Escaping the Uncanny Valley of Synthetic Faces

WALTER J. SCHEIRER
UNIVERSITY OF NOTRE DAME

The Uncanny Valley

By Masahiro Mori (Translated by Karl F. MacDorman and Norri Kageki)

Editor's note: More than 40 years ago, Masahiro Mori, a robotics professor at the Tokyo Institute of Technology, wrote an essay [1] on how he envisioned people's reactions to robots that looked and acted almost like a human. In particular, he hypothesized that a person's response to a humanlike robot would abruptly shift from empathy to revulsion as it approached, but failed to attain, a lifelike appearance. This descent into eeriness is known as the *uncanny valley*. The essay appeared in an obscure Japanese journal called *Energy* in 1970, and in subsequent years, it received almost no attention. However, more recently, the concept of the uncanny valley has rapidly attracted interest in robotics and other scientific circles as well as in popular culture. Some researchers have explored its implications for human–robot interaction and computer-graphics animation, whereas others have investigated its biological and social roots. Now interest in the uncanny valley should only intensify, as technology evolves and researchers build robots that look human. Although copies of Mori's essay have circulated among researchers, a complete version hasn't been widely available. The following is the first publication of an English translation that has been authorized and reviewed by Mori. (See "Turning Point" in this issue for an interview with Mori.)

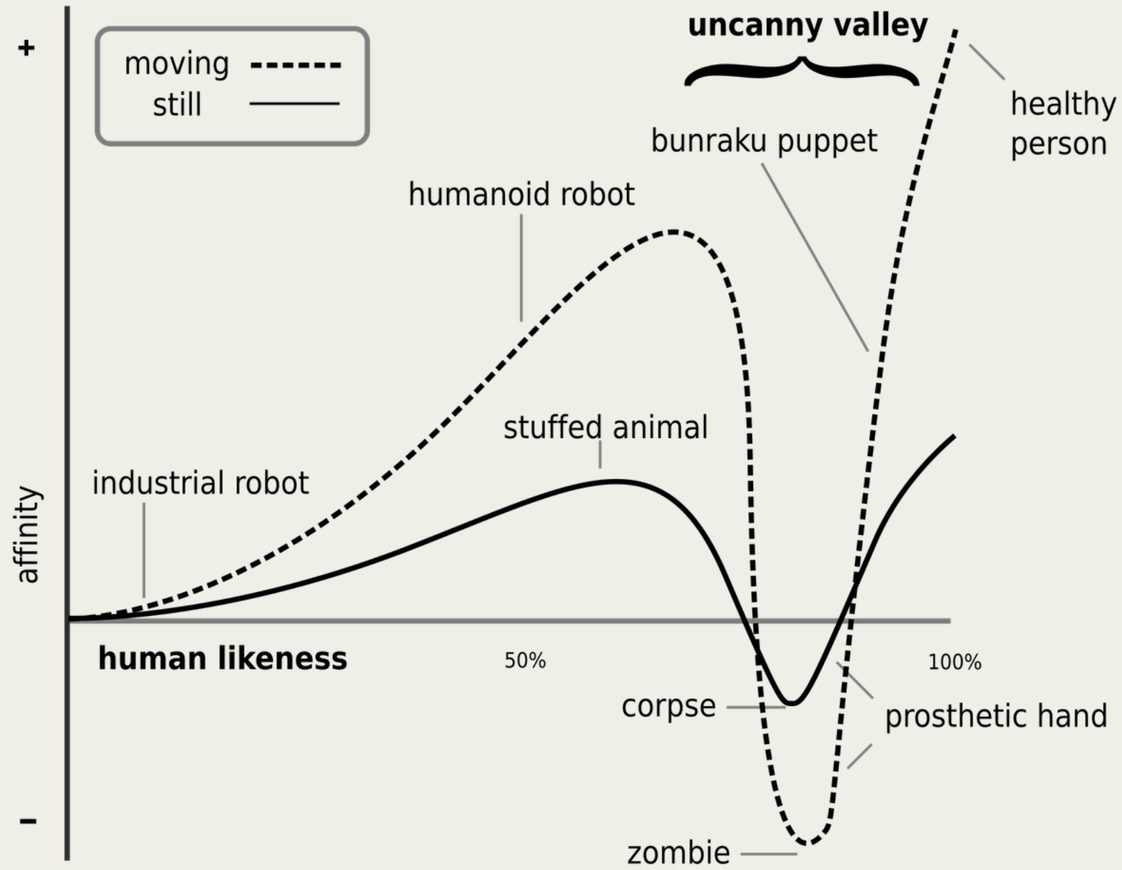
A Valley in One's Sense of Affinity

The mathematical term *monotonically increasing function* describes a relation in which the function $y = f(x)$ increases continuously with the variable x. For example, as effort x grows, income y increases, or as a car's accelerator is pressed, the car moves faster. This kind of relation is ubiquitous and easily understood. In fact, because such monotonically increasing functions cover most phenomena (y)—owing to the intervening hills and valleys. I have noticed that, in climbing toward the goal of making robots appear like a human, our affinity for them increases until we come to a valley (Figure 1), which I call the *uncanny valley*.

Nowadays, industrial robots are increasingly recognized as the driving force behind reductions in factory personnel. However, as is well known, these robots just extend, contract, and rotate their arms; without faces or legs, figure, the robot will start to have a roughly human-looking external form with a face, two arms, two legs, and a torso. Children seem to feel deeply attached to these toy robots. Hence, the toy robot is shown more than halfway up the first hill in Figure 1.

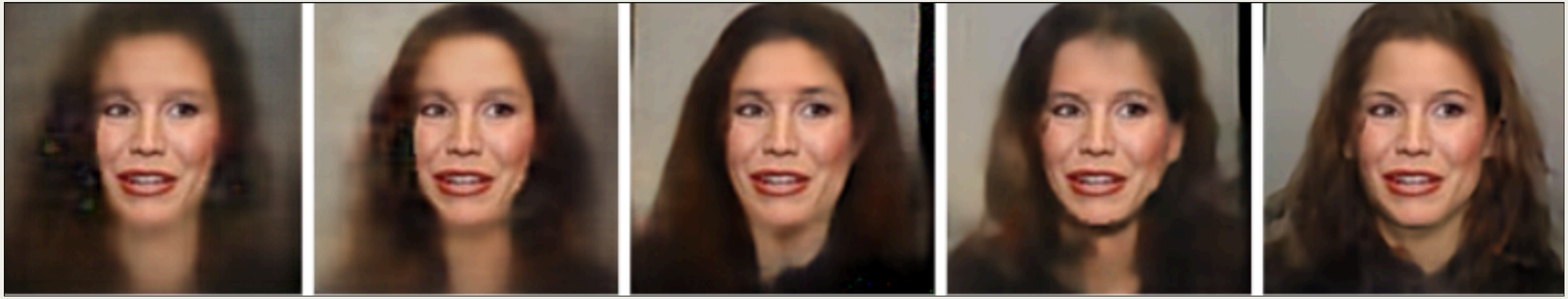
Since creating an artificial human is itself one of the objectives of robotics, various efforts are underway to build humanlike robots. (Note: Others believe that the true appeal of robots is their potential to exceed and augment humans.)

The Uncanny Valley



The Uncanny Valley of Synthetic AI Faces

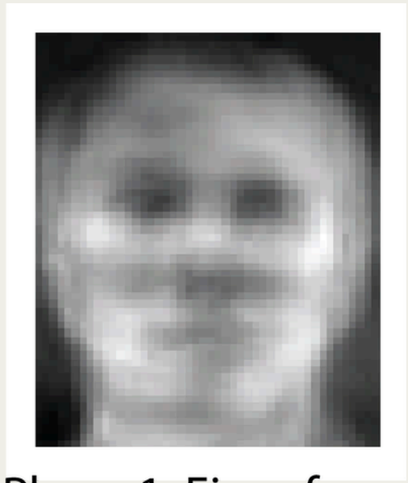
Progressive Growing of GANs (2018)



T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. ICLR, 2018.

B. Sandipan, W. Scheirer, K. Bowyer, and P. Flynn. On Hallucinating Context and Background Pixels from a Face Mask Using Multi-scale GANs. WACV, 2020.

PROGRESS IN FACE SYNTHESIS



Phase 1: Eigenfaces
(Generative)

1990s

Phase 2: FaceGen
(Procedural)

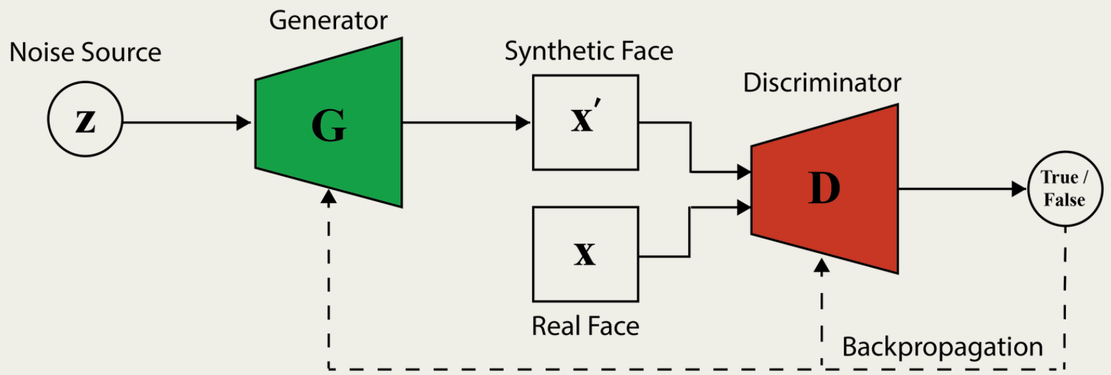
2000s - Mid-2010s

Phase 3: GANs
(Generative)

Mid-2010s - present

CURRENT GENERATIVE MODELS

Generative Adversarial Nets (GANs)



Diffusion

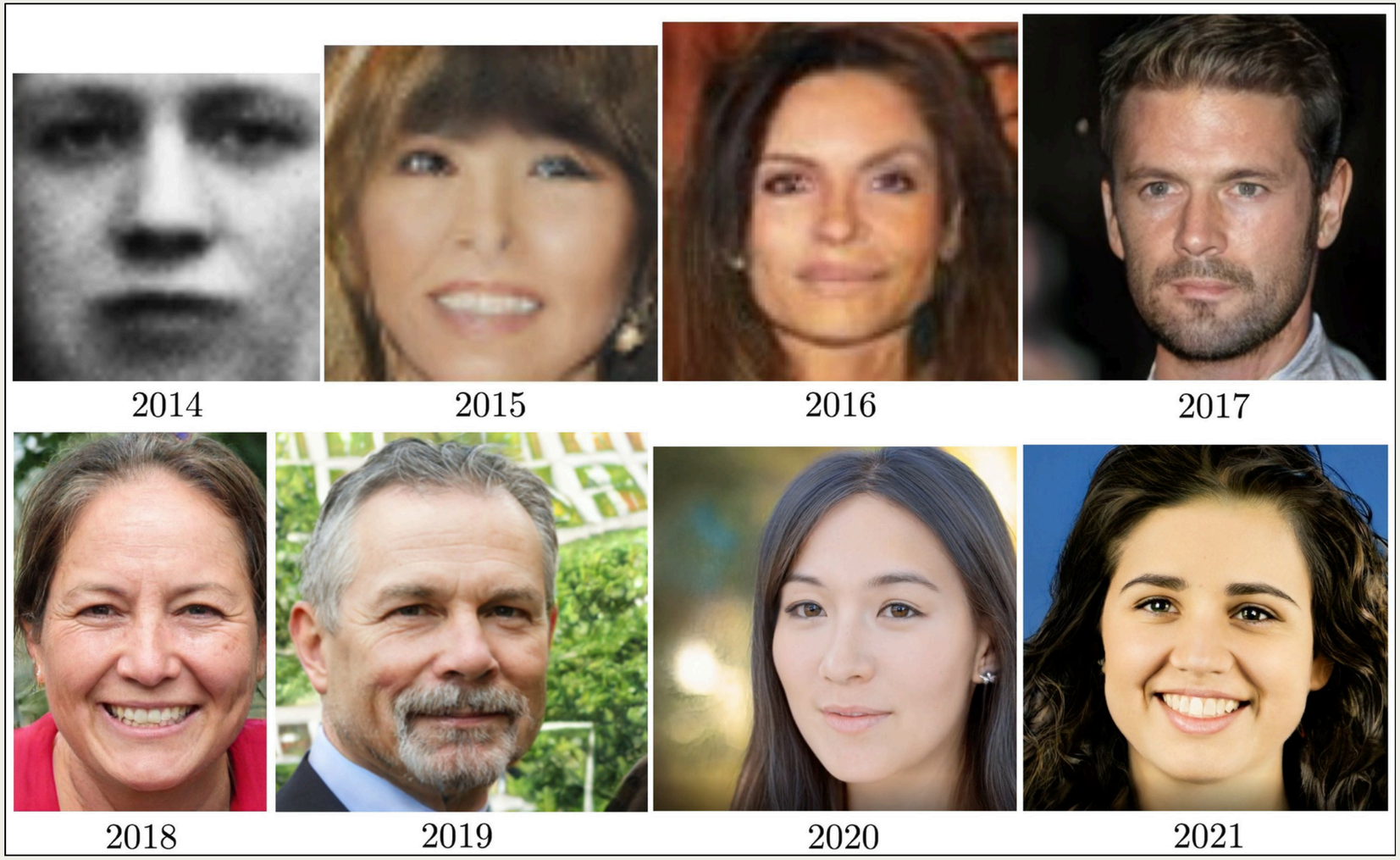
Sora Prompt: "generate a photo of a male face with the English countryside in the background"

Sora Prompt: "generate a photo of a female face with the English countryside in the background"

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Nets. NeurIPS, 2014.

R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-Resolution Image Synthesis With Latent Diffusion Models. CVPR, 2022.

PROGRESSION OF GAN IMPROVEMENT



https://davidleonfdez.github.io/gan/2022/05/17/gan-convergence-stability.html

STYLEGAN2: ALMOST THERE

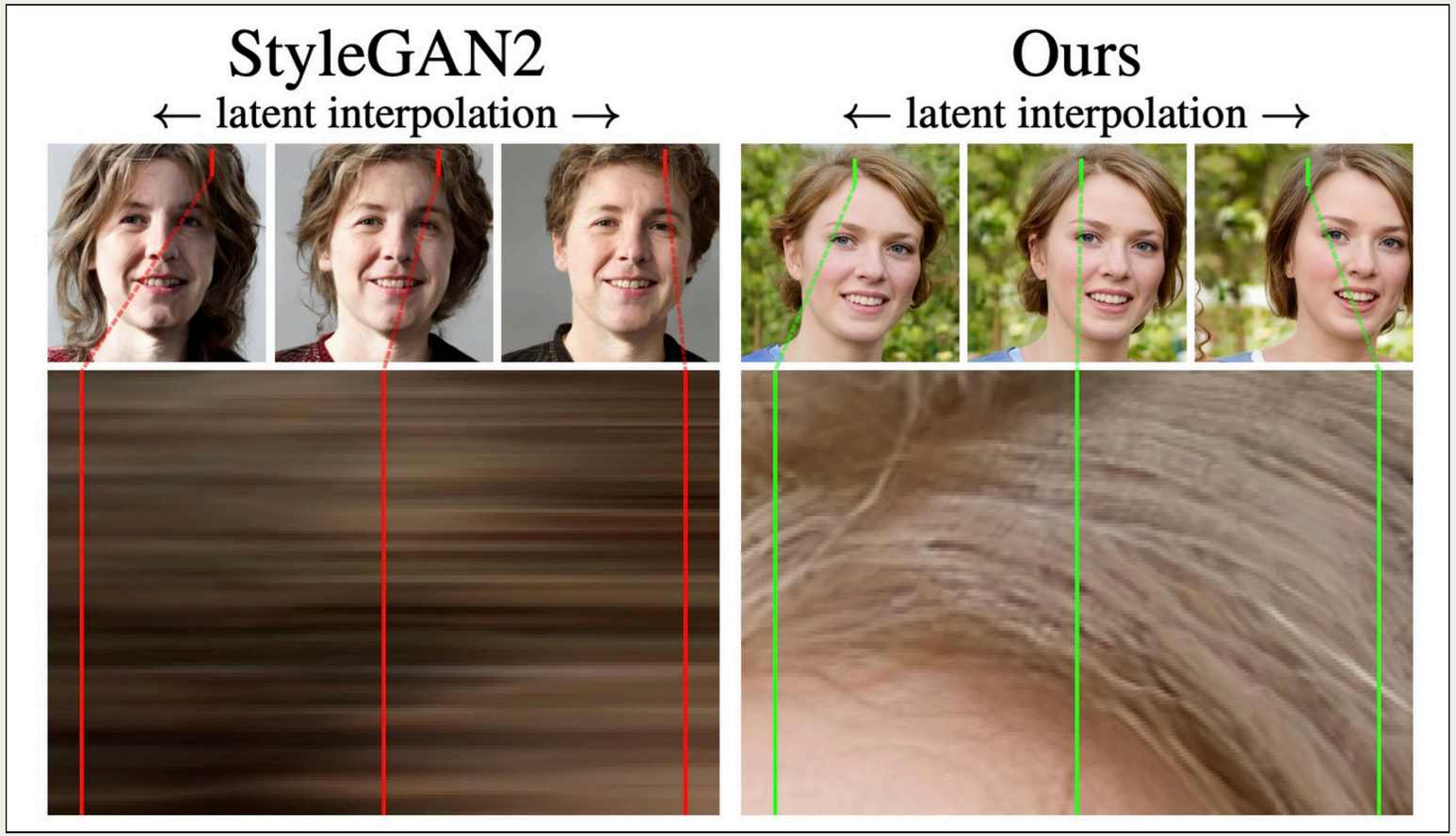
https://thispersondoesnotexist.com/

Capabilities of StyleGAN2:

- Control of the degree to which features are present at different scales (“style” of the image)
  - Automatic, unsupervised separation of high-level attributes (e.g., pose, identity) from stochastic variation (e.g., freckles, hair)
- Scale-specific mixing and interpolation operations
- Generator normalization
- Enhanced strategy for progressive growing
- Regularization scheme for the generator for better mapping from latent codes to images

T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and Improving the Image Quality of StyleGAN. CVPR, 2020.

STYLEGAN3: BEST AVAILABLE FACE SYNTHESIS



Capabilities of StyleGAN3:

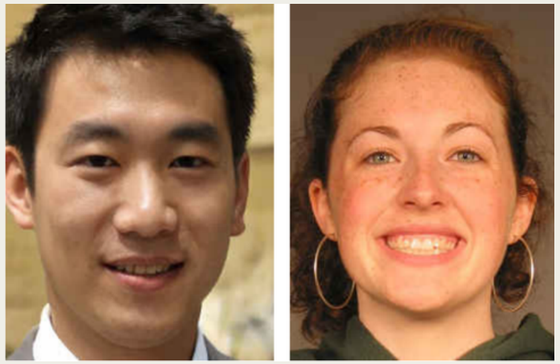
- Addresses aliasing artifacts in generated images

T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila. Alias-free Generative Adversarial Networks. NeurIPS, 2021.

Key Research Question: Can People Tell the Difference Between Real and Synthetic Faces?

STUDY #1: SHEN ET AL. 2021

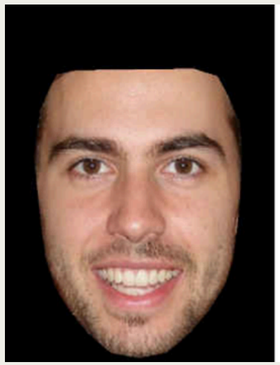
StyleGAN2 **Real**



Experiment 1: Distinguishing faces with full context

176 participants, 1,000 pairs of faces

Each image pair viewed by at least 10 participants (true of all 3 experiments)




Experiment 2: Distinguishing faces with only the face region visible

174 participants, 1,000 pairs of faces

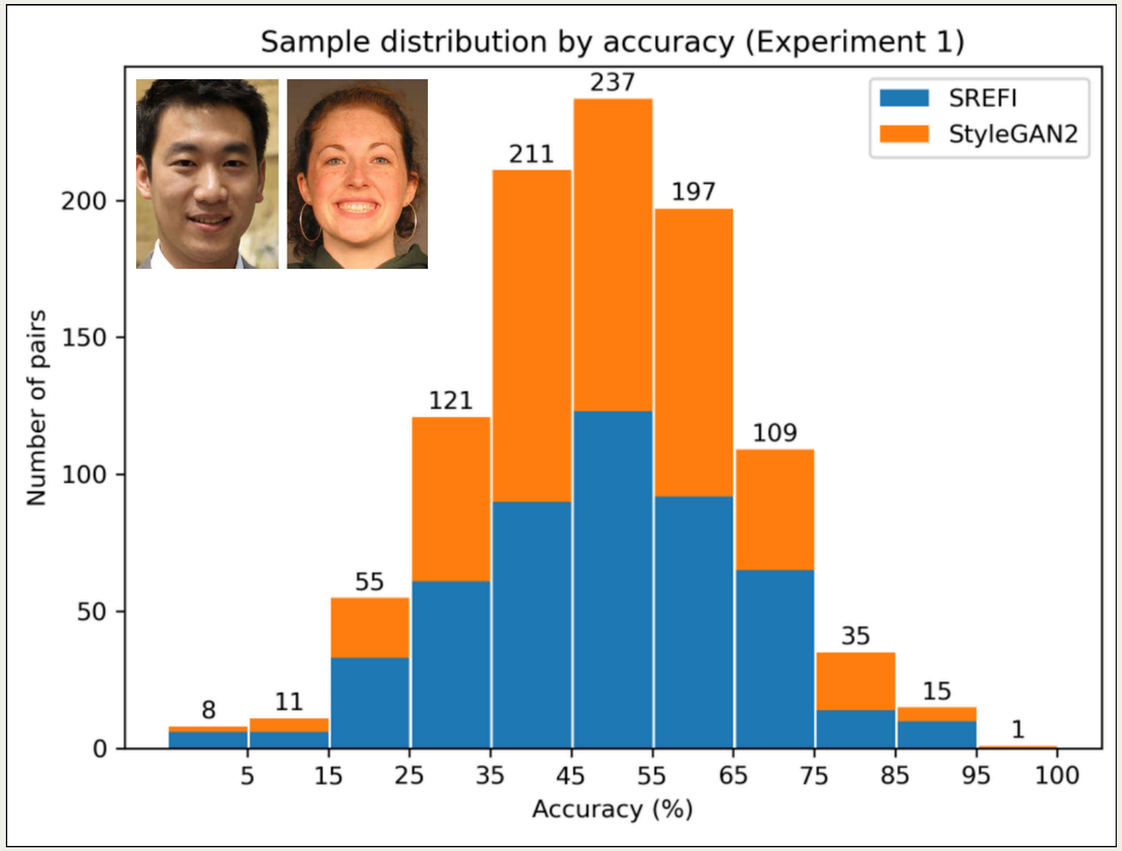

Experiment 3: Distinguishing faces under varied lighting

172 participants, 1,000 pairs of faces

B. Shen, B. RichardWebster, A. O'Toole, K. Bowyer, and W. Scheirer. A Study of the Human Perception of Synthetic Faces. FG, 2021.

STUDY #1: SHEN ET AL. 2021

Experiment 1: Distinguishing faces with full context



Shen et al. 2021

Considering the t-statistic for the right tail, the null hypothesis of “human raters do not tend to identify synthetic faces correctly” is not rejected ($p > 0.05$).

Considering the t-statistic for the left tail, the null hypothesis of “human raters tend not to mistake synthetic faces as real faces,” is rejected ($p < 0.05$).

There is a human tendency to label synthetic faces as real.

S. Banerjee, J. Bernhard, W. Scheirer, K. Bowyer, and P. Flynn. SREFI: Synthesis of Realistic Example Face Images. IJCB, 2017.

STUDY #1: SHEN ET AL. 2021

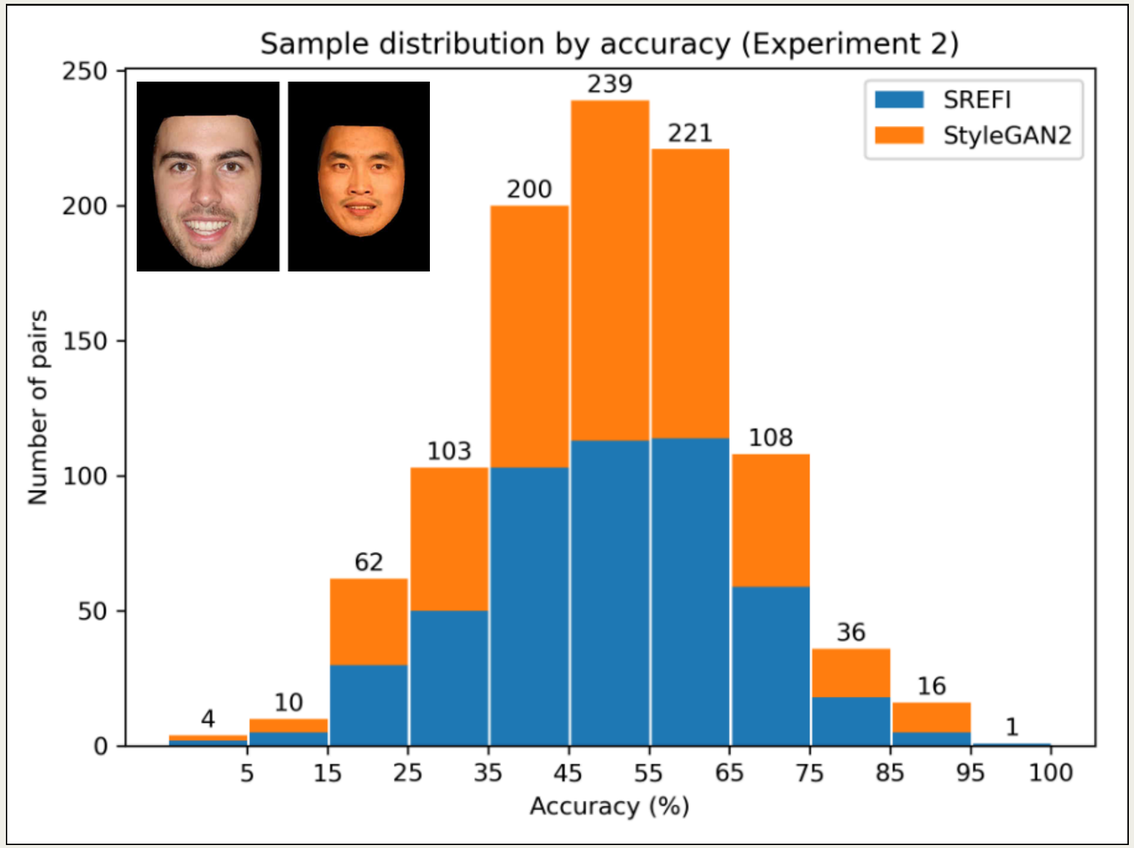
https://thispersondoesnotexist.com/

There are a few possibilities for why this occurs

- The first two are related to the facial features:
  - the synthetic images have very smooth skin under the eyes (blurring effect)
  - eyes are lined up on the exact horizontal (symmetry)
- Possibility that the background is attracting undue attention, causing participants to make incorrect decisions based on irrelevant information

STUDY #1: SHEN ET AL. 2021

Experiment 2: Distinguishing faces with only the face region visible



Shen et al. 2021

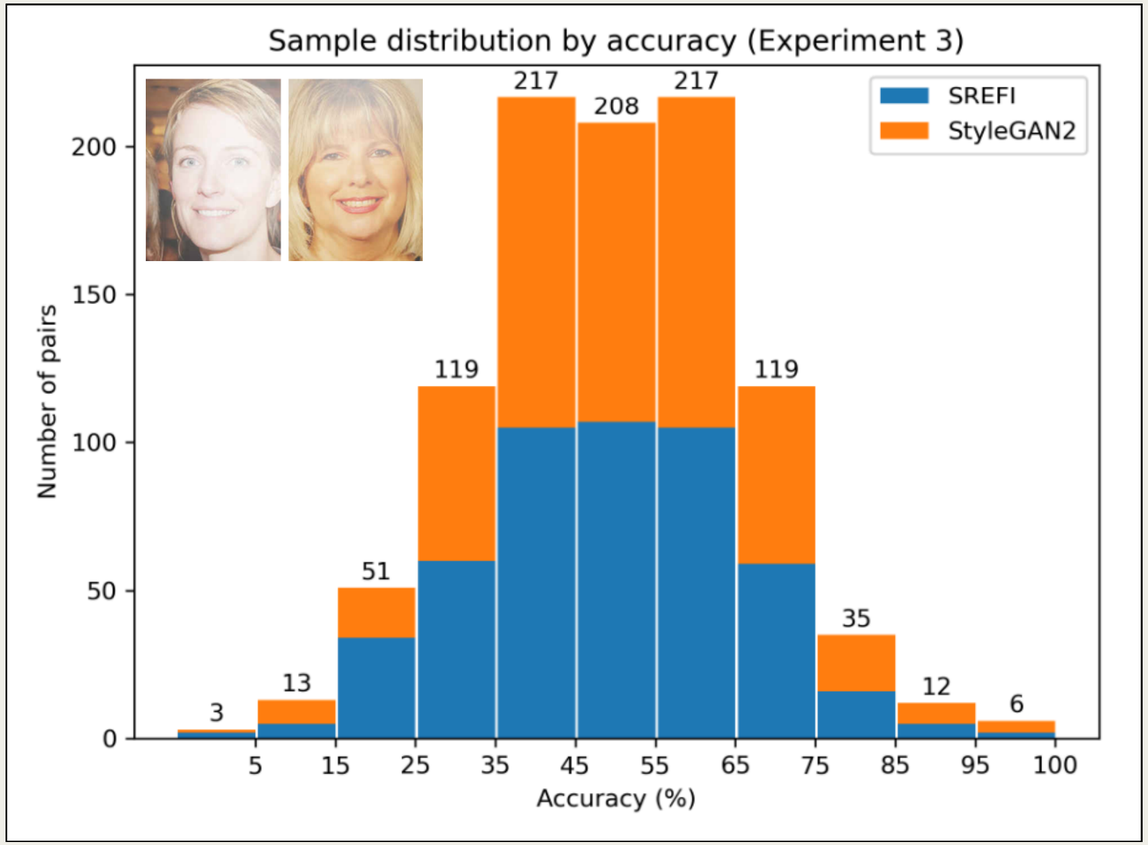
Considering the t-statistic for the right tail, the null hypothesis of “human raters do not tend to identify synthetic faces correctly” is not rejected ($p > 0.05$).

Considering the t-statistic for the left tail, the null hypothesis of “human raters tend not to mistake synthetic faces as real faces,” is not rejected ($p > 0.05$).

There is no human tendency to label synthetic faces as real.

STUDY #1: SHEN ET AL. 2021

Experiment 3: Distinguishing faces under varied lighting



Shen et al. 2021

Considering the t-statistic for the right tail, the null hypothesis of “human raters do not tend to identify synthetic faces correctly” is not rejected ($p > 0.05$).

Considering the t-statistic for the left tail, the null hypothesis of “human raters tend not to mistake synthetic faces as real faces,” is not rejected ($p > 0.05$).

There is no human tendency to label synthetic faces as real.

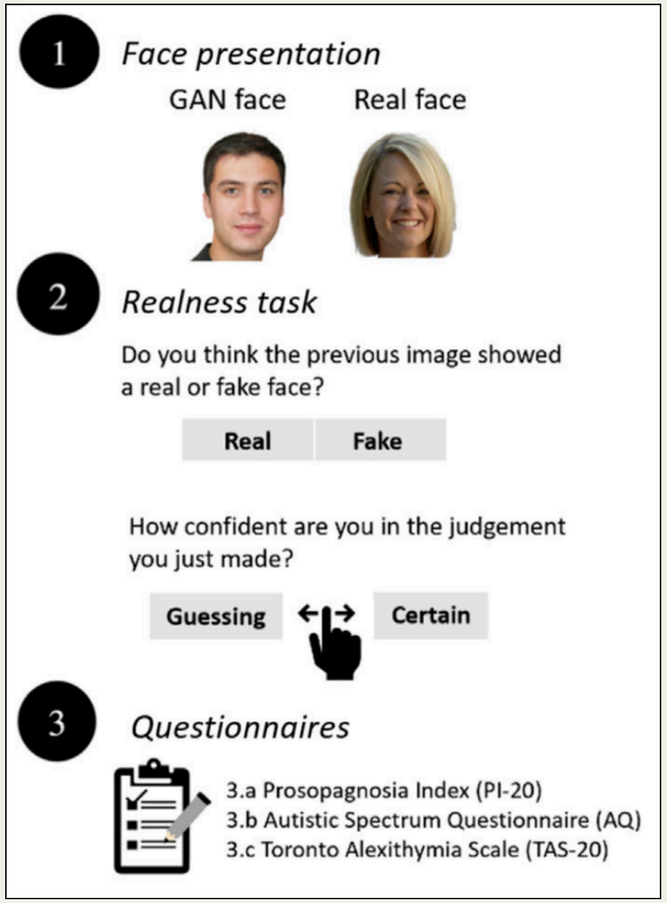
STUDY #2: TUCCIARELLI ET AL. 2022

Two-alterative forced-choice task for measuring the human ability to detect synthetic faces; participant confidence also collected.

Participants also given questionnaires to assess prosopagnosic, autistic, and alexithymic traits, which might influence how certain participants perceive realness.

Each participant was shown 100 faces (50 real and 50 synthetic), with 107 participants involoved in the study.

“GAN faces were judged to be more real than the REAL faces, even when controlling for the specified stimulus characteristics and participants’ demographics” with statistical significance.



R. Tucciarelli, N. Vehar, S. Chandaria, and M. Tsakiris. On the Realness of People Who do Not Exist: The Social Processing of Artificial Faces. Iscience 25, no. 12 (2022).

STUDY #3: NIGHTINGALE AND FARID 2022

Two experiments: two-alternative forced-choice tasks designed to gauge human ability to distinguish between real and AI faces, with and without training. Concluded that people are unable to distinguish between real and AI faces.

A third experiment examined the trustworthiness of real and AI faces. 223 subjects provided Likert ratings of 128 faces. The average rating of synthetic faces was 7.7% more trustworthy than real faces (a significant difference $p < 0.001$).

“This may be because synthesized faces tend to look more like average faces which themselves are deemed more trustworthy”

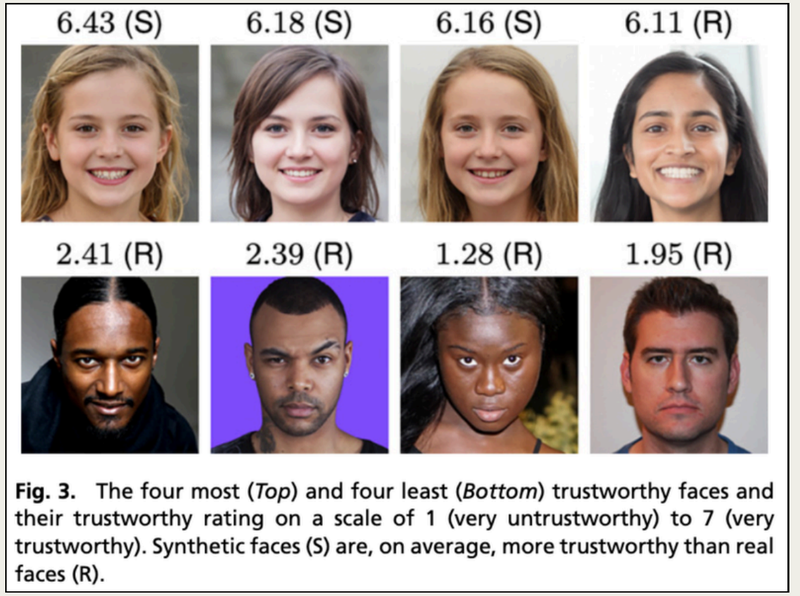


Fig. 3. The four most (*Top*) and four least (*Bottom*) trustworthy faces and their trustworthy rating on a scale of 1 (very untrustworthy) to 7 (very trustworthy). Synthetic faces (S) are, on average, more trustworthy than real faces (R).

S. Nightingale and H. Farid. AI-synthesized Faces are Indistinguishable From Real Faces and More Trustworthy. Proceedings of the National Academy of Sciences 119, no. 8 (2022): e2120481119.

STUDY #4: MILLER ET AL. 2023

AI Hyperrealism: When AI faces are judged as human more often than actual human faces

E. Miller, B. Steward, Z. Witkower, C. Sutherland, E. Krumhuber, and A. Dawel. AI Hyperrealism: Why AI Faces are Perceived as More Real than Human Ones. Psychological Science 34, no. 12 (2023b): 1390-1403.

STUDY #4: MILLER ET AL. 2023

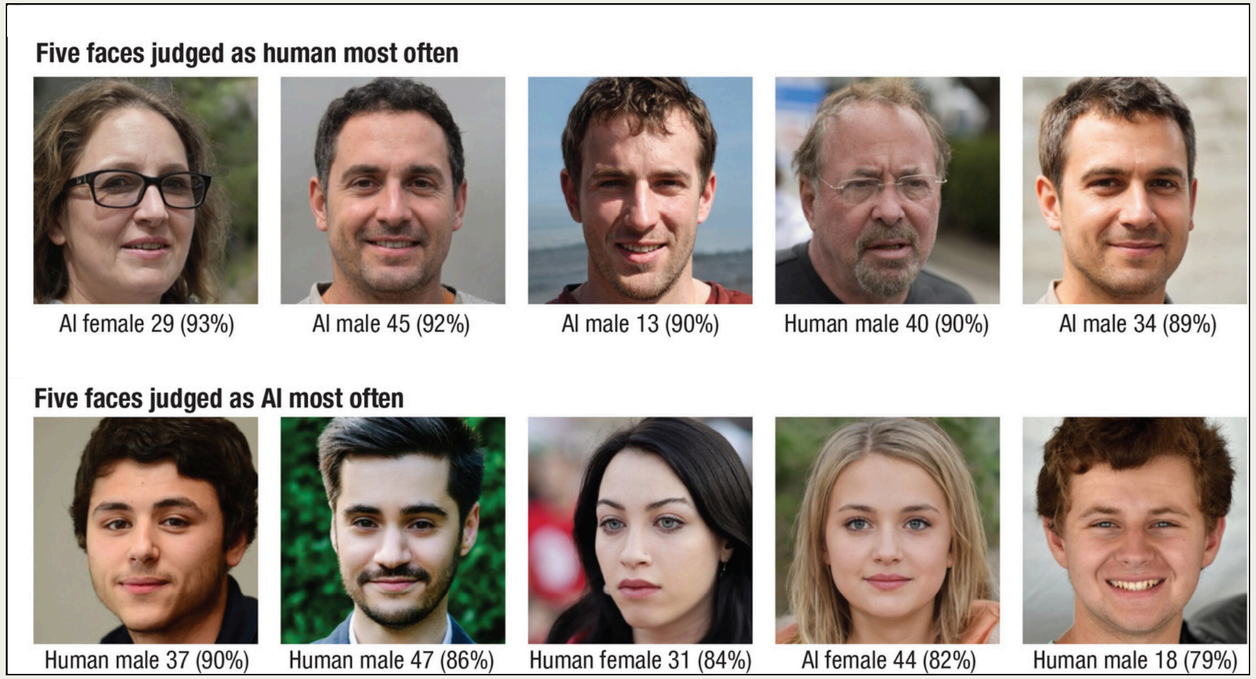
The inverse of the uncanny valley effect: faces so perfect that they are more human than human to participants.

https://thispersondoesnotexist.com/

STUDY #4: MILLER ET AL. 2023

Preliminary reanalysis of the data from Nightingale and Farid 2022

Significant effect of hyperrealism for White AI faces, but not for non-White AI faces ($p < 0.001$ in the former cases and $p = 0.682$ in the latter case)

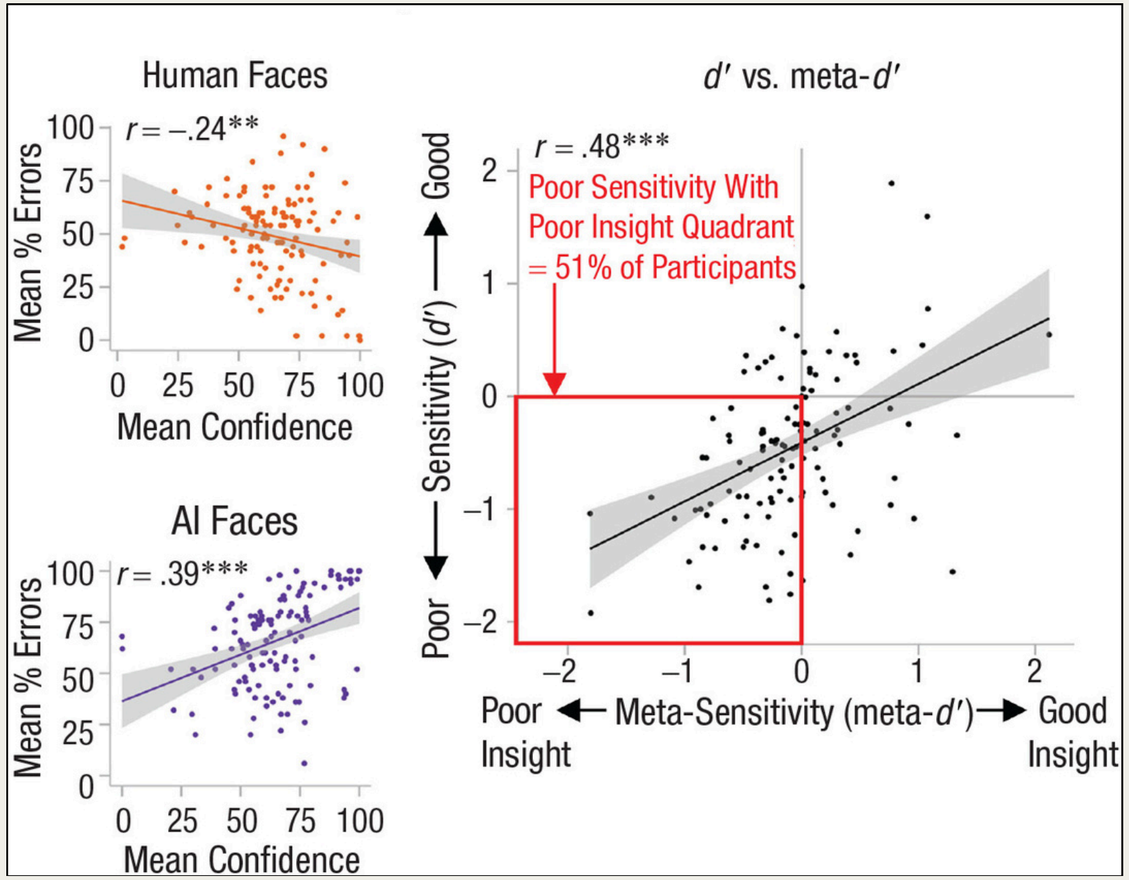


Miller et al. 2023

STUDY #4: MILLER ET AL. 2023

1st new experiment: determine if the hyperrealism effect replicates when subjects report their confidence and free responses describing what information they used to make their decision

- 124 subjects viewed 100 AI and 100 human faces in each trial
- Hyperrealism effect was replicated in this new study ($p < 0.001$)
- Dunning-Kruger effect was present across the subjects who made the most mistakes. They tended to be the most confident in their decisions

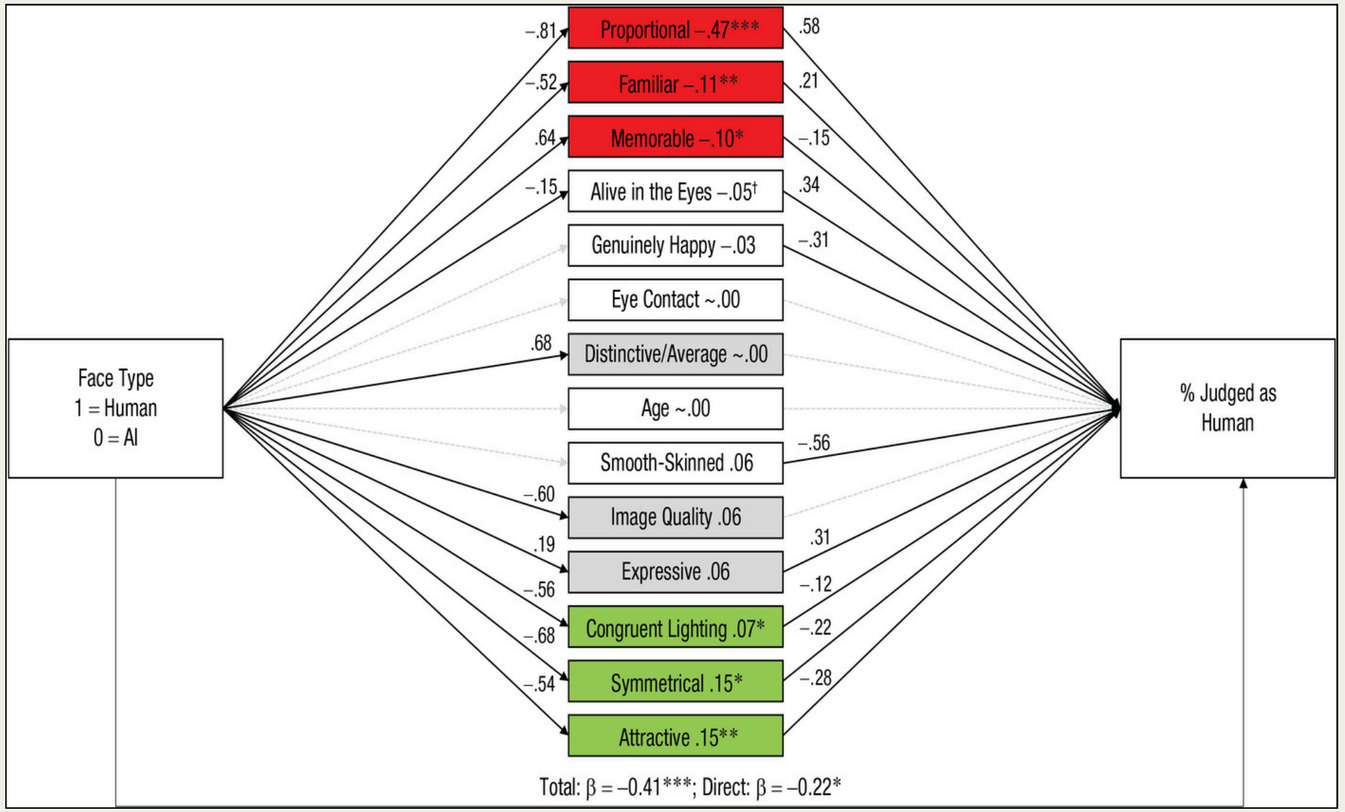


Miller et al. 2023

STUDY #4: MILLER ET AL. 2023

2nd new experiment: assess 14 face-specific attributes that might explain the hyperrealism effect

- Four attributes from face space theory, nine of the most common attributes mentioned in the free responses from the first experiment, and perceived age
- 610 subjects rated the faces from the first experiment on one of the 14 attributes
- Using a Brunswikian lens model, AI faces were significantly more average (less distinctive), familiar, and attractive

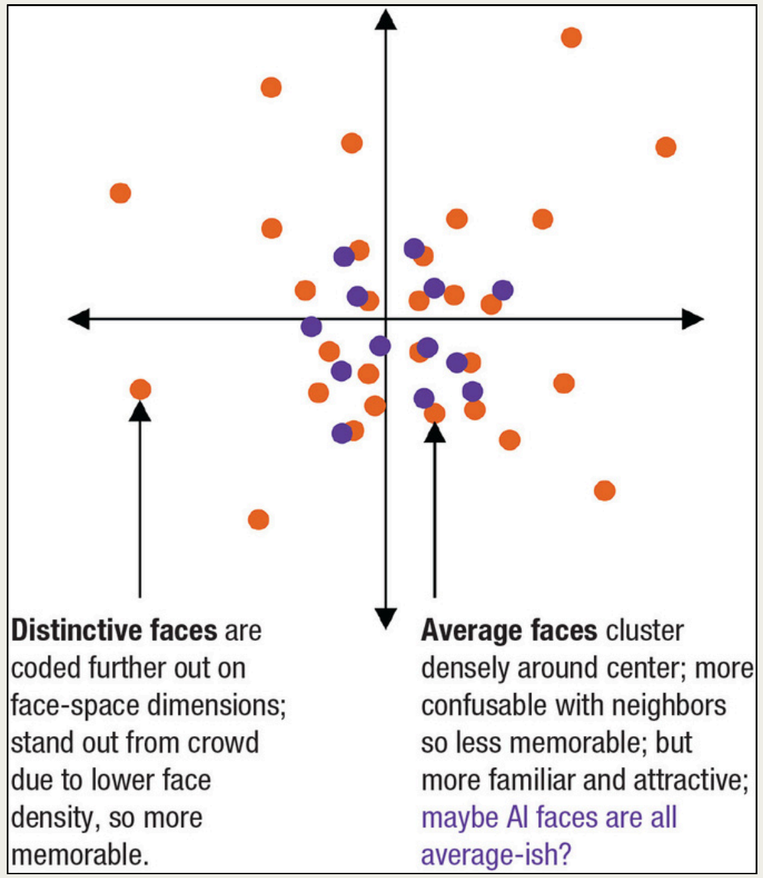


Miller et al. 2023

STUDY #4: MILLER ET AL. 2023

Two biases that emerged from experimentation:

1. StyleGAN2 generates faces that resemble average faces
    - The same observation made by Nightingale and Farid (2022)
    - The claim here is that average faces tend to be judged as more attractive
    - This might also happen with other diffusion-based generators such as OpenAI's DALL-E

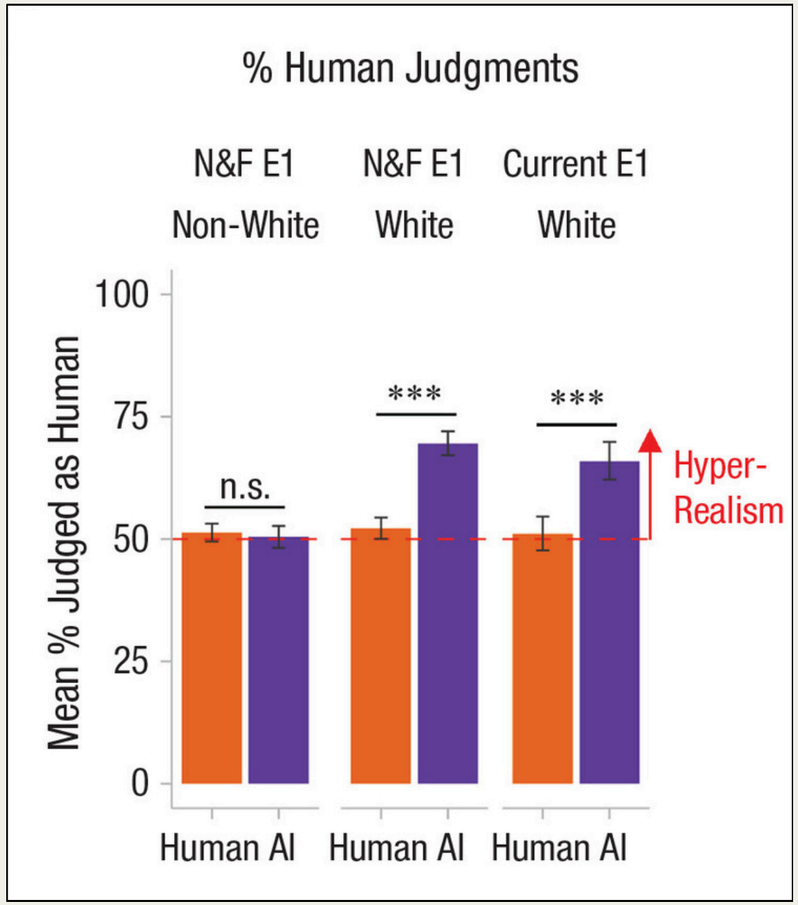


Miller et al. 2023

STUDY #4: MILLER ET AL. 2023

Two biases that emerged from experimentation:

2. White racial bias in the training data of commonly used generative models
    - The hyperrealism effect may go away when considering other races that are not as well represented in the training data
    - Not all synthetic faces are created equally, but in principle, could be with proper data curation



Miller et al. 2023

A QUESTION FOR WHICH THERE IS NO CURRENT CONSENSUS

Are synthetic faces good enough to use for studies of facial perception or other practical software applications?

- Di Natale et al. (2023) argue that the uncanny valley phenomenon persists, with even the best GAN-based generators producing flawed images
  - While the bulk of generated images in a dataset may be acceptably photorealistic, individual images may still contain problematic artifacts
- Miller et al. (2023a) are more mixed in their analysis of the utility of synthetic faces for a number of applications, highlighting strengths and weaknesses when considering the domains of e-health, social companionship, video-gaming, and scientific work
- Groh et al. (2022) and Huang et al. (2025) also cast some doubt, but they looked at specialized circumstances (DeepFake detection and forensic analysis), not free viewing

A. Di Natale, M. Ellen Simonetti, S. La Rocca, and E. Bricolo. Uncanny Valley Effect: A Qualitative Synthesis of Empirical Research to Assess the Suitability of Using Virtual Faces in Psychological Research. Computers in Human Behavior Reports 10 (2023): 100288.

E. Miller, Y. Zhi Foo, P. Mewton, and A. Dawel. How Do People Respond to Computer-generated Versus Human Faces? A Systematic Review and Meta-analyses. Computers in Human Behavior Reports 10 (2023a): 100283.

M. Groh, Z. Epstein, C. Firestone, and R. Picard. Deepfake Detection by Human Crowds, Machines, and Machine-informed Crowds. Proceedings of the National Academy of Sciences 119, no. 1 (2022): e2110013119.

J. Huang, S. Gopalakrishnan, T. Mittal, J. Zuena, and J. Pytlarz. Analysis of Human Perception in Distinguishing Real and AI-Generated Faces: An Eye-Tracking Based Study. IEEE FG. 2025.

RELATIVELY UNEXPLORED RESEARCH AVENUES

How is human perception changing as more of life is lived online?

Are there behavioral changes associated with hyperrealism that we should be aware of?

Changes to the visual environment are being driven by the Internet, opening new doors of inquiry for vision scientists.

https://thispersondoesnotexist.com/

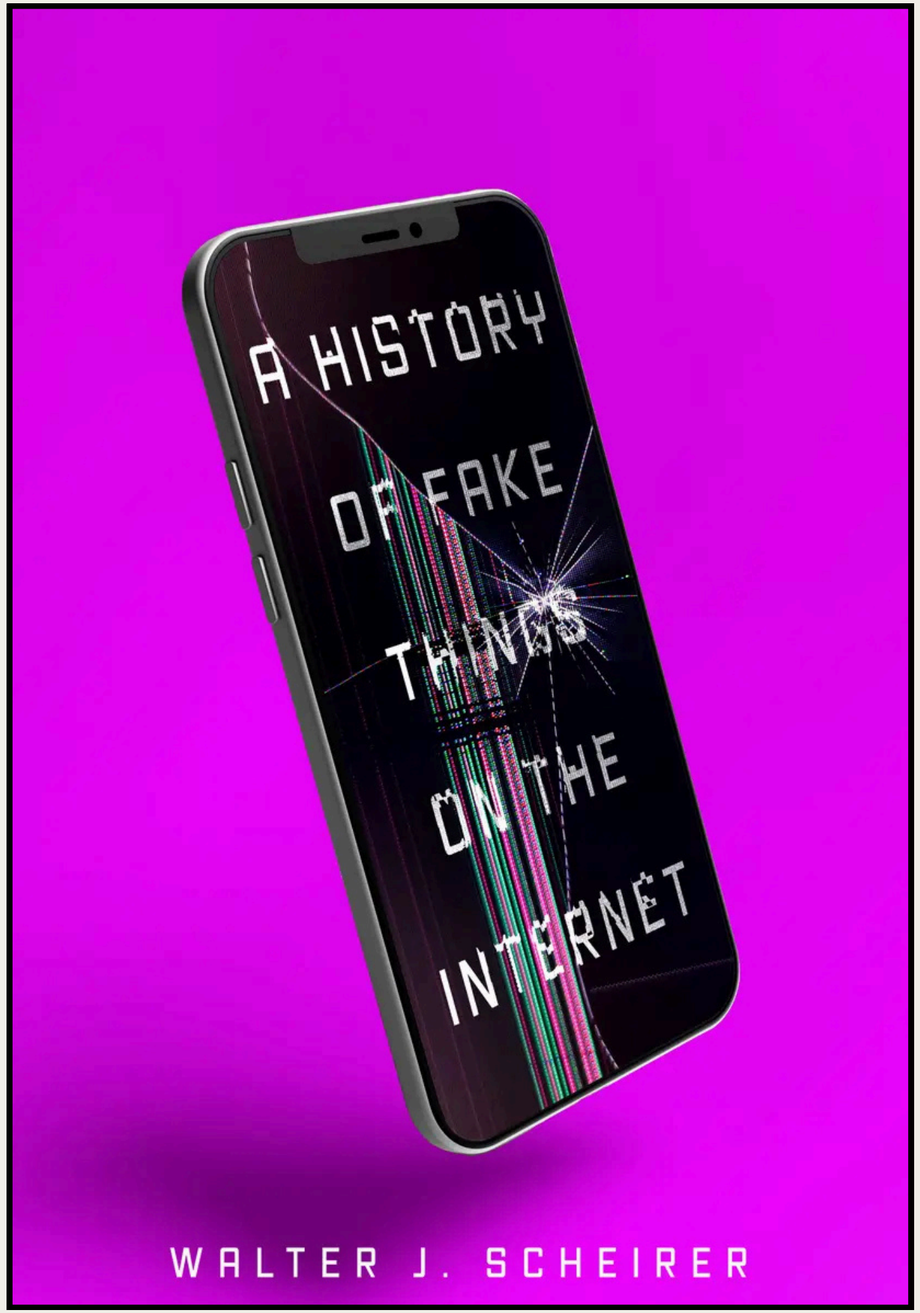
Want to learn more about AI bending human *and* machine perception?

Check out Louisa Conwill's poster at this workshop!

See the forthcoming book chapter "Studies with Synthetic and Filtered Faces" in *The Face Book* (Oxford University Press 2026)

Interested in the social history of synthetic imagery?

Read this book!



(Stanford University Press 2023)

Questions?