Brain-computer interface for generating personally attractive images

Michiel Spapé, Keith M. Davis III, Lauri Kangassalo, Niklas Ravaja, Zania Sovijärvi-Spapé, and Tuukka Ruotsalo

Abstract—While we instantaneously recognize a face as attractive, it is much harder to explain what exactly defines personal attraction. This suggests that attraction depends on implicit processing of complex, culturally and individually defined features. Generative adversarial neural networks (GANs), which learn to mimic complex data distributions, can potentially model subjective preferences unconstrained by pre-defined model parameterization. Here, we present generative brain-computer interfaces (GBCI), coupling GANs with brain-computer interfaces. GBCI first presents a selection of images and captures personalized attractiveness reactions toward the images via electroencephalography. These reactions are then used to control a GAN model, finding a representation that matches the features constituting an attractive image for an individual. We conducted an experiment (N=30) to validate GBCI using a face-generating GAN and producing images that are hypothesized to be individually attractive. In double-blind evaluation of the GBCI-produced images against matched controls, we found GBCI yielded highly accurate results. Thus, the use of EEG responses to control a GAN presents a valid tool for interactive information-generation. Furthermore, the GBCI-derived images visually replicated known effects from social neuroscience, suggesting that the individually responsive, generative nature of GBCI provides a powerful, new tool in mapping individual differences and visualizing cognitive-affective processing.

Index Terms—Brain-computer interfaces, Electroencephalography (EEG), Generative Adversarial Networks (GAN), Image generation, Attraction, Personal preferences, Individual differences.

✦

1 INTRODUCTION

WHAT is beauty? Although in daily life we can instantly judge whether a picture looks attractive, we commonly find it hard to explain the reasons behind such a decision and harder still to create beauty without extensive skill and experience. The difficulty in describing and portraying attractiveness stems from two interrelated problems. First, ratings of attractiveness vary significantly between individuals, for example in terms of age, culture, and gender [13], [45], [68]. Second, describing what one finds aesthetically pleasing requires awareness of what is thought to be an implicit evaluation of a complex configuration of features [34]. Therefore, attractiveness is rather a subjective, personal characteristic than a objective, visual feature: Beauty is in the eye of the beholder.

The creation of personally attractive images has thus far been a challenge due to the complex nature of attractiveness. Previous studies sought to determine the attractiveness of a face by relying on simple, predefined features computed from a picture [49], [56], [63], [65], but in so doing have likely underestimated the complexity of attractiveness judgments. Models relying on hand-crafted features through simple methods (e.g. length of nose measurements, golden

- K. M. Davis, L. Kangassalo, and Z. Sovijärvi-Spapé are with the Department of Computer Science, University of Helsinki, Finland. T. Ruotsalo is with the Department of Computer Science, University of Helsinki, Finland and Department of Computer Science, University of Copenhagen, Denmark.
- M. Spapé and N. Ravaja are with the Department of Psychology and Logopedics, University of Helsinki, Finland.
  E-mail: michiel.spape@helsinki.fi

ratio, symmetry) or more higher-level algorithms (e.g. Gabor wavelet transformations) do not reflect an individual's understanding of aesthetics [34]. Thus, such approaches fall short in modeling psychologically relevant sources of attractiveness and cannot enable inverse inference by producing novel, attractive images. This is because the automatically captured features are limited to the observable distribution of images within the models, which likely represents only a small portion of the true distribution of features that constitute variance in the visual image features.

Besides attractiveness being a complex, personal perceptual decision, it may also be better judged implicitly than explicitly. Humans typically respond emotionally to attractive images [9] rather than purely on the basis of rational, visually salient reasons. In order to generate attractive images, an ideal system should therefore rely on early, implicit responses rather than explicit ratings. This could be implemented as a physiologically adaptive system, such as a brain-computer interface (BCI), which would utilize implicit signals to generate personally attractive images. However, the aforementioned complexity problem limits the use of BCIs in this field. For example, if one face is implicitly evaluated as more attractive than another, it will likely differ in multiple ways. How do we infer which of the features is important, and how can we generate another face that is expected to be similarly attractive to the target face?

Here, we present a new paradigm, which utilizes implicit reactions to perceived attractiveness to steer a generative adversarial network (GAN) [30], thereby producing images that are expected to be personally attractive. We refer to the approach as generative brain-computer interfacing (GBCI). By training a GAN with images, it learns to mimic the

underlying visual distribution, which enables us to draw new, unobserved samples from that distribution [7], [38], [47]. GBCI unites a GAN with BCI: in response to a series of evoked brain responses to images, the GBCI iteratively produces novel (previously unseen), photorealistic images that match a user's individual aesthetic preferences. As further explained in Figure 1, the GBCI works by classifying brain activity evoked by faces being perceived as either attractive or unattractive. Each face is represented as a coordinate within the GAN space, so that with multiple attractive faces being detected, we can triangulate GAN vectors that are expected to be subjectively attractive. This expected-attractive localization is iteratively updated whenever more evidence is detected, each time producing a novel image for the position. The final position is expected to match the participant's sense of personal attractiveness, which we empirically test in the present study.

We report an experiment with 30 participants to validate the GBCI in its ability to generate attractive faces. A GAN model trained with celebrity faces was used to create a sample of fictional faces. These were then presented to each participant while their EEG was recorded. The GBCI paradigm was then applied to iteratively generate the image that was predicted to be the best match for a user's personal attraction. To empirically test the GBCI's efficacy, we requested users to blindly evaluate the best personal match, expecting higher personal attractiveness ratings than for control images.

In summary, we present generative brain-computer interfacing for generating personally attractive images:

1) This is, to our knowledge, the first successful approach utilizing brain responses as an interactive feedback to a generative neural network.
2) The approach was validated in a face generation task and found to generate novel, personalized, and highly attractive images.

2 BACKGROUND

In the following section, we will first review the current state of research on the psychology of attraction before we focus on brain processes involved in aesthetic judgements. We then move our focus to research showing that general, but not personal, attraction can be predicted from computer vision algorithms. Likewise, the brain processes reviewed earlier can be harnessed to enable automatic detection of aesthetic relevance. In the subsequent section, we will propose GBCI as a new approach to unite the psychology of aesthetics, the cognitive neuroscience of personal attraction, and the computer science of generative adversarial models.

2.1 The psychology of personal attraction

The study of aesthetics, or the perception of beauty and experience of attractiveness, has a long tradition within psychology and related disciplines. Despite the common idea that taste is intensely individual, psychological research consistently shows a strong consensus on the visual features that are considered attractive [46]. Symmetry in faces is known to be seen as attractive, perhaps because symmetry in general is an important evolutionary signifier. Visual

symmetry, for example, may point towards the nearby presence of fruit, flowers, and animals [70]. Indeed, it is even thought that positive affect is a consequence of the computational ease of processing due to symmetry necessarily including a redundant set of visual features, streamlining perception [55]. Another common theory in evolutionary psychology is the sexual dimorphism account of facial attractiveness, which holds that feminine displays in females and masculine in males are attractive by signifying mate quality [41]. It seems, however, that evolutionary factors shaping visual feature perception do not entirely predict attraction: Computer simulations that optimize averageness or sexual dimorphism generate faces that are judged attractive but not maximally attractive faces for individuals [61].

The common consensus on what is beautiful notwithstanding, individual differences in attractiveness judgments do exist. Interestingly these differences have been found to be larger for female ratings of male faces than the other way around [36]. It is clear, however, that attractiveness is not solely due to biological, genetic factors: Subjective levels of attractiveness vary widely as a function of social learning [49]. Typically, debates of biological and cultural determinants of individual differences in faces are presented in nature-vs-nurture terms, but cultural differences do not negate evolutionary theory, or vice versa. Indeed, Darwin himself already noted the tremendous differences in what people find beautiful [14]. A lack of common consensus in beauty may itself be adaptive, as variations within and between cultures present an evolutionary advantage [50].

Yet, however much people differ in what they find personally attractive, cognitive neuroscience suggests their brains process attraction in very similar ways. Consequently, if the cognitive and affective processes underlying aesthetic relevance are comparable between individuals, then it should allow us to detect whenever a stimulus is deemed attractive, which can then be used as the GBCI's implicit measure. In the next section, we will discuss the relevant literature on the neural response to detecting attractive stimuli.

2.2 The cognitive neuroscience of personal attraction

The event-related potential (ERP) technique within electroencephalography research presents a useful method for detecting when a user perceives a beautiful face. ERPs are electrophysiological recordings of brain activity occurring in response to known physical events. Since EEG provides high temporal precision, it is possible to functionally dissociate perceptual from cognitive operations. Thus, by detecting whether an ERP is in response to an attended letter, the classic brain-computer interface (BCI) allows people to spell letters using EEG [25]. More controversially, this approach was extended to situations in which participants would rather hide their thoughts, but failed to disguise their implicit guilt by having guilty knowledge [24] (but see [53]). Due to this enticing possibility that hidden evaluations might be detected from EEG, ERP research is more and more conducted in the context of human-computer interaction research [66].

In the present work, we focus on using properties of the ERP that can inform us as to the aesthetic preference of

users. As preference entails a stimulus being found intrinsically relevant, we targeted the ERP component that has been particularly associated with relevance detection, the P300. The P300 in general is characterised by a late parietal positivity occurring from ca. 300 ms after the onset of stimuli in any modality if they are infrequent, attended, novel, and relevant [12].

Later research, however, suggested the P300 can be further divided into multiple subcomponents, of which at least three seem related to aesthetic relevance detection. The *P3a* predominantly affects frontal-central electrodes and usually leads the other subcomponents in terms of latency, and has functionally been related to exogenous relevance, responding to stimuli that are novel or affectively evocative [11]. The *P3b*, meanwhile, is the more commonly targeted component in EEG research, and has generally been related to top-down, endogenous, task-relevance related attention, being modulated if stimuli of any modality [31] require additional processing [42]. As such, it has theoretically been seen as corresponding to a neurodynamic mechanism related to working memory updating [19], or a process in between attention and further memory processing [58]. It is also the more reliable part of the P300, especially if stimuli are improbable and require a mental or physical response [57], [73]. Consequently, this component is often the one targeted by BCIs, such as with the classic BCI-speller [25] or applications utilizing relevance effects [22], [23], [35].

Finally, a further sub-component following the P3a and P3b is sometimes referred to as the *LPP*, the late positive potential, which has particular importance to attractiveness research as it was shown enhanced on seeing beloved partners [44], and aesthetically pleasing images [32].

To maximize the degree to which implicit attractiveness evoked P300s, we employed three strategies based on the literature. First, attractive images were made relatively improbable by presenting images of both the participants preferred and non-preferred gender. Thus, for the heterosexual majority, fewer than 50 % of faces were both of preferred gender *and* attractive. Second, we requested participants to select unattractive images as reminders. These were shown to the left and right of a screen to focus users on the central, target image. Third, we asked participants to focus particularly on attractive images by mentally counting their occurrence. This showed that only about 1 in 5 images was found attractive, matching relevant probability in traditional BCIs [25].

2.3 Affective computing of personal attraction

Our work aims to detect personal attractiveness of images based on implicit brain signals in order to optimize a prediction of personal attraction. Our neuroadaptive interface is designed to detect and predict *personal attraction* to visual images, which is related to the affective computation of attractiveness. Traditionally, computer vision has been used to extract complex features related to aesthetics from images and using these to predict user evaluations. The attractiveness of images in general has been done based on automatic computation of low-level features, such as quality [64], or more cognitively significant features, such as based on *Gestalt* principles, rules of thirds, and visual weight [15],

[40], [52]. For images of faces, visual features such as nose-to-forehead and nose-to-chin ratio were found to predict how attractive a face is found with an accuracy of about 25% [20]. Recent years have enhanced the efficacy of this approach by extracting a combination of visual features from image input and applying more advanced machine learning algorithms to predict general attractiveness and affective ratings [2], [3], [28], [62]. However, extracting invariable features from a stimulus input necessarily optimizes prediction of *general attractiveness*, dismissing interindividual variance in what counts as attractive as merely noise.

Instead of solely relying on the objective visual features of images, our work aims to infer what a particular individual finds personally attractive and therefore targets the subjective qualities evoked by an image. Here, we rely on the growing literature on neuroadaptive computing [43], which has previously focused on adapting a system via neurofeedback in a very limited set of pre-defined states. Examples of such neurofeedback systems include changing the difficulty of a learning task to avoid overwhelming the user's cognitive capacities [75], or transforming a player's character based on brain signals [72]. In contrast, our work is based on the neuroadaptive framework by [37] towards optimization of an inference of personal attraction within a complex, multidimensional space represented using a generative adversarial network (GAN). Adapting the hypothetical "best guess" of what each individual finds attractive, iteratively retrieving this point within the GAN generates a visualization of the personal attraction.

To summarize, neuroadaptive systems have shown solid promise in adapting an inference based on brain activity. The advent of GANs now enables extension of this framework by allowing complex adaptations within ill-constrained problem spaces [71]. Additionally, this provides the possibility to generate photorealistic, yet artificial images of human faces, producing a novel window into mental processes. In other words, a generative BCI is a neuroadaptive system with a deep-learning representation of image data operating on individual users, which we believe to represent the next natural step in predicting and visualizing personal attraction.

3 A GBCI FOR PERSONAL ATTRACTION

Personal attraction as modeled via generative brain-computer interfacing consists of four phases, reflecting the example illustrated in Figure 1. In this section, we first provide a general overview of how the GBCI functions. In the following four subsections, we formally define each of these phases using mathematical terms.

A GAN is first trained (A) using the CelebA-HQ dataset. Next, a participant is asked to assess images randomly sampled from the GAN while their EEG is recorded; a classifier is then trained to associate their EEG with their subjective assessments of the images (B). After calibration, new images sampled from the latent space are shown to the participants while their EEG is recorded. The EEG signals are then classified to determine which images the participants found attractive (C). In the example, images 1, 3, and 4 are found attractive and image 2 is not found attractive. The GBCI then tries to find optimal values for latent features
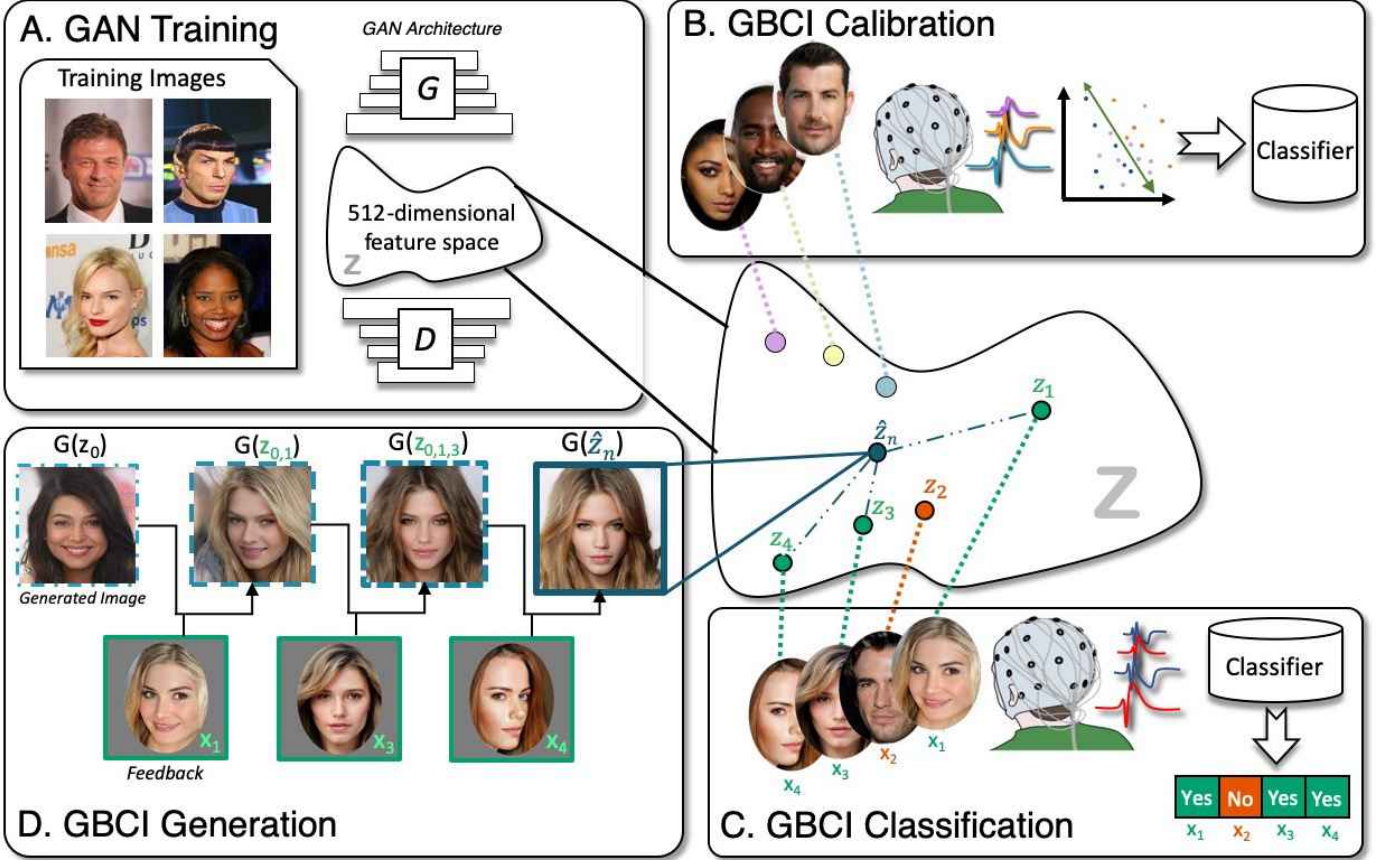
Fig. 1: The GBCI approach. **A:** A GAN model with generator G and discriminator D is trained using ca. 200k images of celebrity faces, resulting in a 512-dimensional latent space from which sampled feature vectors used as Generator input produce artificial images; **B:** Participants are shown images produced from sampled feature vectors while their EEG is measured; Following, they are shown the same images and select based on personal attractiveness; These collected data are then used to train an LDA classifier for each participant; **C:** Participants are shown new images produced using the same generative procedure as in **B**; Now, their measured EEG responses are classified as attractive/unattractive using their personal classifier; **D:** New images are generated from the latent representations (i.e. feature vectors) of images labeled by the classifier as attractive. An image $G(\hat{z})$, estimated as personally attractive, is iteratively generated as more images are classified as attractive and their combined feature vectors z_i are used as inputs for the Generator.

within the GAN model that encode for attractive features. Finally, a new image containing optimal values for attractive features and non-attractive features is generated (D). The key idea is that certain stimuli images have some features that the participant finds attractive, but these features are not necessarily all contained within a single stimulus image. On the other hand, some images contain features which the participant may find unattractive. In the example, image 2 is an image of a male that is not attractive for the participant and thus has some features that are to be avoided. Now, the GBCI is able to parameterize a vector that combines features from the latent vectors of images 1, 3, 4, and features not in the latent vector of image 2. The resulting latent vector then corresponds to an image that is hypothesized to be maximally attractive for the participant.

3.1 Phase A: GAN Training

First, a latent image space is created by training a progressively growing generative adversarial network [38]. This results in a 512-dimensional feature space Z. The generative model provides a mapping $G : Z \to X$, where $z \in Z$ is a point in the latent feature space Z, and $x \in X$ is an individual face in the set of faces. The goal is to produce a feature space where it is possible to select a point $\hat{z}_n \in Z$, for which $G(\hat{z}_n) = \hat{x}_n$, and determine if it matches the attractiveness criteria of the participant.

3.2 Phase B: GBCI Calibration

Next, the feature space Z is sampled to produce images of artificial faces using the GAN architecture. These images are presented to a participant. The brain activity $S_n = \{s_1, ..., s_n\}$ associated with the images is used to calibrate a regularized Linear Discriminant Analysis (LDA) classifier [5] with shrinkage chosen with the Ledoit-Wolf lemma [48]. The classifier learns a classification function $f : S \to Y$, where Y is a binary value discriminating target/non-target stimuli. In this case, detecting that there were some features in the presented face x_i that was attractive for the participant.

3.3 Phase C: GBCI Classification

In Phase C, a set of n images $X_n = \{x_1, \ldots, x_n\}$ that were not in the training set of the classifier are generated from a set of latent representations $Z_n = \{z_1, \ldots, z_n\}$. This information is displayed to the participant, whose brain responses evoked by the presented information are measured. These responses are then classified using the trained classifier, and their associated presented images x_i and the latent vectors z_i used to generate these images are assigned labels corresponding to the classifier outputs. The classifiers are personalized and a separate classifier is trained per participant. In the example case presented in Figure 1, images x_1, x_3 and x_4 are found to be attractive, while image x_2 is not found attractive. This results in a set of images x_i, their associated latent vectors z_i, and a set of binary labels of whether an image is found to be attractive. Let us refer to the set of latent vectors classified as attractive as Z_{POS}.

3.4 Phase D: GBCI Generation

Finally, in phase D, $\hat{z}_n$ is updated by using a simple model updating function $h : Z, Y \to Z$ to generate a final image from the latent model. Formally, this average vector was computed with $\hat{z}_n = \frac{1}{|Z_{POS}|} \sum_{z_j \in Z_{POS}} z_j$. This updating procedure is a special case of the Rocchio algorithm [60].

The resulting $\hat{z}_n$ is then used as an input to the generator of the latent GAN model G to generate a new image, $G(\hat{z}_n) = \hat{x}_n$. The resulting image represents the point in the latent space that is a novel, unseen face image, which is expected to contain the attractive facial features. The process starts from the first positively classified image, such that the vector $\hat{z}$ is initialized with the corresponding latent vector $\hat{z}_0$ of that positively classified image

4 USER EXPERIMENT

To evaluate the approach, we present an experiment that tested whether GBCI could generate images that were evaluated as personally attractive by 30 participants. The experiment was run in two stages. In stage I, the pre-trained GAN was used to produce 240 face images that were used as stimuli. Participants viewed the images in a rapid serial visual presentation (oddball) paradigm, particularly concentrating on attractive images (relevant targets). Based on their data, we trained a classifier to detect relevance from their brain responses. This information was then used as a positive model for generating individual, novel images that were expected to be personally attractive to the participants. In stage II, the participants explicitly evaluated the generated images randomly placed along with matched controls in a blind test. We hypothesized that the positive-model generated images would be evaluated as more personally attractive than matched controls.

4.1 Participants

Thirty-one volunteers were recruited from the student and staff population of the University of Helsinki. During the recruitment procedure, volunteers were informed that participation required they state their sexual gender preference. They were fully informed as to the nature of the study, and signed informed consent to acknowledge understanding their rights as participants in accordance with the Declaration of Helsinki, including the right to withdraw at any time without fear of negative consequences. One volunteer did withdraw due to lack of time and was removed from data analysis. The full sample thereafter included 17 males and 13 females, aged 28.23 (SD = 7.14, range = 18 to 45) years on average. The study was approved by the University of Helsinki's Ethical Review Board in the Humanities and Social and Behavioural Sciences. Participants received one cinema voucher for participating in the acquisition phase of the experiment, and two more for completion of the validation phase.

4.2 Stimuli

A pre-trained Generative Adversarial Network (GAN) [38][1] was used to generate all stimuli used in this study. The GAN was pre-trained with the CelebA-HQ dataset, which consists of 30 000 1024×1024 images of celebrity faces. The CelebA-HQ dataset is a resolution-enhanced version of the CelebA-dataset [51]. The generator part of the GAN provided a mapping from a 512-dimensional latent space to a 1024×1024 image. Only visual features (i.e. pixel data) from the dataset were used to train the GAN. Other data included in the Celeb-HQ dataset, including manually annotated labels describing various features contained within an image, were *NOT* used in any way to train the GAN model.

Training images were initially generated via a random process that sampled latent vectors from a 512-dimensional multivariate normal distribution. These were then used to produce corresponding images with the GAN generator. Following, images were manually categorized by a human (female) assessor into male and female-looking faces, without regard for other visual attributes (e.g. age, ethnicity, emotional expression), other than looking convincing and being without significant artifacts. We then selected the first 120 male and 120 female faces for use in the present study. To standardize the images with regards to face-unrelated attributes such as the background, we removed the surrounding area of the original 1024 x 1024 sized images using a 746 x 980 pixels elliptic mask, replacing this with uniform gray. To further improve presentation timing accuracy, we then downsampled the images by a factor of 2. The images were displayed on a 24" LCD monitor placed ca 60 cm from the participant, running at 60 Hz with a resolution of 1920 x 1080, its timing and EEG synchronization optimized using E-Prime 3.0.3.60 [67].

For the evaluation procedure, three sets of stimuli were generated: positive, negative and random images. Positives and negatives were generated by averaging the latent vectors representing images detected as attractive or unattractive, respectively. The matching controls were generated similarly, but by averaging over randomly chosen latent vectors. This procedure simulated a random-feedback classifier. The formal averaging and generation process is provided in the Image generation section (4.6).
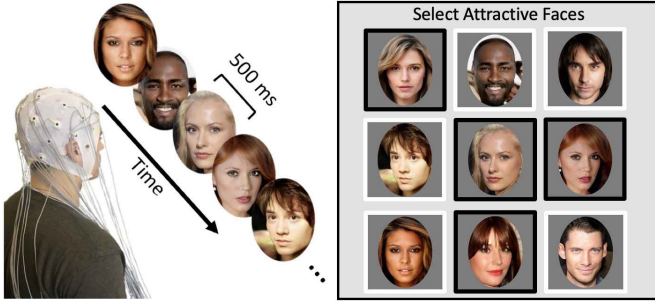
Fig. 2: Data acquisition procedure. During the RSVP, 8 x 60 images were presented at a rate of 2 stimuli / s. Following the RSVP, users provided explicit feedback by clicking on attractive images.

4.3 Stage I: Data acquisition procedure

In stage I of the experiment, participants viewed the images in a rapid serial visual presentation paradigm. Following EEG setup and signing of forms, the experiment was started by the lab assistant. The participants were asked in private to provide their sexual gender preference with options given between male, female, or either. The experiment included two parts: a presentation and a feedback. During the former part, participants undertook 8 rapid serial visual presentation (RSVP) trials. Each trial started by displaying 4 random images of a male or female face (order randomised), asking participants to click on the face they found the least attractive. This image was subsequently used as mismatch "flanker". Following this, an instruction screen appeared to remind the participants of the task, which was to concentrate on attractive images by keeping a mental count. After acknowledgment, the flankers were presented left and right of the central, target location, for 1 s before the RSVP started, which involved sequential presentation of 60 (30 male, 30 female) images being presented in the central location at a rate of 2 per second without inter-stimulus interval. After the last image, participants were requested to enter the number of attractive images using the keyboard. One block had 8 trials, such that after a block (480 images) all 240 images were presented twice, once flanked by males, and once by females.

The feedback part was started after each of the three blocks. Here, participants selected, or "voted for", images they found attractive. Participants were shown all 240 images in random order, using four screens of 60 small buttons laid out in 5 rows of 12 images, and asked to click on the ones they had previously found attractive. To avoid changes in decisional preference criterion, an estimate was shown in the corner of the screen, indicating the number of images they had provided previously, counting down as participants selected images. After 3 blocks, taking on average 37.15 (SD = 6.97) minutes in total, the experiment was complete.

4.4 EEG acquisition and preprocessing

A BrainProducts QuickAmp USB was used to record EEG from 32 Ag/AgCl electrodes positioned at equidistant locations of the 10-20 system by means of an elastic cap, using a single AFz electrode as online reference. The time series voltage amplitudes were then digitised with Brain Vision Recorder running at a samplerate of 1000 Hz, with a high-pass filter at 0.01 Hz, and a re-referencing to the common average reference. Furthermore, two pairs of bipolar electrodes – one pair placed lateral to the eyes and the other above and below the right eye – were used to capture EOG. Offline preprocessing included application of a band-pass filter between 0.2 – 35 Hz to remove slow signal fluctuations and line noise, after which the data were time-locked to stimulus-onset and segmented into 900 ms epochs of post- and 200 ms of pre-baseline activity. After removing the average baseline activity from each channel, epochs contaminated by artefacts such as eyeblinks were tagged with an individually adjusted, threshold-based heuristic. As a result, approximately 11% of each participants' epochs with the highest absolute maximum voltage was removed from analysis. In order to speed up classifier training procedures, the data were decimated by a factor of four. The final dataset consisted of on average 1265 (SD = 109) epochs per participant.

Standard feature engineering procedures were followed to form a vectorized representation of the EEG data [5]. After preprocessing, the measured scalp voltages of each participant were available as a $X^{n \times m \times t}$ tensor, with n epochs, m channels and t sampled time points . Each epoch was split into $t' = 7$ equidistant time windows on the 50 - 800 ms post-stimulus period, and the measurements in each window were averaged. To generate spatio-temporal feature vectors all available channels and the t' averaged time points were concatenated, resulting in a data matrix $X^{n \times m \cdot t'}$, where $m \cdot t' = 32 \cdot 7 = 224$.

4.5 Classifier

The attractiveness of faces was predicted in a single-trial ERP classification scenario [5]. In detail, a regularized Linear Discriminant Analysis (LDA) [27] classifier with shrinkage chosen with the Ledoit-Wolf lemma [48] was trained for each of the participants with the vectorized ERPs. LDA has been shown to perform robustly for EEG classification [5]. Also included were binary labels indicating attractiveness of the faces associated with the vectorized ERPs (attractive/unattractive). The label was assigned based on the attractiveness votes (clicks) of each face, which were given by the participant during the feedback phases: zero votes for a given face labelled the face as unattractive, while two or three votes labelled the face as attractive. Faces with only one vote were deemed to be of unknown attractiveness and removed from further analysis. This led to the average participant having 1168 (SD = 123) data points prior to splitting the datasets to training and test sets. The split was done so that the first 80% of the data points were used for training and the remaining 20% for testing. The test set thus contained approximately 233 ERPs per participants.

After the training, the vectorized ERPs in the test set were assigned to either the attractive or unattractive class based on classifier confidence. For this purpose, two per-participant thresholds were computed: one for the attractive class (positive prediction threshold) and one for the unattractive class (negative prediction threshold); a face was predicted to be attractive if the classifier confidence for the

attractive class exceeded the positive prediction threshold and unattractive if it fell below the negative prediction threshold. To ensure that the images generated from the unattractive and attractive predictions were of equal quality, the thresholds were computed so that the amount of positive and negative (attractive/unattractive) predictions made by the classifier for a participant were equal. This resulted in a grand average of 41.13 (SD = 42.36) positive and negative predictions by the classifiers.

The classifier performance was measured with an Area Under the ROC Curve (AUC), and evaluated by permutation-based p-values acquired by comparing the AUC scores to those of classifiers trained with randomly permutated class labels [54]. For each participant, $n = 100$ permutations were run, meaning that the smallest achievable p-value was 0.01 [29].

4.6 Image generation

Using the same GAN architecture that produced the face images used as stimuli (see 4.2), five different configurations of the GBCI model were designed for generating the evaluation images. The first configuration, POS, was a positive feedback model that used only the latent vectors of the face images classified as attractive A_{POS}. The second configuration, NEG, used only negatively classified vectors A_{NEG} (i.e. the latent vectors of the face images classified as unattractive). The third configuration, RND, used random feedback A_{RND}, where the labels for vectors used for the POS and NEG models were shuffled. The fourth configurations, POS-NEG, subtracted the latent vectors used in the NEG model from the latent vectors used in the POS model. The fifth configuration, NEG-POS, subtracted the latent vectors used in the POS model from those of the NEG model.

Operationalizing our hypothesis towards these models, we expected that the POS model would generate faces that were evaluated as more attractive than those generated by the RND and NEG models. Furthermore, we expected that the NEG model would produce faces that were evaluated as less attractive than the baseline provided by the RND model. As we had no clear a-priori hypothesis with regards to the POS-NEG model, or the NEG-POS model, we left these out of the confirmatory tests and analysis.

4.7 Stage II: Image evaluation procedure

In stage II, a blind evaluation procedure was used to test the hypothesis that GBCI generated images were more personally attractive than matched controls. Two months after initial participation, we recalled the participants for the follow-up validation procedure, in which they evaluated their custom-generated images. Single generated images from each of the models described in 4.6, along with 20 matched controls generated from the RND model, were empirically tested for personal attraction using two tasks that were presented in a set order. Following, an interview was conduced with the participants in order to obtain qualitative data.

In the *Free-selection* task, the images were simultaneously presented in 2 rows of 12 (similar to the feedback phase described earlier) randomly arranged buttons, and participants were requested to click on all which they found attractive. We analyzed the percentage of times images expected to be found personally attractive and unattractive were selected.

In the *explicit evaluation* task, the 24 images were sequentially presented in random order, and participants were requested to rate the attractiveness of each using a 1 (very unattractive) to 5 (very attractive) Likert-type scale. To distinguish personal preference from general judgments related to cultural norms or demand characteristics, we asked participants to perform the task twice. During the second run, we asked participants to estimate how attractive the general population, given compatible orientation, would rate the person. Thus, the explicit evaluation provided both measurements of personal attractiveness, and estimations of population attractiveness. We analyzed the average ratings for the three types of images – expected to be attractive, unattractive, and neutral – using repeated measures ANOVAs.

Finally, the personal predictions were revealed to the participants and a semi-structured *user interview* was conducted with the aim of determining whether participants felt the generated images matched their personal attraction. Guiding interviewees to reflect on the process of the study, we explored the phenomenology of attraction in the context of the experiment using thematic analysis [6]. In particular, the participants reflected on the epistemology of attraction in terms of how they defined and experienced attraction. The interview was digitally recorded and answers were transcribed for 8 randomly selected users. The complete validation procedure, including free selection and explicit evaluation tasks took ca. 20 minutes.

5 RESULTS

The generative brain-computer interface (GBCI) used event related potentials (ERPs) to create an attractiveness classifier, which was then used to generate images that were empirically tested for matching personal attraction. In the results, we first present the generalized effect of perceiving attractive images on ERPs to confirm the expected pattern was visible on the P3. However, this general analysis did not affect individualized classifiers, the effectiveness of which we describe in the subsequent section. The classifier was then used to generate a set of novel images that were hypothesized to match (positive generated images) and not match (negative generated imagers) personal attraction. The third section presents the results of the empirical test of the GBCI to produce personally attractive images. The final section summarizes how participants experienced the decision-making process and the perceived efficacy of the GBCI.

5.1 ERP results

ERPs to images voted as unattractive, attractive, and inconsistent were averaged per participant and analyzed for Fz and Pz channels. As effects were predicted for both P3a (commonly earlier and frontal), and P3b (usually parietal and later) potentials, we first performed a confirmatory
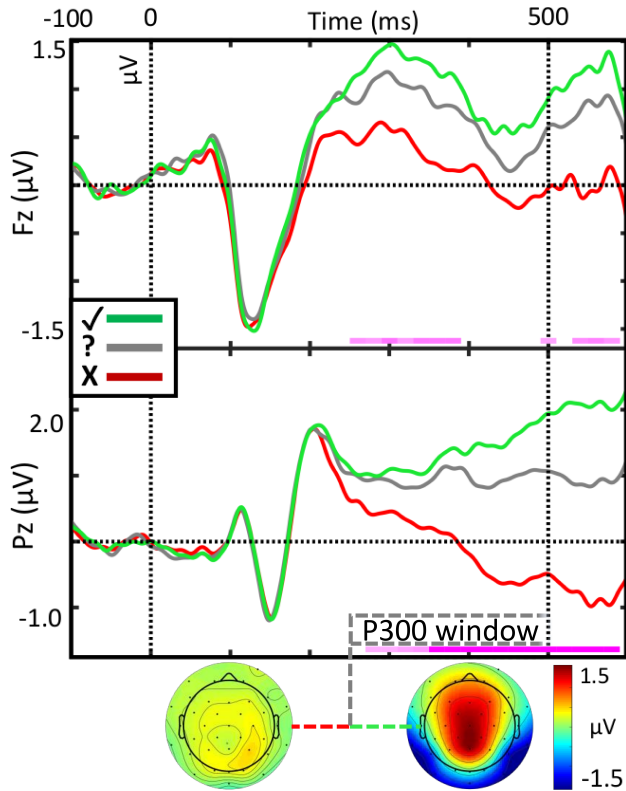
Fig. 3: Grand average ERP from the Fz (top) and Pz (bottom) showing evoked responses to faces consistently deemed attractive (green) or unattractive (red) and inconsistently rated faces (grey). Pink lines to the bottom of each graph show significant differences between conditions (Bonferroni corrected p < .05). Scalp topographies displaying the difference between unattractive (left, red), and attractive (right, green) between 250-500 ms is shown below the panel.

analysis on the average amplitude between 250-350 for Fz (P3a) and between 350-500 for Pz (P3b) using two repeated measures ANOVAs with attractiveness (unattractive, inconsistent, attractive) as factor and component (P3a, P3b) as measures to replicate the effect that observing attractive, relevant images evokes a predictable pattern on average. A significant effect of attractiveness was observed for the P3a, $F (2, 58) = 15.51$, MSE = 0.33, $p < .0001$, $\eta^2 = .12$. Post-hoc comparisons showed inconsistent and attractive images evoked larger P3as than unattractive images, $ps < .01$, and that attractive images evoked higher amplitudes than inconsistent ones, $p = .01$. The effect was also significant for the P3b, $F (2, 58) = 51.24$, MSE = 0.53, $p < .0001$, $\eta^2 = 0.50$. Again, attractive images significantly amplified the P3b relative to unattractive, $p < .001$, and inconsistent images, $p < .001$. Here, inconsistent showed P3bs roughly in between unattractive and attractive images, as it was found to also evoke amplified P3bs vs unattractive images, $p = .005$.

To provide a more comprehensive analysis, we furthermore explored the univariate effect of attractiveness on the entire time-series of Fz and Pz activity using two windowed repeated measures ANOVAs with the average amplitude of Fz and Pz between 100-600 ms in bins of 20 ms as measure. Figure 3 shows the result of these tests with short pink lines under any interval in which a significant (Bonferroni-

corrected $p^* < .05$ equals $p < .00096$) effect is observed. This indicates a somewhat earlier effect of preference on Fz (between ca 240-400 ms) than Pz (290-600 ms), likely coinciding with a difference between P3a and P3b. For both electrodes, attractive faces evoked more positivity, with inconsistent faces roughly in between non-preferred and preferred. Given that effects were observed mainly after 250 ms, we can infer that the GBCI likely did not benefit from a flanker-induced N2 effect, and instead primarily relied on patterns of activity within the P300 range.

5.2 Classification results

Classification results for all participants are shown in Figure 4. Within-subject permutation-based significance tests at $p < .05$ showed that the classifiers performed significantly better than random baselines for 28 out of 30 participants, far better than chance level (less than 2 out of 30). Across subjects, classifier performance had an average AUC of 0.76, min = .61, max = .93. This suggests that the classifiers were able to find significant structure in the data that discriminated brain responses for attractive and unattractive faces. These classifications were then employed to generate the preferred/non-preferred faces, as shown in Figure 5.
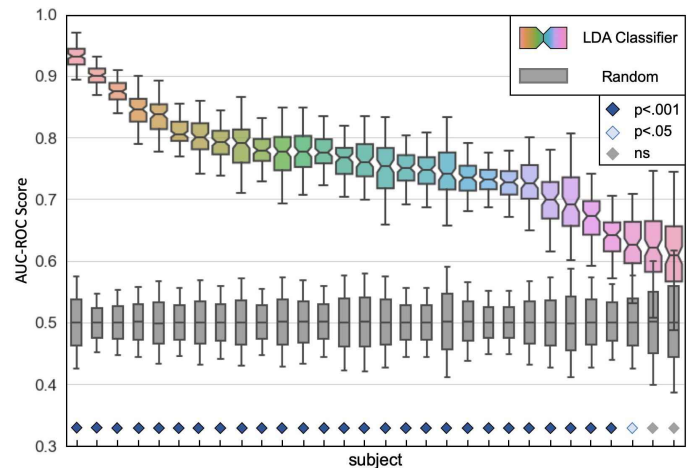

Fig. 4: AUC scores across participants, with scores for random baselines computed using label permutation in gray. Diamonds placed below boxplots indicate within-subject significance of classifier performance against random.

5.3 Generated image evaluation results

The latent vectors of the faces expected to be found attractive were used to generate images such as those displayed in Figure 5. To validate the GBCI in its ability to generate personally attractive images, we performed two empirical tests.

As can be seen in Figure 5 panel B, the results from the *free selection task* showed that the positive generated image was selected as attractive in 86.7% of cases from among 24 images (i.e. for 26/30 participants), while the negative was selected in 20.0% (SE = 3.2%). In other words, there were 86.7% true positives, 80.0% true negatives, 20.0 % false negatives, and 13.3% false positives. Therefore, generative

A: Individually generated faces for 16 individuals
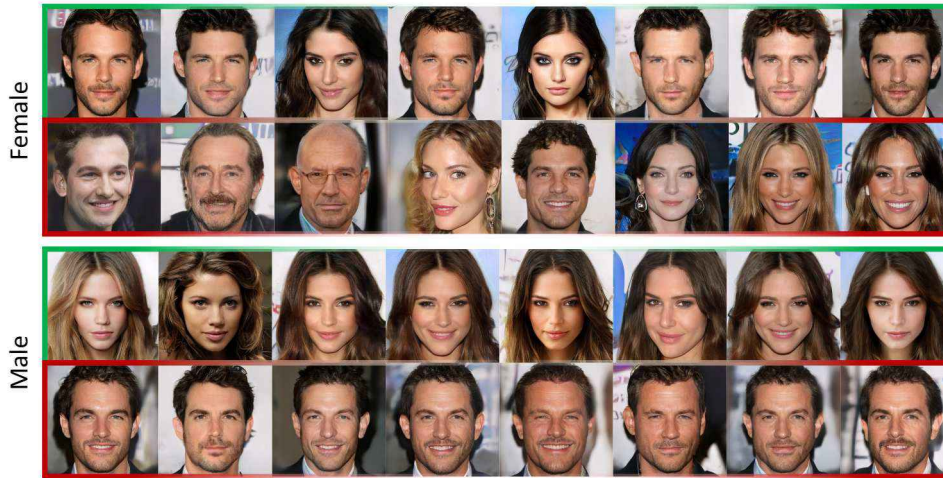
B: Evaluation results
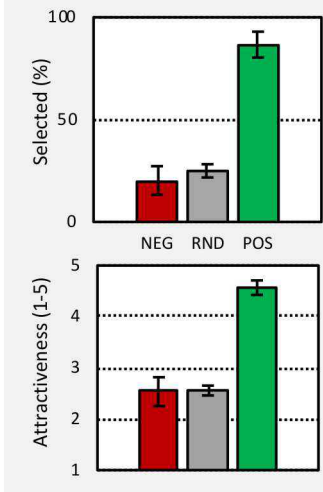

Fig. 5: Individually generated faces and their evaluation. Panel A shows for eight female and eight male participants (full overview available here) the individual faces expected to be evaluated positively (in green framing) and negatively (in red). Panel B shows the evaluation results averaged across participants for both the free selection (upper-right) and explicit evaluation (lower-right) tasks. In the free selection task, the images that were expected to be found attractive (POS) and unattractive (NEG) were randomly inserted with 20 matched controls (RND = random expected attractiveness), and participants made a free selection of attractive faces. In the explicit evaluation task, participants rated each generated (POS, NEG, RND) image on a Likert-type scale of personal attractiveness.

performance (Accuracy = 83.33%) could be described as high.

Figure 5 panel B furthermore shows the results of the second, *explicit evaluation task*. To analyse these results we conducted two Bonferroni-corrected repeated measures ANOVAs with *image* (unattractive, random, attractive) as factor and average rating on personal attractiveness (measure 1) and population attractiveness (measure 2) as dependents. This showed a significant effect of *image* on personal attractiveness, $F(2, 58) = 40.83$, MSE = 1.00, $p < .0001$, $\eta^2 = .46$. As shown in 5, positive images were rated higher than either negative ($p < .0001$) or random ($p < .0001$) images, while negative and random images did not show a difference. For the second measure, of population attractiveness, the same analysis again showed a significant effect of *image*, $F(2, 58) = 43.88$, MSE = 0.45, $p < .0001$, $\eta^2 = .52$. In contrast with personal attractiveness, post-hoc comparisons of population attractiveness showed both positive ($p < .0001$) and negative ($p < .0001$) to be significantly higher evaluated than random images. Indeed, Bonferroni-corrected t-tests showed no significant difference between positive and negative images on popular attractiveness, $p = .07$. Thus, negative generated images were evaluated as highly attractive for other people, but not for the participant themselves.

Taken together, the results suggest that the GBCI was highly accurate in generating personally attractive images (83.33%). They also show that while both negative and positive generated images were evaluated as highly attractive for the general population (respectively M = 4.43 and 4.90 on a scale of 1-5), only the positive generated images (M = 4.57) were evaluated as highly *personally* attractive.

5.4 Qualitative results

In semi-structured post-test interviews, participants were shown the generated images that were expected to be found attractive/ unattractive. Thematic analysis found predictions of positive attractiveness were experienced as *accurate*: There were no false positives (generated unattractive found personally attractive). The participants also expressed being pleased with results (eg."Quite an ideal beauty for a male!"; "I would be really attracted to this!"; "Can I have a copy of this? It looks just like my girlfriend!").

However, there was some ambiguity regarding the accuracy of images that were expected to be seen as unattractive. This may have to do with inviting the social discourse of the outside world into the lab. Despite being well aware that the faces were not real, the subjects would not be *rude to their face*. They would use conversational mitigation strategies, roughly grouped into three categories:

First, when a participant would discuss an image that was expected to be found unattractive, they volunteered that others might: the "it's not you, it's me!" strategy ("I don't find this person attractive... but I see people would think so"). Second, they would couple the (fictional) person in the image with negative personality traits ("I don't like his smile...too bossy"). This could be seen as "shifting the blame" away from themselves: The reason for not liking the image is that the image is unlikable. Only as a last option, might they deny the image personhood entirely, blaming "weird artifacts" or "bad source material" for finding an image unattractive. Otherwise, they would assign the faces personalities, even getting upset if they were predicted to be found unattractive ("He's quite good, if I were him I would be a trifle annoyed!"). This may have the effect of confusing just how negative a judgment they would admit to.

Various themes were discerned in the phenomenological deconstruction of the preferences of the participants. They took into account not only a variety of *visual, physical features*, but also speculated, without being invited to do so, on what these features signified. Some features were very concrete, such as hair color, with a general preference towards blonder generated images. Others were slightly more complex: Male participants uniformly expressed a preference for younger faces while female participants tied perceived youth to specific attributes of the faces, such as the presence or lack of hair ("he looks old...bald"). Participants, however, moved beyond simple *physical features* to inferring *personality traits* ("he looks too bossy") and judging these ("I cannot say anything against him. He is charming"). Here again, they had a tendency to apply real-world social discourse when disapproving. For example, one female expressed dislike for a generated image due to assumption she was required to like it, despite it not being presented as her "perfect match": ("he is too Hollywood. I feel pressured to like him").

In sum, the qualitative results confirm quantitative tests in that the predictions were experienced as accurately matching personal preferences. Moreover, despite the tendency to play down negative attraction, participants agreed on what did not match their personal preferences. The qualitative analysis further enriched our understanding of how participants determined attractiveness, and suggests GBCI could present an effective tool for discussing preference.

6 DISCUSSION AND CONCLUSION

We presented generative brain-computer interfacing for visualizing personal attraction. Our approach used brain responses related to aesthetic preference as feedback for a generative adversarial network (GAN) that generated novel images matching personal attraction. To empirically test this, we recorded EEG responses to GAN produced faces, training a model to classify ERPs. This model was applied to a new sample of artificially produced faces, detecting latent vectors expected to have personally attractive features. Combining these vectors, we generated novel images that were expected to be evaluated as personally attractive in a blind test against matched controls. The results show that GBCI produces highly attractive, personalized images with high (83.33%) accuracy.

6.1 Summary of contributions

The generative brain-computer interface (GBCI) is able to generate *a priori* non-existing images of faces that are seen as personally attractive. Uniting BCI methods with a GAN allowed us to generate photorealistic images based on brain activity. Importantly, the generated images did not rely on external assumptions of the underlying data (such as what attributes make a face beautiful). Thus, the GBCI is able to generate attractive images in a data-driven way unaffected by current theories and opinions of beauty. The contributions of our work to affective computing are both methodological and empirical and are summarized as follows:

The GBCI shows personalized, affective decisions related to visual aesthetics can be revealed in interaction between a human user and a generative neural network. Intentions are latent mental functions that can be hard to verbalize or even unconscious. Aesthetic judgments are a prime example of a common cognitive function we engage in without full understanding: We make snap judgments that something is beautiful, but are ill equipped to explain why. The GBCI operates on *implicit signals* in response to complex configurations of visual features to visualizing the aesthetic decision-making.

The GBCI's generates images by operating on *individual* processes of subjective features, resulting in higher performance than computer vision approaches based on general features. Unlike existing systems that are able to predict and generate generally attractive-looking images, the neuroadaptive BCI allows personalized predictions. While images generated by the GAN model are generically attractive, GBCI was found to produce images that specifically match individual user preferences.

6.2 Limitations

While the GBCI shows clear capability of generating personally attractive images, we would not go so far as to suggest that the generated images correspond to a mental representation of the participant's ideal attraction. This would naively assume that our participants engaged in the task with a "Platonic ideal" image in mind that was also localizable within the GAN space, which they then repeatedly matched against the displayed stimuli so that the GBCI would converge upon this point. Even if such a single image existed either in the mind – a standpoint few in the literature would endorse – or in the GAN, the final generated images in the present study were simply weighted averages of the 240 vectors representing the shown images within a high-dimensional artificial neural network. Such a sample is unlikely to provide representative coverage of the GAN, let alone human face perception.

The approach may thus be limited by the structure of mental representations of personal attraction, as well as the GAN characteristics, such as coverage within this space by the selected initial images and the algorithms used to navigate this space. Further work could present more images, provide an interactive path based on ongoing neural feedback to correct and incorrect predictions, and develop more sophisticated methods to iteratively identify and explore dimensions of the GAN space so that attractive images can be more reliably generated given a user's unique preferences [76].

On the other hand, the constrained sample space within the GAN provides commensurability, enabling inference across subjects [17] and allowing us to compare and contrast them in terms of personal attraction [16]. For example, as can be seen in Figure 5, the male positive matches show generated images with a marked family resemblance, suggesting high consensus in their personal attraction. This presents an independent confirmation of a well-known finding concerning attractiveness ratings in that males are generally more in agreement than females as to what they find attractive [74].

As the GBCI output is determined by the GAN architecture, it is critical to consider how the particular train-

ing set initially used to create the network, i.e. images of thousands of celebrities [38], affected the results. Biased data is a common problem throughout the machine learning community, and even datasets widely used as benchmarks for image classification tasks of everyday objects have been shown to be biased in some manner [69]. Clearly, given that images of famous people include people like fashion models and movie stars, it is reasonable to assume that the GBCI's neuroadaptivity was biased towards producing attractive-looking images. The results clearly indicate that this was indeed the case: negative predictions were evaluated and discussed as conventionally, even if not personally attractive. Indeed, in blind tests, only positive predictions were selected as matching the user's individual aesthetics. In other words, the use of a GAN trained on generically attractive people increased the difficulty of demonstrating validity, and yet, positive generated images were highly preferred over matched controls. Thus, while a regular GAN and computer vision based attraction detection algorithm provides ample means for generating attractive images, the GBCI performs better for personal attraction.

A similar limitation from the use of a celebrity images GAN is the degree this over-emphasizes inequality and lack of diversity in the popular media. For example, considering the degree to which people of color are generally underrepresented in celebrity populations [21], [33], they are also less likely to be generated by the GBCI, as can be seen in the lack of ethnic diversity presented in Figure 5. On the other hand, this may well correspond to the degree representation influences social perception, given that our mainly white sample demonstrated racial preferences favoring own-race images over other-race images, replicating existing findings from the social-psychological literature [8], [26], [59]. It is possible that the GBCI's high performance metrics are due to the social inequality within the GAN model mirroring cultural psychological biases. Thus, much like other machine learning applications, GBCI does not represent an "objective", "fair" technology.

Another aspect related to the representation learned by the GAN model is the face domain. Our experiments are based on a latent structure learned in an unsupervised manner, but are limited to a dataset of images of celebrity faces. Therefore, the broader generalizability across different domains, such as landscapes, animals, or even art, depends on the ability of GAN architecture to capture the semantics [4] and different affective dimensions that the learned representations capture. Recent research has shown that GAN models are, indeed, able to perform on a variety of domains [39] and varying tasks, such as style transfer [1] and even capture object-level semantics [10].

6.3 Implications and future work

As a novel paradigm, GBCI produces insights into human cognitive and affective processing by visualizing individual differences and portraying implicit processing. In the previous section, we already discussed how the uniform procedure used to generate the individual images presents interesting insights: The generated positive and negative images of the male group of participants seems more similar to one another than the female ones. While cognitive neuroscience typically looks at the similarity between individuals

in terms of their averages, the reverse inference achieved by GBCI allows us to speculate about interindividual variance: What causes groups of people to relate in terms of GBCI generated faces, or to differ drastically between one another?

Furthermore, we are interested in exploring the degree to which GBCI visualizes implicit bias. Noting the lack of people of color in Figure 5 despite ethnicity being task irrelevant, we speculated in the previous section how this could either reflect a GAN limitation or an implicit bias in participants. If the latter is true, then we would expect even with more diversely oriented GANs, GBCI would produce similar results. Furthermore, GBCI could provide qualitatively different insights with other subjective categorization tasks, such as recognizing images as trustworthy, benevolent, or powerful. What kind of person would the GBCI generate, but more importantly, what would this tell us about the individual? While this remains speculation, the results presented in the present study cause us to believe the GBCI could be a significant step forward in social cognition.

Given the far-reaching importance the GBCI could have if it were to present a tool for visualizing implicit bias, it is critical for future work to address the degree to which the results are affected by the GAN's original source material and/or by the diversity of the GBCI participants. To this end, we envision two separate paths. Firstly, it is critical to investigate whether GANs based on more diverse source material may better represent diverse participant's personal attraction. Secondly, cross-cultural endeavors should be made to estimate the efficacy of GBCI with groups that are underrepresented within the GAN. For both, it may seem obvious that disparity between GAN and sample diversity limits the efficacy of the GAN. However, it is important to keep in mind that the overrepresentation of certain demographies in the popular media may itself affect mental representations, thus causing implicit bias. As we see, the present results suggest high performance of the GBCI across genders and even though the sample pool hardly reflected the original celebrity database.

Finally, while future work must identify to what extent GBCI output corresponds to mental processes, we envision the GBCI as a practical, creative tool that generates images optimized towards personal aesthetic preferences. By mapping brain activity to a representational space learned from training data without predefined model parameterization, we demonstrate the feasibility of visualizing attractiveness without bias towards a-priori assumptions of the internal structure underlying attractive images. Since the representational space is infinite, the model is theoretically capable of reproducing mental images within the limits of the data used to train the generative model [18]. The GBCI furthermore operates on brain signals alone and therefore requires no artistic skills to create aesthetically pleasing portraits. Thus, our approach may enable users to realize creative customization of images using brain signals. Even though future studies will be needed to determine the relationship between GBCI-produced images, mental imagery, and mental representation, we can already confirm that the GBCI is capable in creating novel images that are evaluated as personally attractive.

ACKNOWLEDGMENTS

REFERENCES

[1] Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2stylegan: How to embed images into the stylegan latent space? In *Proceedings of the IEEE international conference on computer vision*, pages 4432–4441, 2019.

[2] Oswald Barral, Manuel JA Eugster, Tuukka Ruotsalo, Michiel M Spapé, Ilkka Kosunen, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. Exploring peripheral physiology as a predictor of perceived relevance in information retrieval. In *Proceedings of the 20th international conference on intelligent user interfaces*, pages 389–399, 2015.

[3] Oswald Barral, Ilkka Kosunen, Tuukka Ruotsalo, Michiel M Spapé, Manuel JA Eugster, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. Extracting relevance and affect information from physiological text annotation. *User Modeling and User-Adapted Interaction*, 26(5):493–520, 2016.

[4] David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B Tenenbaum, William T Freeman, and Antonio Torralba. Gan dissection: Visualizing and understanding generative adversarial networks. *arXiv preprint arXiv:1811.10597*, 2018.

[5] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus-Robert Müller. Single-trial analysis and classification of erp components - a tutorial. *NeuroImage*, 56(2):814–825, 2011.

[6] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[7] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. *CoRR*, abs/1809.11096, 2018.

[8] David M Buss. Human mate selection: Opposites are sometimes said to attract, but in fact we are likely to marry someone who is similar to us in almost every variable. *American scientist*, 73(1):47–51, 1985.

[9] Danilo Bzdok, Robert Langner, Svenja Caspers, F Kurth, U Habel, Karl Zilles, A Laird, and Simon B Eickhoff. Ale meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Structure and Function*, 215(3-4):209–223, 2011.

[10] Xinyuan Chen, Chang Xu, Xiaokang Yang, and Dacheng Tao. Attention-gan for object transfiguration in wild images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 164–180, 2018.

[11] Matthew A Conroy and John Polich. Affective valence and p300 when stimulus arousal level is controlled. *Cognition and emotion*, 21(4):891–901, 2007.

[12] Eric Courchesne, Steven A Hillyard, and Robert Galambos. Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and clinical neurophysiology*, 39(2):131–143, 1975.

[13] Michael R Cunningham, Alan R Roberts, Anita P Barbee, Perri B Druen, and Cheng-Huan Wu. " their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, 68(2):261, 1995.

[14] Charles Darwin. *The Descent of Man, and Selection in Relation to Sex*, volume 2. D. Appleton, 1872.

[15] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Studying aesthetics in photographic images using a computational approach. In *European conference on computer vision*, pages 288–301. Springer, 2006.

[16] Keith M Davis, Michiel Spapé, and Tuukka Ruotsalo. Collaborative filtering with preferences inferred from brain signals. In *WWW '21: Proceedings of The Web Conference 2021*, 2021. To appear.

[17] Keith M Davis III, Lauri Kangassalo, Michiel Spapé, and Tuukka Ruotsalo. Brainsourcing: Crowdsourcing recognition tasks via collaborative brain-computer interfacing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.

[18] Carlos de la Torre-Ortiz, Michiel M Spapé, Lauri Kangassalo, and Tuukka Ruotsalo. Brain relevance feedback for interactive image generation. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pages 1060–1070, 2020.

[19] Emanuel Donchin and Michael GH Coles. Is the p300 component a manifestation of context updating? *Behavioral and brain sciences*, 11(3):357–374, 1988.

[20] Yael Eisenthal, Gideon Dror, and Eytan Ruppin. Facial attractiveness: Beauty and the machine. *Neural Computation*, 18(1):119–142, 2006.

[21] Morgan E Ellithorpe and Amy Bleakley. Wanting to see people like me? racial and gender diversity in popular adolescent television. *Journal of youth and adolescence*, 45(7):1426–1437, 2016.

[22] Manuel J. A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. Natural brain-computer interfaces: Recommending information by relevance inferred from human brain signals. *Scientific Reports*, 6:38580, December 2016.

[23] Manuel J.A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. Predicting term-relevance from brain signals. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '14, page 425–434, New York, NY, USA, 2014. Association for Computing Machinery.

[24] Lawrence A Farwell and Emanuel Donchin. The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology*, 28(5):531–547, 1991.

[25] Lawrence Ashley Farwell and Emanuel Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, 70(6):510–523, 1988.

[26] Raymond Fisman, Sheena S Iyengar, Emir Kamenica, and Itamar Simonson. Racial preferences in dating. *The Review of Economic Studies*, 75(1):117–132, 2008.

[27] Jerome H. Friedman. Regularized Discriminant Analysis. *Journal of the American Statistical Association*, 84(405):165–175, March 1989.

[28] Junying Gan, Lichen Li, Yikui Zhai, and Yinhua Liu. Deep self-taught learning for facial beauty prediction. *Neurocomputing*, 144:295 – 303, 2014.

[29] P.I. Good. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer, 2nd edition, 2000.

[30] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[31] Ville J Harjunen, Imtiaj Ahmed, Giulio Jacucci, Niklas Ravaja, and Michiel M Spapé. Manipulating bodily presence affects cross-modal spatial attention: A virtual-reality-based erp study. *Frontiers in human neuroscience*, 11:79, 2017.

[32] Lea Höfel and Thomas Jacobsen. Electrophysiological indices of processing aesthetics: Spontaneous or intentional processes? *International Journal of Psychophysiology*, 65(1):20–31, 2007.

[33] Darnell Hunt, Ana-Christina Ramón, Michael Tran, Amberia Sargent, and Debanjan Roychoudhury. Hollywood diversity report 2018: Five years of progress and missed opportunities. *UCLA College of Social Sciences, February. https://socialsciences. ucla. edu/wp-content/uploads/2018/02/UCLA-Hollywood-Diversity-Report-2018-2-27-18. pdf*, 2018.

[34] Miguel Ibáñez-Berganza, Ambra Amico, and Vittorio Loreto. Subjectivity and complexity of facial attractiveness. *Scientific reports*, 9(1):1–12, 2019.

[35] Giulio Jacucci, Oswald Barral, Pedram Daee, Markus Wenzel, Baris Serim, Tuukka Ruotsalo, Patrik Pluchino, Jonathan Freeman, Luciano Gamberini, Samuel Kaski, et al. Integrating neurophysiologic relevance feedback in intent modeling for information retrieval. *Journal of the Association for Information Science and Technology*, 70(9):917–930, 2019.

[36] Victor S Johnston, Rebecca Hagel, Melissa Franklin, Bernhard Fink, and Karl Grammer. Male facial attractiveness: Evidence for hormone-mediated adaptive design. *Evolution and human behavior*, 22(4):251–267, 2001.

[37] Lauri Kangassalo, Michiel Spapé, and Tuukka Ruotsalo. Neuroadaptive modelling for generating images matching perceptual categories. *Scientific Reports*, 10(1):1–10, 2020.

[38] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *arXiv:1710.10196 [cs, stat]*, October 2017. arXiv: 1710.10196.

[39] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image

quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020.

[40] Yan Ke, Xiaoou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 419–426. IEEE, 2006.

[41] Nicole Koehler, Leigh W Simmons, Gillian Rhodes, and Marianne Peters. The relationship between sexual dimorphism in human faces and fluctuating asymmetry. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(suppl_4):S233–S236, 2004.

[42] Albert Kok. On the utility of p3 amplitude as a measure of processing capacity. *Psychophysiology*, 38(3):557–577, 2001.

[43] Laurens R Krol and Thorsten O Zander. Passive bci-based neuroadaptive systems. In *GBCIC*, 2017.

[44] Sandra JE Langeslag, Bernadette M Jansma, Ingmar HA Franken, and Jan W Van Strien. Event-related potential responses to love-related facial stimuli. *Biological psychology*, 76(1-2):109–115, 2007.

[45] Judith H Langlois, Jean M Ritter, Lori A Roggman, and Lesley S Vaughn. Facial diversity and infant preferences for attractive faces. *Developmental Psychology*, 27(1):79, 1991.

[46] Judith H Langlois and Lori A Roggman. Attractive faces are only average. *Psychological science*, 1(2):115–121, 1990.

[47] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photorealistic single image super-resolution using a generative adversarial network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[48] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, February 2004.

[49] Anthony C Little, Benedict C Jones, and Lisa M DeBruine. Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571):1638–1659, 2011.

[50] Anthony C Little, Benedict C Jones, Lisa M DeBruine, and Christine A Caldwell. Social learning and human mate preferences: a potential mechanism for generating and maintaining between-population diversity in attraction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1563):366–375, 2011.

[51] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep Learning Face Attributes in the Wild. *arXiv:1411.7766 [cs]*, November 2014. arXiv: 1411.7766.

[52] Yiwen Luo and Xiaoou Tang. Photo and video quality evaluation: Focusing on the subject. In *European Conference on Computer Vision*, pages 386–399. Springer, 2008.

[53] Ewout H Meijer, Gershon Ben-Shakhar, Bruno Verschuere, and Emanuel Donchin. A comment on farwell (2012): brain fingerprinting: a comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cognitive neurodynamics*, 7(2):155–158, 2013.

[54] Markus Ojala, Niko Vuokko, Aleksi Kallio, Niina Haiminen, and Heikki Mannila. Randomization methods for assessing data analysis results on real-valued matrices. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 2(4):209–230, 2009.

[55] Letizia Palumbo, Marco Bertamini, and Alexis Makin. Scaling of the extrastriate neural response to symmetry. *Vision Research*, 117:1–8, 2015.

[56] David Perrett, KA May, and Sakiko Yoshikawa. Facial shape and judgements of female attractiveness. *Nature*, 368:239–42, 04 1994.

[57] John Polich. Attention, probability, and task demands as determinants of p300 latency from auditory stimuli. *Electroencephalography and clinical neurophysiology*, 63(3):251–259, 1986.

[58] John Polich. Updating p300: an integrative theory of p3a and p3b. *Clinical neurophysiology*, 118(10):2128–2148, 2007.

[59] Belinda Robnett and Cynthia Feliciano. Patterns of racial-ethnic exclusion by internet daters. *Social Forces*, 89(3):807–828, 2011.

[60] J. J. Rocchio. Relevance feedback in information retrieval. In Gerard Salton, editor, *The SMART Retrieval System - Experiments in Automatic Document Processing*, pages 313–323. Prentice Hall, Englewood, Cliffs, New Jersey, 1971.

[61] Christopher P Said and Alexander Todorov. A statistical model of facial attractiveness. *Psychological Science*, 22(9):1183–1190, 2011.

[62] Jose San Pedro and Stefan Siersdorfer. Ranking and classifying attractiveness of photos in folksonomies. In *Proceedings of the 18th international conference on World wide web*, pages 771–780. ACM, 2009.

[63] Kendra Schmid, David Marx, and Ashok Samal. Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios. *Pattern Recognition*, 41(8):2710 – 2717, 2008.

[64] Hamid R Sheikh, Alan C Bovik, and Lawrence Cormack. No-reference quality assessment using natural scene statistics: Jpeg2000. *IEEE Transactions on Image Processing*, 14(11):1918–1927, 2005.

[65] Jiazheng Shi, Ashok Samal, and David Marx. How effective are landmarks and their geometry for face recognition? *Computer vision and image understanding*, 102(2):117–133, 2006.

[66] Michiel M Spapé, Marco Filetti, Manuel JA Eugster, Giulio Jacucci, and Niklas Ravaja. Human computer interaction meets psychophysiology: a critical perspective. In *International Workshop on Symbiotic Interaction*, pages 145–158. Springer, 2015.

[67] Michiel Spapé, Rinus Verdonschot, and Henk van Steenbergen. *The E-Primer: An introduction to creating psychological experiments in E-Prime.* Leiden Univerisity Press, 2019.

[68] Randy Thornhill and Steven W Gangestad. Facial attractiveness. *Trends in cognitive sciences*, 3(12):452–460, 1999.

[69] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *CVPR 2011*, pages 1521–1528. IEEE, 2011.

[70] Christopher W Tyler. Empirical aspects of symmetry perception. *Spatial Vision*, 9(1):1–8, 1995.

[71] Antti Ukkonen, Pyry Joona, and Tuukka Ruotsalo. *Generating Images Instead of Retrieving Them: Relevance Feedback on Generative Adversarial Networks*, page 1329–1338. Association for Computing Machinery, New York, NY, USA, 2020.

[72] Bram van de Laar, Hayrettin Gürkök, Danny Plass-Oude Bos, Mannes Poel, and Anton Nijholt. Experiencing bci control in a popular computer game. *IEEE Transactions on Computational Intelligence and AI in Games*, 5(2):176–184, 2013.

[73] Rolf Verleger. On the utility of p3 latency as an index of mental chronometry. *Psychophysiology*, 34(2):131–156, 1997.

[74] Dustin Wood and Claudia Chloe Brumbaugh. Using revealed mate preferences to evaluate market force and differential preference explanations for mate selection. *Journal of personality and social psychology*, 96(6):1226, 2009.

[75] Beste F. Yuksel, Kurt B. Oleson, Lane Harrison, Evan M. Peck, Daniel Afergan, Remco Chang, and Robert JK Jacob. Learn piano with bach: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5372–5384, New York, NY, USA, 2016. ACM.

[76] Thorsten O Zander, Lauens R Krol, and Klaus Gramann. Towards neuroadaptive technology: Implicitly controlling a cursor though a passive brain–computer interface. In *Neuroergonomics*, pages 301–302. Elsevier, 2018.

Michiel Spapé received his PhD in Psychology from Leiden University in 2009. After working as a postdoc (Nottingham, Helsinki), and lecturer (Liverpool), he is now docent in cognitive neuroscience at Helsinki University, focusing on emotion, perception/action, and EEG.

Lauri Kangassalo received his MSc in Computer Science from the University of Helsinki in 2019. As a specialist in machine learning, he aided the present study in psychophysiology analysis. He currently works at the Finnish Meteorological Institute.

Keith M. Davis III is a PhD student at the University of Helsinki. He specializes in machine learning and collaborative brain-computer interfaces.

Niklas Ravaja received a PhD degree in psychology from the University of Helsinki, Finland, in 1996. He is professor of eHealth and well-being, and expert on emotional and physiological processes during mediated social interaction.

Zania Sovijärvi-Spapé studied at the Universities of Amsterdam and Helsinki. She is a lab assistant at the Interaction Lab and the Cognitive Computing Lab and provided expertise in qualitative data analysis.

Tuukka Ruotsalo is an Academy Research Fellow at the University of Helsinki, and an Associate Professor at the University of Copenhagen, He is an expert on machine-learning and cognitive computing.