The Effect of Fictional Reappraisal on Subjective Ratings Toward Images

Dominique Makowski[2,1] and Ana Neves[2]
[1]Sussex Centre for Consciousness Science, University of Sussex
[2]School of Psychology, University of Sussex

Blabla the abstract blabla.

Keywords: keyword1, keyword2, keyword3

Recent advances in artificial intelligence (AI) have introduced new challenges for human cognition, particularly regarding the ability to distinguish between authentic and fabricated experiences (Miller et al., 2023). The stakes of such perceptual uncertainty are considerable, with misinformation representing one of the most pressing societal consequences (Kreps et al., 2022). For instance, deepfakes (face-swapping technologies that enable the creation of realistic fake images and videos) have already been employed to fabricate convincing political speeches that appear genuine (Meskys et al., 2020). When deployed in electoral contexts, such content has the potential to distort public opinion and undermine democratic processes by rendering the truth unclear (Graber-Mitchell, 2021). Yet political disinformation is only one domain in which synthetic media exerts influence. The increased accessibility of generative technologies means that manipulated or entirely artificial content now permeates social media, entertainment, immersive environments, and interpersonal communication (Nightingale & Farid, 2022). In these settings, the boundaries between the "real" and the "artificial" are becoming progressively less discernible, raising fundamental questions about how individuals perceive and evaluate reality.

A key challenge lies in the prevalence of ambiguous stimuli, that is images, texts, videos, or environments whose authenticity cannot be easily established. While artificial stimuli once carried perceptual markers that betrayed their inauthenticity, such as distortions in early computer-generated imagery (Corvi et al., 2023; McDonnell & Breidt, 2010), these limitations are rapidly disappearing. In the domain of face generation especially, current algorithms now produce synthetic images that are virtually indistinguishable from genuine photographs (Nightingale & Farid, 2022; **?**). As these technologies approach perfect realism, the study of reality perception becomes not only a theoretical concern but also a practical imperative, with direct implications for information security, social trust, and the ethical deployment of AI.

Study 1

Methods

Participants

The initial sample comprised 1,067 participants recruited via multiple channels, including Prolific©, Sona (ref), social media platforms, university classrooms, and snowball sampling. This strategy yielded a heterogeneous pool of both incentivised and non-incentivised individuals, including students and members of the general population from England, France, Italy, Colombia, and Spain.

To ensure data quality, several exclusion criteria were applied. Participants were removed if they (a) showed no variation in arousal ratings across trials (N = 8), (b) displayed a negative correlation between arousal and enticement alongside lower arousal ratings for erotic compared to neutral stimuli, suggesting a possible misunderstanding of scale direction (N = 4), or (c) self-identified with a gender or sexual orienta-

Dominique Makowski
Ana Neves

This preprint is a non-peer-reviewed work from the **Reality Bending Lab**.

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: Dominique Makowski: Project administration, Data curation, Formal Analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing; Ana Neves: Data curation, Formal Analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing

Correspondence concerning this article should be addressed to Dominique Makowski, Email: D.Makowski@sussex.ac.uk

tion incompatible with the aims of the analysis (N = 350). For the latter, only self-reported heterosexual individuals were retained.

The final sample consisted of 705 participants (Mean age = 30.2 years ± 11.8; 35.7% female). Participants were primarily from the United Kingdom (28.23%), Italy (18.72%), the United States (14.33%), and Colombia (11.06%), with the remaining 27.66% distributed across other countries. Ethical approval for this study was obtained from the School of Psychology Ethics Committee at the University of Sussex (ER/MHHE20/1).

Materials

All written materials in this study were translated into the participants' native languages: English, Italian, French, and Spanish.

Questionnaires. The Beliefs about Artificial Image Technology (BAIT) scale assesses general attitudes toward artificial intelligence (AI) and beliefs about AI-generated media. It includes six items adapted from the General Attitudes towards Artificial Intelligence Scale (GAAIS, Schepman & Rodway, 2020, 2023), comprising three positively valenced (e.g., "Artificial Intelligence is exciting") and three negatively valenced items (e.g., "Artificial Intelligence might take control of people"). In addition, several items were developed to evaluate beliefs about computer-generated imagery, such as "Current Artificial Intelligence algorithms can generate very realistic images" and "Images of faces or people generated by Artificial Intelligence always contain errors and artifacts." All items were rated on a continuous scale from strongly disagree (0) to strongly agree (1). One item was included to assess self-reported AI knowledge, with anchors ranging from Not at all (0) to Expert (6).

The Consumption of Pornography Scale – General (COPS, Hatch et al., 2023) is a 34-itemmeasure assessing pornography use across multiple dimensions, including frequency, duration and recency of sexual activity. Participants reported how often they had viewed pornography in the past 30 days (e.g., not at all, once or twice, weekly, daily, multiple times per day) and the typical duration of viewing sessions (less than 5 minutes to 46+ minutes). An additional item assessed the recency of any sexual activity, with response options ranging from within the past 24 hours to more than a year ago.

Affective Measures.

Arousal. Subjective sexual arousal was assessed following each image with the question, "How much did you feel sexually aroused?" Responses were recorded on a continuous scale from Not at all (0) to Very much (1).

Enticement. Perceived enticement was measured after each image using the question, "How enticing would you rate this image to be?" with the same scale ranging from Not at all (0) to Very much (1).

Valence. Emotional valence was evaluated by asking, "The feeling evoked by the image was…" rated on a scale from Unpleasant (-1) to Pleasant (1).

Realism. In a final stage of the experiment, each image was shown again, and participants rated its perceived realism with the question, "How realistic was this image?" using a continuous scale anchored at AI-generated (0) and Photograph (1).

Feedback. Procedure

The study was conducted in line with the born-open principle (Leeuw, 2024), ensuring transparency and reproducibility at every stage. The experiment was implemented entirely in jsPsych (De Leeuw, 2015), with the full code hosted publicly on GitHub, which also served as the platform for running the online study. Raw data were automatically stored in a private Open Science Framework (OSF) repository. Anonymized data, together with all preprocessing and analysis scripts, will be openly released on GitHub to facilitate complete reproducibility. Participants first provided informed consent before completing a short demographic questionnaire covering gender, age, ethnicity, country of residence, education, and English proficiency. Optional questions on birth control use were also included. They then proceeded to the experimental tasks.

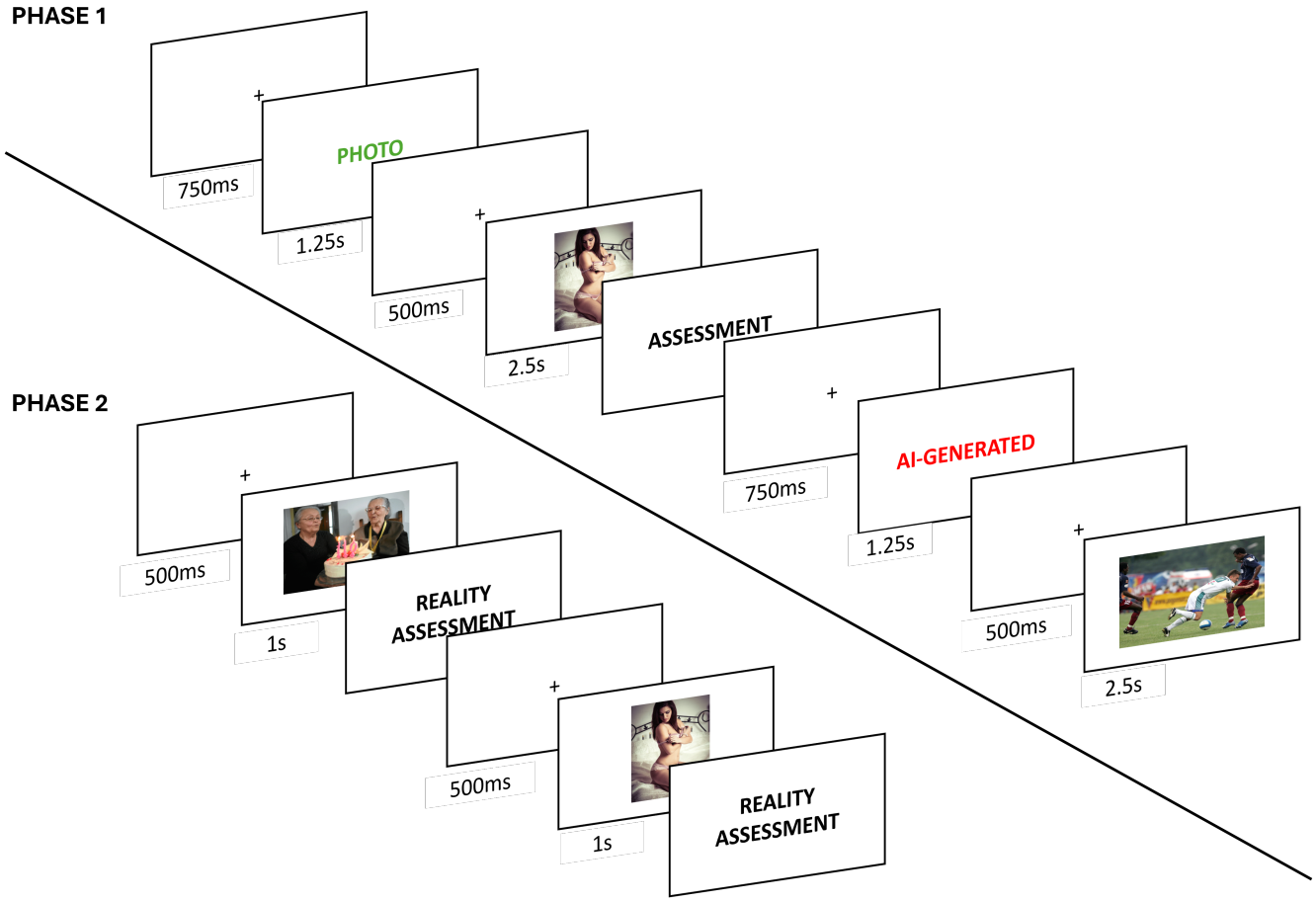
In the first phase, participants were told that the study aimed to validate a new image-generation algorithm. They were informed that they would see images allegedly produced by this algorithm intermixed with real photographs, each preceded by a cue indicating whether the upcoming image was of an "AI-generated" or "Photograph" origin. Their task was to rate each image on arousal, enticement, and valence. Each participant viewed 60 images in total: 40 erotic images (20 male and 20 female) from the Erotic subset of the Nencki Affective Picture System (NAPS ERO, Wierzba et al., 2015), and 20 additional images (10 neutral, 10 positively arousing) from the original NAPS database (Marchewka et al., 2014). Each trial followed a fixed timing sequence: a fixation cross (750 ms), a color-coded textual cue (1,250 ms), another fixation cross (500 ms), then the image (2,500 ms). Cues were presented in red, green, or blue, with colors randomly assigned across participants.

Following each image, participants rated their emotional response using three continuous sliders assessing sexual arousal, enticement, and valence. This phase was self-paced, with responses required before continuing. After completing the image-rating phase, participants filled out two self-report questionnaires: first the BAIT scale, followed by the COPS questionnaire.

In the final phase, participants viewed the same 60 images, presented in a new randomized order. Each was preceded by a 500 ms fixation cross and displayed for 1,000 ms. This time, participants rated each image on perceived realism—how photographic or lifelike it appeared.

Figure 1

Paradigm 2



At the end of the experiment, participants completed a feedback form. They were asked whether they could distinguish AI-generated from real images, whether AI images appeared more or less arousing, whether cue labels seemed accurate or reversed, and whether specific images stood out as particularly arousing or unarousing. Finally, participants were debriefed on the true purpose of the study: to examine how image labels (AI-generated vs. real photograph) influence emotional responses. Importantly, they were informed that all images were real photographs, and that the "AI-generated" label was used solely to test the effect of belief on affective reactions. A shareable link to the experiment was also provided.

Data Analysis

Study 2

Methods

Participants

The initial sample comprised 279 participants recruited via Prolific©. Inclusion criteria required participants to be native English speakers or residents of countries with high levels of English proficiency. Participant exclusions were applied as follows: five participants were removed for showing no variability in arousal ratings (i.e., they did not move the response scales). An additional five participants were excluded for completing the study on a mobile device. One participant was excluded due to displaying negative correlations between arousal and both enticement and valence. Furthermore, five participants who self-identified as neither female nor male, and two participants who reported a sexual orientation other than heterosexual, homosexual, or bisexual, were excluded from further analyses. Finally, one participant was removed because the stimuli presented were not relevant to their gen-

der and sexual orientation.

The final sample consisted of 261 participants (Mean = 37.4 ± 12.7, 48.7% Female). 56.32% of participants were from the United Kingdom, 26.82% from South Africa, 10.34% from the United States, and the remaining 6.51% were from other countries.

Ethical approval for this study was obtained from the School of Psychology Ethics Committee at the University of Sussex (ER/EB672/2).

Materials

Questionnaires. The questionnaires used in Study 2 were largely the same as those in Study 1, with minor modifications. In the BAIT, two items, assessing beliefs that AI might take control of people and interest in using AI systems in daily life, were removed. Additionally, the wording was streamlined by replacing "Artificial Intelligence" with "AI" throughout the scale. In the COPS, the item assessing the typical duration of pornography viewing sessions was omitted, retaining only items measuring frequency of pornography viewing and recency of sexual activity.

Affective Measures.

Arousal. Subjective sexual arousal was assessed following each image with the question, "How much did you feel sexually aroused?" Responses were recorded on 6-point Likert scale from Not at all (0) to Very much (6).

Enticement. Perceived enticement was measured after each image using the question, "How enticing would you rate this image to be?" with the same scale ranging from 6-point Likert scale from Not at all (0) to Very much (6).

Valence. Emotional valence was evaluated by asking, "The feeling evoked by the image was…" rated on a scale from Unpleasant (0) to Pleasant (6).

Reality. In a final stage of the experiment, each image was shown again, and participants rated on the images authenticity with the question, "I think this face is…Indicate your confidence that the image is fake or real" using a continuous scale anchored at AI-generated (-3) and Photograph (3).

Feedback.

Procedure

Consistent with Study 1, Study 2 was conducted in jsPsych following born-open principles (De Leeuw, 2015; Leeuw, 2024).Participants first provided informed consent, being informed that they could withdraw at any time; however, once the experiment was completed, withdrawal was not possible because the data were anonymized prior to storage. They then completed the same demographic questions as in Study 1, with the exception of items regarding birth control use.

In the first phase, participants were informed that the researchers were collaborating with a young AI start-up based in Brighton, intended to enhance the believability of the study. Participants were told that they would view images generated by this algorithm intermixed with "real" photographs, each preceded by a label indicating whether the image was AI-generated or a photograph. They were asked to rate each image on sexual arousal, enticement, and valence. A total of 50 images were presented, drawn from two categories of the NAPS-ERO database (25 images of couples and 25 images of individuals).

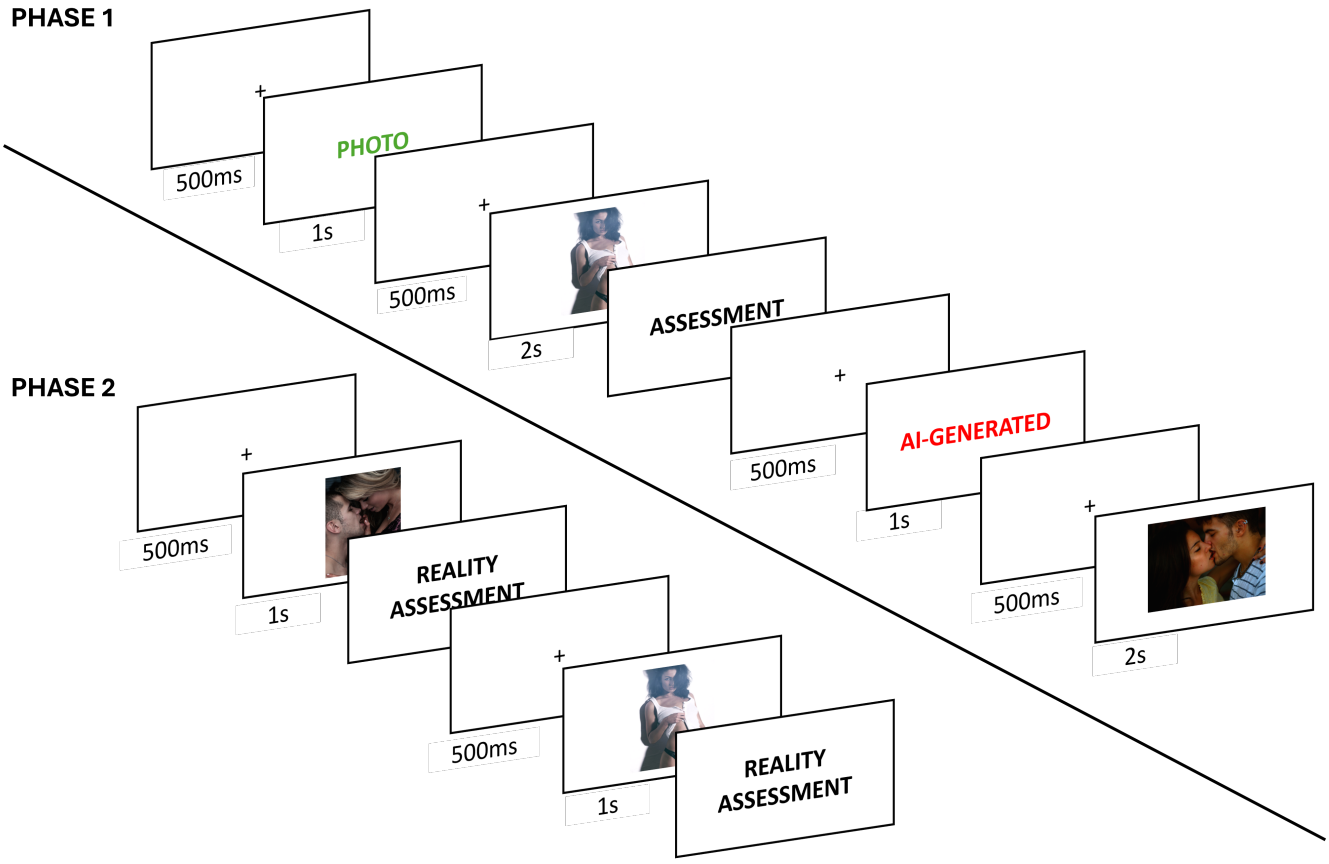
Images were assigned to be relevant to participants' self-reported gender and sexual orientation; for example, male participants identifying as homosexual viewed male individuals and male couples. Each trial followed a fixed timing sequence: a fixation cross (500 ms), a color-coded textual cue displayed for 1,000 ms, a second fixation cross (500 ms), and then the image presented for 2,000 ms. Cue colors were the same as in Study 1 and were randomly assigned across trials. Participants identifying as other in gender or bisexual/other in sexual orientation were asked which type of images they preferred, with options including "Women (and heterosexual couples)," "Men (and heterosexual couples)," "Only women (and lesbian couples)," and "Only men (and gay couples)."

Midway through the 50 images, participants were provided with a break and instructed to continue when ready. At this point, they completed a brief feedback survey assessing their subjective impressions of the images and AI-generation labels. This survey asked whether certain images were particularly arousing, whether AI-generated images were more or less arousing than the photographs, and participants' perceptions of the AI-generation algorithm. Specifically, they indicated whether differences between AI-generated and real images were obvious or subtle, whether they perceived inconsistencies or reversals in labeling ("Photograph" vs. "AI-Generated"), and whether they believed all images were either photos or AI-generated. If participants indicated that all images were real or AI-generated, they rated their confidence on a scale from "Not at all" to "Completely certain." This feedback captured participants' explicit beliefs and subjective reactions regarding both the content and labeling of the images.Following the first phase, participants completed the BAIT and COPS questionnaires.

In the second phase, participants were informed that some images had been intentionally mislabeled and were asked to judge whether each image was AI-generated or a real photograph, expressing their confidence at the extremes of the scale. The timing procedure in this phase was identical to Study 1. After each image, participants provided general feedback regarding their experience and any additional comments. Finally, participants were presented with a debrief page informing them that all images were, in fact, real photographs (see Figure 2).

Figure 2

Paradigm 2



Data Analysis

References

Corvi, R., Cozzolino, D., Zingarini, G., Poggi, G., Nagano, K., & Verdoliva, L. (2023). On the detection of synthetic images generated by diffusion models. *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5.

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

Graber-Mitchell, N. (2021). *Artificial illusions: Deepfakes as speech*.

Hatch, S. G., Esplin, C. R., Hatch, H. D., Halstead, A., Olsen, J., & Braithwaite, S. R. (2023). The consumption of pornography scale–general (COPS–g). *Sexual and Relationship Therapy*, *38*(2), 194–218.

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, *9*(1), 104–117.

Leeuw, J. R. de. (2024). DataPipe: Born-open data collection for online experiments. *Behavior Research Methods*, *56*(3), 2499–2506.

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The nencki affective picture system (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, *46*(2), 596–610.

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

Meskys, E., Kalpokiene, J., Jurcys, P., & Liaudanskas, A. (2020). Regulating deep fakes: Legal and ethical considerations. *Journal of Intellectual Property Law & Practice*, *15*(1), 24–31.

Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A., Krumhuber, E. G., & Dawel, A. (2023). AI hyperrealism: Why AI faces are perceived as more real than human ones. *Psychological Science*, *34*(12), 1390–1403.

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.