefficient: Standard face identity matching tests fail to identify prosopagnosia. *Neuropsychologia, 165*.
Fysh, M., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: Implications for theory, research and personnel selection. *Royal Society Open Science, 7*(9). https://doi.org/10.1098/rsos.200233
Geissbühler, D., Shahreza, H. O., & Marcel, S. (2024). Synthetic face datasets generation via latent space exploration from brownian identity diffusion. *arXiv preprint arXiv: 2405.00228*.
Groh, M., et al. (2021). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences of the United States of America, 119*.
Guo, Y., et al. (2016). MS-Celeb-1M: A dataset and benchmark for large-scale face recognition. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 87–102). Springer.
Jain, A. K., Ross, A., & Pankanti, S. (2006). Biometrics: A tool for information security. *IEEE Transactions on Information Forensics and Security, 1.2*, 125–143.
Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology, 14.1*, 4–20.
Kammoun, A., et al. (2022). Generative adversarial networks for face generation: A survey. *ACM Computing Surveys, 55*, 1–37.
Karras, T., et al. (2021). Alias-free generative adversarial networks. In *Neural information processing systems*.
Kim, M., et al. (2023). Dcface: Synthetic face generation with dual condition diffusion model. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 12715–12725).
Kim, M., Jain, A. K., & Liu, X. (2022). Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18750–18759).
Lab, E. I. (2022). *Facing reality? Law enforcement and the challenge of deepfakes*. https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes.
Lago, F., et al. (2021). More real than real: A study on human visual perception of synthetic faces [applications corner]. *IEEE Signal Processing Magazine, 39*, 109–116.
Li, H. (2024). *Facing the future: Implementing AI-powered digital humans across disciplines*.
Melzi, P., et al. (2023). GANDiffFace: Controllable generation of synthetic datasets for face recognition with realistic variations. In *2023 IEEE/CVF international conference on computer vision workshops (ICCVW)* (pp. 3078–3087).
Mildenhall, B., et al. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM, 65.1*, 99–106.
Nador, J., et al. (2021a). Image or identity? Only super-recognizers' (Memor)Ability is consistently viewpoint-invariant. *Swiss Psychology Open, 1*(1). https://doi.org/10.5334/spo.28
Nador, J., et al. (2021b). Psychophysical profiles in super-recognizers. *Scientific Reports, 11*, Article 13184. https://doi.org/10.1038/s41598-021-92549-6
Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences of the United States of America, 119*.
Otroshi-Shahreza, H., et al. (2024). Sdfr: Synthetic data for face recognition competition. In *2024 IEEE 18th international conference on automatic face and gesture recognition (FG)* (pp. 1–9).
Papesh, M. H. (2018). Photo ID verification remains challenging despite years of practice. *Cognitive Research: Principles and Implications, 3*.
Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
Ramon, M. (2021). Super-Recognizers – a novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia, 158*. https://doi.org/10.1016/j.neuropsychologia.2021.107809
Ramon, M., Bobak, A., & White, D. (2019). Super-recognizers: From the lab to the world and back again. *British Journal of Psychology, 110*(3). https://doi.org/10.1111/bjop.12368
Ramon, M., et al. (2015). Neural microgenesis of personally familiar face recognition. *Proceedings of the National Academy of Sciences, 112*, E4835–E4844.
Ramon, M., et al. (2016). All new kids on the block? Impaired holistic processing of personally familiar faces in a kindergarten teacher with acquired prosopagnosia. *Visual Cognition, 24*, 321–355.
Ramon, M., & Rjosk, S. (2022). beSure - Berlin test for super-recognizer identification Part I: Development. In *Verlag fur Polizeiwissenschaft*.
Ramon, M., & van Belle, G. (2016). Real-life experience with personally familiar faces enhances discrimination based on global information. *PeerJ, 4*.
Ramon, M., Vowels, M. J., & Groh, M. (2024). Deepfake detection in super-recognizers and police officers. *IEEE Security & Privacy, 22*, 68–76.
Rezlescu, C., et al. (2020). More time for science: Using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. *Progress in Brain Research, 253*, 243–262.
The rise and fall (and rise) of datasets. *Nature Machine Intelligence, 4*(1), (2022), 1–2. https://doi.org/10.1038/s42256-022-00442-2
Rombach, R., et al. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684–10695).
Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychometric Bulletin and Review, 16*, 252–257. https://doi.org/10.3758/PBR.16.2.252
Shahreza, H. O., & Marcel, S. (2024). HyperFace: Generating synthetic face recognition datasets by exploring face embedding hypersphere. In *Neurips safe generative AI workshop 2024*.
Snodgrass, J. G., & Vanderwart, M. L. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*(2), 174–215.
Stacchi, L., et al. (2020). Normative data for two challenging tests of face matching under ecological conditions. *Cognitive Research: Principles and Implications, 5*(8), 2041–2050. https://doi.org/10.1080/17470218.2014.1003949
Tummon, H. M., Allen, J., & Bindemann, M. (2019). Facial identification at a virtual reality airport. *I-Perception, 10*.
Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society, 6*.
Valentine, T. (2001). Face–space models of face recognition. In M. J. Wenger, & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 83–113).
Van Doorn, J., et al. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review, 28*, 813–826.
Wang, X., et al. (2022). *GAN-generated faces detection: A survey and new perspectives. ArXiv* abs/2202.07145.
Weber, E. H., Ross, H. E., & Murray, D. J. (2018). *E.H. Weber on the tactile senses*.
Wichmann, F., & Geirhos, R. (2023). Are deep neural networks adequate behavioural models of human visual perception?. In *Annual review of vision science*.
Zheng, Y., et al. (2022). General facial representation learning in a visual-linguistic manner. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18697–18709).