Beautification of Images by Generative Adversarial Networks (GANs)

Amar Music
Master of Psychology: Theory and Research

KU LEUVEN
Faculty of Psychology and Educational Sciences

Supervisor: Prof. Johan Wagemens
Co-supervisor: Anne-Sofie Maerten
July 27, 2022

Contents

Abstract

...

Beautification of Images by Generative Adversarial Networks (GANs)

For millennia, philosophers have been arguing about the nature of beauty. Yet after all these years, we still can't confidently claim to understand what beauty even *is*. Furthermore, the fundamental question whether beauty is subjective or objective remains unsettled (Sartwell, 2022). In this paper, we will use state-of-the-art machine learning techniques to try and create a so-called 'visual definition' of beauty.

Defining Beauty

Throughout history, there have been different conceptions of beauty. The classical conception of beauty, often embodied in classical art, typically concerns the arrangement of independent parts into a coherent whole through properties such as symmetry and order (Sartwell, 2022; Wölfflin, 1932). Another important conception is the idealist conception of beauty described by Plato as a perfect unity. This is in contrast to the classical conception which emphasizes individual parts (Sartwell, 2022). 18th century empiricist philosophers on the other hand looked at beauty in a more instrumental, hedonistic sense. Hume (1740) for example argues that beauty exists to give pleasure and satisfaction to the soul.

While most classical philosophers would have argued that beauty is a property of an object that exists outside the mind, Renaissance and later philosophers argue more for a combination of some objective properties with unique subjective appreciations (Sartwell, 2017). The assumption that there are indeed some objective properties to beauty allows us to examine what exactly it means for something to be beautiful.

How we Perceive Beauty

The shift from a pure objective conception of beauty towards a more nuanced combination of objective properties with subjective perception preceded the start of psychology as an empirical science in the 19th century. Among these early psychologists a school of thought known as Gestalt psychology came to fruition in the early 20th century. The Gestalt psychologists created a framework of visual perception that emphasizes patterns and configurations of objects rather than individual components (Wagemans et al., 2012). Examples of perceptual grouping principles are proximity, similarity, continuity, closure, and connectedness (Goldstein, 2009). In the Gestalt framework, principles such as these are the building blocks for our visual system (Wagemans et al., 2012).

Later psychological research focusing on art and aesthetic appreciation has found that there are certain principles that tend to evoke a certain aesthetic quality across cultures such as for example symmetry (Bode et al., 2017), order and complexity (Van Geert & Wagemans, 2021), and figure-ground contrast (Reber et al., 2004). Mirroring the progression in philosophy, modern views on aesthetics from neuroscience and experimental psychology support an interactionist interpretation of beauty: objects do contain intrinsic properties conveying beauty, but the final aesthetic appraisal comes from an interaction between these objective properties and a subjective observer (Valenzise et al., 2022).

One influential framework that incorporates the interactionist interpretation of beauty is that of processing fluency formulated by Reber et al. (2004). In this framework, the aesthetic experience is determined by the fluency with which a perceived object is processed. Processing fluency relies on an interaction of objective properties such as the principles from Gestalt psychology with subjective factors such as personal experiences and expectations.

Machine Learning and Beauty

One of the biggest technological advances in recent years is in the field of machine learning. This 'Deep Learning Revolution' was sparked by the continuous efforts of researchers and engineers, the increasing availability of big data, and the boost in computing power (Sejnowski, 2018). Because these networks have some organizational similarities to the human brain, they can potentially be used as a means to study obscure mental representations in humans (Goetschalckx et al., 2021; Guo et al., 2016).

Computational aesthetics is defined as the research of human aesthetic decision making through computational methods (Hoenig, 2005; Valenzise et al., 2022). As a result of the deep learning revolution, more and more types of artificial neural networks have become available to be used as a means to study human aesthetic behavior.

Generative Adversarial Networks (GANs) are a relatively recent development characterized by two competing neural networks (Goodfellow et al., 2014). The first network, the generator, starts generating random outputs and the second network, the discriminator, receives inputs from the generator and compares these to a real dataset (Creswell et al., 2018). In the context of image generation, the inputs and outputs are images. Importantly, the discriminator has been trained on a real dataset and will judge the output generated by the generator. The generator receives the feedback and over many iterations will slowly learn to create better outputs that start to resemble the real dataset the discriminator had been trained on. Both the generator and the discriminator improve over time and as a result, the generator will be able to generate more convincing outputs and the discriminator will be able to differentiate much better between real and fake.

Goetschalckx et al. (2019) have developed GANalyze, a framework using GANs that allows us to study cognitive properties by means of so-called visual definitions. In their paper, Goetschalckx et al. (2019) utilized their network to generate images of varying levels of memorability which were validated in a behavioral task. GANalyze makes this possible by using as the discriminator a convolutional neural network (CNN) designed for predicting image memorability (Khosla et al., 2015).

Present Study

The goal of this study is to investigate the factors underlying the aesthetic value of images using GANalyze. We will use GANalyze to generate sequences of images that we hypothesize possess either *more*, or *less* aesthetic quality depending on the internal variable used for their respective generations. By comparing the low-aesthetic images to the high-aesthetic images, we will be able to examine the underlying factors responsible for making an image either more or less beautiful. We will also perform a finer analysis of the aesthetics-related parameters by looking into the hidden layers of the trained GANalyze network.

However, if we wish to learn something about human aesthetics appraisal from a GAN, we must first validate whether the GAN truly 'understands' what it means for an image to possess aesthetic quality. To do so, we will validate whether human participants tend to agree with GANalyze on which image appears to be more beautiful. We are also interested whether a person's experience with art and their demographic characteristics affect their proportion of agreement with the GAN. A relation between art experience and agreement with the GAN could have important implications for the validation of the network and the generalizability of the final results.

Methods

Stimulus Generation

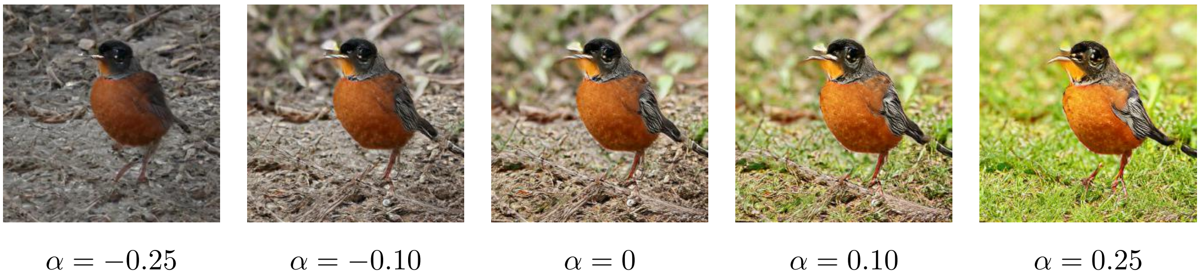
We used GANalyze with AestheticsNet (Kong et al., 2016) as the discriminator and BigGAN-256 (Brock et al., 2019) pretrained on ImageNet (Russakovsky et al., 2015) as the generator to train the GANalyze model for 400,000 iterations. To do this, we used `Python 3.6` (van Rossum & Team, 2016) with the `TensorFlow` libraries (Abadi et al., 2016). The training resulted in GANalyze being able to produce image sequences based on 1,000 ImageNet categories. For each ImageNet category, we initially generated three different seeds (e.g., three different images of Siamese cats) to test for an effect of idiosyncratic effects of images on top semantic category. Within each seed, GANalyze produced 21 images with a supposedly increasing amount of aesthetic value (Figure 1). We picked a broad range of α-values with increasing density close to zero to obtain stimuli with obvious changes (e.g., $\alpha = 0.25$), but also subtle changes (e.g., $\alpha = 0.0025$). We chose these values because we hypothesized there to be a ceiling effect where there would be no behavioral differences for extreme α-values. We also hypothesized that for very low α-values, the perceptual difference between stimuli would be too small to detect. The specific α-values used in the final experiment were chosen based on the results of a pilot study with a very broad range. In addition, the pilot showed us that we would be justified to only use one of the three seeds for each category.

511 ImageNet categories were manually excluded from the stimulus set because they either contained human faces (which are deliberately warped by BigGAN), insects such as spiders which might make some participants feel uncomfortable, or for redundancy such as the disproportionately high number of categories about dog breeds. In addition,

stimuli that were unrecognizable were also excluded. This resulted in 489 categories, each with one seed and 15 α-values, equaling a total of 7,335 stimuli.

Figure 1

Truncated Sample of an Image Sequence Produced by GANalyze



Questionnaire

To measure the aesthetic experience of each participant, we used a shortened version of the Aesthetic Experience Questionnaire (AEQ; Wanzer et al., 2020). We have chosen this particular questionnaire based on a number of factors. First of all, the theoretical background of an aesthetic flow experience (Csikszentmihalyi & Robinson, 1990) is valid under our assumed interactionist conception of beauty. Second, this questionnaire seems fitting because the AEQ is designed to be generalizable to all visual aesthetic experiences and is not limited to famous pieces in a museum like many other questionnaires. Finally, we picked the AEQ because the authors suggest that it should be possible to reduce the items on the six scales and still end up with a sufficient measurement of aesthetic experience.

As the experience one has with art was not the primary focus of our study, we only used the highest loading items from each of the six scales, resulting in a total of six items at the start of our study in order to save time. Each item was presented with a five-point Likert scale, ranging from 0 (*Strongly disagree*) to 4 (*Strongly agree*). To test whether this shortened version of the AEQ was sufficient to explain the latent aesthetic experience, we used confirmatory factor analysis (CFA).

Behavioral Experiment

To determine whether GANalyze can be used to study what it means for an image to be aesthetic, we conducted a behavioral study to test whether human participants indeed preferred the images that GANalyze generated to be more aesthetic over a corresponding base image.

The study was hosted on Prolific.co to obtain a large number of data points in a simple online task, ensuring each of the 7,335 stimuli was evaluated around 20 times on average. Individuals younger than 17 or those without normal or corrected-to-normal vision were not allowed to participate. The online experiment started an informed consent document which had to be signed, followed the AEQ. After the AEQ, the experiment started. The

Figure 2

Example of a Sequence of Trials in the Study

participants carried out a spatial two-alternatives forced choice (2AFC) task in which they had to indicate which of the two presented GAN-generated images they considered to be most aesthetically pleasing. They did this by pressing "F" for the left, and "J" for the right image (Figure 2).

Each trial consisted of two images generated the same ImageNet category. One of the two images was always the base image with $\alpha = 0$ while the comparison image was, according to GANalyze at least, either more aesthetic ($\alpha > 0$) or less aesthetic ($\alpha < 0$). The α-value for the comparison image differed for each trial, leading to differences in similarity between the base and comparison images for each trial. An α-value close to zero indicates high similarity and, consequently, difficulty to discriminate between the images. An α-value much larger or smaller than zero indicates low similarity and therefore easy discrimination. The position of the base and comparison image was randomized for each trial to prevent possible confounding factors. Each 10-minute block was created to contain a roughly uniform distribution of chosen α-values and ImageNet categories to ensure that each participant is presented with a comparable stimulus set. Furthermore, the same category never appeared more than one time in a block.

After 10 minutes, the task automatically turned to a screen thanking the participants for their assistance, after which they were given monetary compensation of £1.38 converted to their local currency.

Data Analysis

Structural Equation Modeling

Our structural model needs to have a latent variable measuring aesthetic experience which has to be properly explained. In addition, we want to know whether age, sex, and nationality affect this. Furthermore, we are interested in the effects of aesthetic experience, age, sex, and nationality on the outcome variable of the behavioral task measured by percentage of agreeing with the neural network. To test this, we used structural equation modeling (SEM) with the `lavaan` package (Rosseel, 2012) in R version 4.1.0 (R Core Team, 2021) to build and evaluate models based on theory that would give an explanation to the data. All models were fitted using a maximum likelihood estimation method and the fits were evaluated with the comparative fit index (CFI), Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA).

First of all, we performed a priori CFA to validate whether our shortened version of the AEQ was indeed sufficient in explaining the latent aesthetic experience variable. We used the 'emotional', 'cultural', 'perceptual', 'understanding', 'flow-proximal', and 'flow-experience' scales as indicators for our latent variable.

Next, we moved on to more complicated structural models to better describe the full dataset. More specifically, we created a multiple indicators multiple causes (MIMIC) model aimed to show whether age, sex, and nationality affected the latent aesthetic experience variable. Crucially, we were interested in the effects of the aesthetic experience latent variable and age, sex, and nationality on the outcome of the image comparison behavioral task. We used the proportion of agreement with the GAN as the measure for aesthetic judgement here. Our initial idea for this model consisted of the latent variable, explained by the indicators, affecting aesthetic judgement in the behavioral task. Age, sex and nationality would have an effect on both the latent variable and the behavioral outcome of judgement. We also made another alternative model based on different theoretical arguments. For this second model, we decided that it did not make sense for nationality to have a direct effect on aesthetic judgement, so we removed that effect and instead added an effect of nationality on the cultural scale from the AEQ, as we believed this would make sense. These models would additionally test the questions Wanzer et al. (2020) had regarding their homogeneous population sample.

Even though the new model MIMIC did not appear to be significantly better than the older one, we decided to keep the newer model as it made more theoretical sense.

Behavioral Experiment

We analyzed the results from the behavioral task using a variety of methods. First of all, we cleaned the data by removing participants that (1) had a bias for the left/right stimulus $> 25\%$, (2) performed too close to chance (i.e., $< 55\%$ agreement for all α-values), (3) responded too fast to reasonably perceive and judge the stimuli (i.e., median RT < 500 ms), and (4) had an exceptionally low number of trials during the 10-minute experiment (i.e., number of trials < 50).

To see whether the different α-values indeed represent a gradual change in stimulus difference, we fitted the results of all participants with a psychometric function using the `quickpsy` package (Linares & López-Moliner, 2016). Analysis on differences between positive and negative α-values was done using second-order polynomial regression. Finally, we tested for potential confounding factors using simple ANOVA.

GANalyze Parameters

The features the transformer takes into account for image modification are a direct consequence of the discriminator's properties. We visualized feature maps of AestheticsNet's hidden layers to find the features it used to evaluate the generator during the training phase. We did this using

Results

Pilot

We conducted a preliminary pilot version of the experiment to find reasonable variables to use for the final experiment. The pilot consisted of 10 participants who completed a version of the experiment with three seeds for each image and 21 α-levels. With a Generalized Linear Model (GLM), we concluded that there was no effect of seed ($p = .38$). With a visual inspection of the effects of α-value on agreement, we removed α-values that appeared redundant in showing the relation in order to have better measurements for the useful values.

Participants

After removing participants that did not satisfy the criteria listed above, our total number of participants was 542 ($M_{age} = 31.17,\ SD_{age} = 11.00,\ 234$ female).

Structural Models

Table 1: Model Fit Indices

Model	χ^2	*df*	CFI	TLI	RMSEA
CFA	33.92***	9	.96	.93	.07
MIMIC 1	68.72***	29	.94	.91	.05
MIMIC 2	68.00***	29	.94	.91	.05

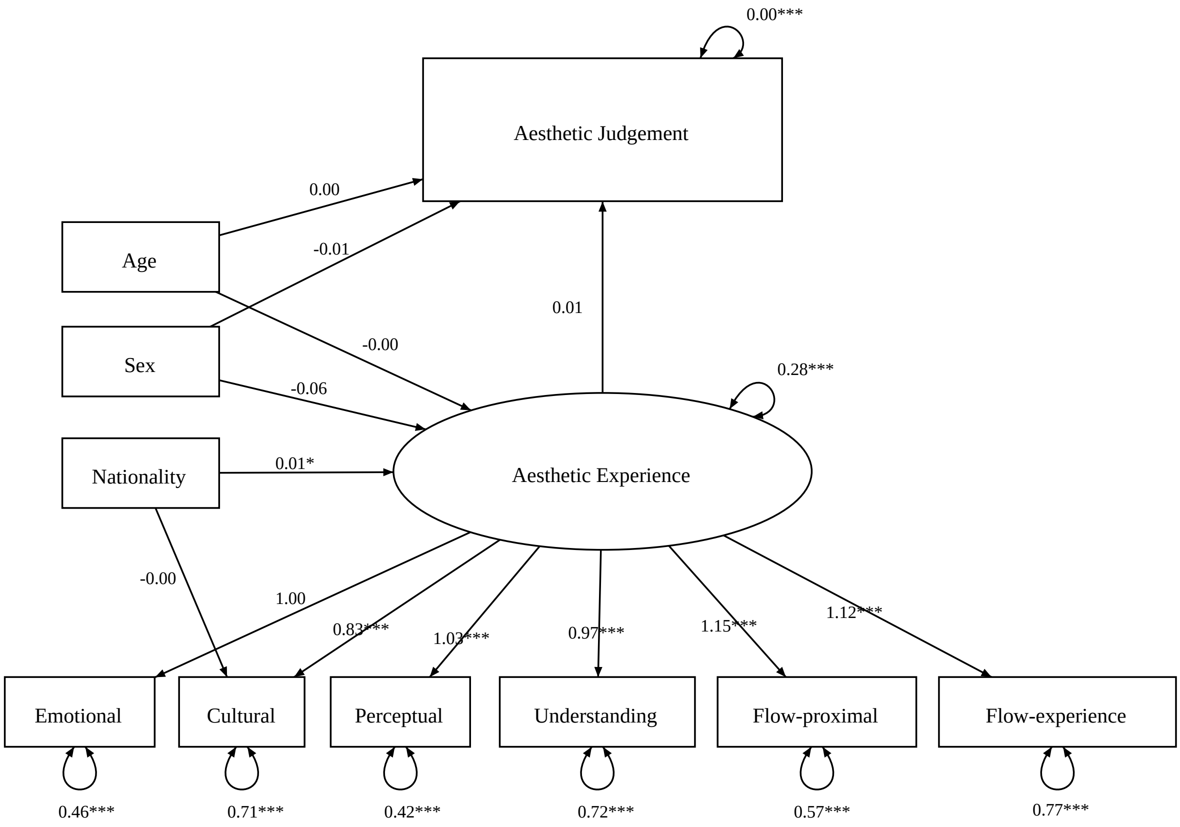
***$p < .001$.

Our first model, the CFA on the shortened AEQ, resulted in a validation of the shortened questionnaire. The model fit can be seen in Table 1 under CFA. All fit indices, apart from the conservative χ^2, were in the acceptable range (i.e., $> .90$ for CFI and TLI, and $< .08$ for RMSEA). All standardized factor loadings from the latent variable on the measured indices were significant ($p < .001$) and strong enough to justify interpretation ($> .80$). We interpret these results as evidence for our shortened version of the AEQ being adequate for measuring an individual's aesthetic experience.

For our first MIMIC model, we also obtained good fit measures, as seen in Table 1 under MIMIC 1. This model included an effect of sex, age, and nationality on aesthetic experience, and an effect of sex, age, and nationality on the proportion of agreement

Figure 3

Final MIMIC Model



with the GAN. While the measurement part of the model remained largely unchanged, the new structural part of the model only showed a significant effect of nationality on aesthetic experience ($\beta = 0.01, p < .05$). This finding may have consequences for the AEQ itself, as Wanzer et al. (2020) had mentioned that their own sample only included US participants. On the other hand, we did not find an effect of age and sex on aesthetic experience, unlike Wanzer et al. (2020). The most important finding here is that there was no effect of the latent aesthetic experience on the behavioral outcome variable of aesthetic judgement, which has important implications for the rest of our study.

For our second MIMIC model, illustrated in Figure 3, we made changes on theoretical bases. The fit indices were equally good, as seen in Table 1 under MIMIC 2, but we believe this second model makes more sense from a theoretical point of view. The standardized factor loadings were again very similar. Again, the effects of age and sex on aesthetic experience and aesthetic judgement were not significant. Nationality still has a significant effect on aesthetic experience here, but not on the cultural indicator. There was also no significant effect of aesthetic experience on aesthetic judgement here.

We can interpret these results as there being no effect of a person's demographic characteristics and their aesthetic experience on the outcome of the behavioral image rating task. This implies that the simple image comparison we used in our experiment is not influenced by seemingly relevant factors, pointing to a potential universal nature of aesthetic appreciation of the images we provided in the task. Furthermore, this will make our interpretations of the factors determining the aesthetic value of an image through GANalyze more generalizable.

Behavioral Task

Psychometric Function

To test the effect of different α-values on image preference, we fitted a psychometric function on the data (Figure 4a). The figure shows that the relation between the α-value used for image generation and the behavioral responses can be reasonably fitted with the typical psychometric function commonly found in psychophysical discrimination tasks. As expected, the estimated point of subjective equality (PSE) for the neutral and non-neutral image is at $\alpha = 0$ (99% CI $[-0.00, 0.00]$). The estimated guess rate (equivalent to the lapse rate of negative α's) is 0.07 (99% CI $[0.07, 0.08]$) and the estimated lapse rate is 0.17 (99% CI [0.17, 0.18]).

We found a difference between the rating of high-$|\alpha|$ images for positive and negative α-values (Figure 4b). People tend to agree more with GANalyze when it generates images that are supposed to be of lower aesthetic value. This difference is especially pronounced for the extreme variations. A second order polynomial multiple regression with the predictors $|\alpha|$, $|\alpha|^2$, and dummy variable `positive` explained 95% of the observed variance ($R^2 = .95, F(3, 10) = 90.59, p < .001$). $|\alpha|$ and $|\alpha|^2$ both significantly predicted the proportion agreement with GANalyze ($\beta = 3.87, p < .001; \beta = -10.03, p < .001$). The `positive` variable was also significant ($\beta = -0.05, p < .05$), indicating an asymmetry in the appreciation of the generated high- and low-aesthetic images.

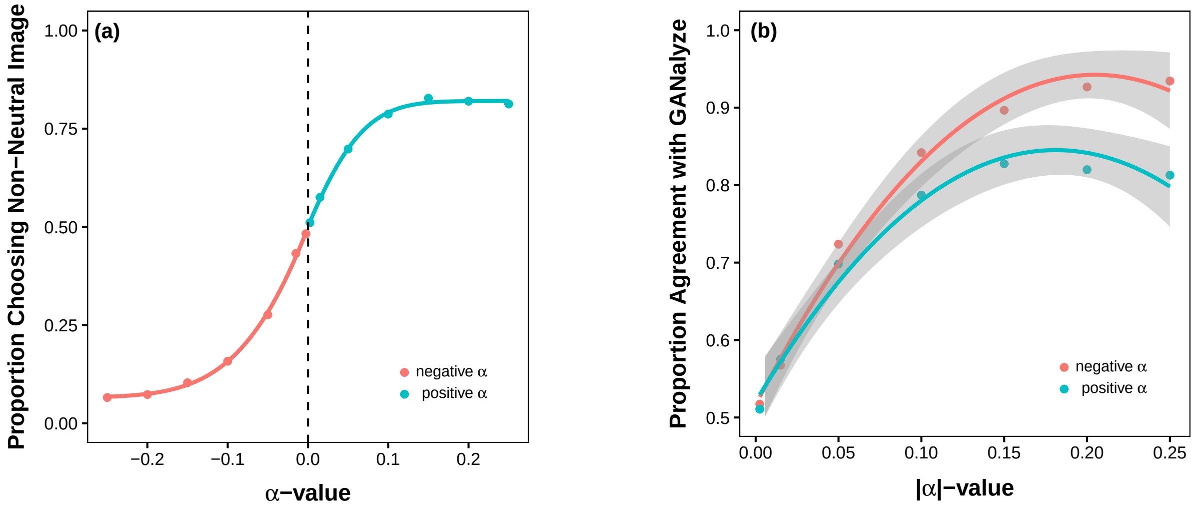
Confounding Factors

There was no effect of broad category (), no effect of image position (), no effect of .

Discussion

Figure 4

Behavioral Results from the Image Rating Task



(a) Psychometric function for the proportion of chosen non-neutral images, illustrating that the estimated threshold value is approximately 0. Choosing the non-neutral image for a negative α-trial corresponds with a disagreement with GANalyze and vice versa. (b) Proportion of agreement with GANalyze with positive and negative α's taken together, fitted with second-order polynomial multiple regression.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., . . . Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

Bode, C., Helmy, M., & Bertamini, M. (2017). A cross-cultural comparison for preference for symmetry: Comparing British and Egyptians non-experts. *Psihologija*, *50*(3), 383–402. https://doi.org/10.2298/PSI1703383B

Brock, A., Donahue, J., & Simonyan, K. (2019, February 25). *Large Scale GAN Training for High Fidelity Natural Image Synthesis*. Retrieved May 17, 2021, from http://arxiv.org/abs/1809.11096

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, *35*(1), 53–65. https://doi.org/10.1109/MSP.2017.2765202

Csikszentmihalyi, M., & Robinson, R. E. (1990). *The art of seeing: An interpretation of the aesthetic encounter*. Getty Publications.

Goetschalckx, L., Andonian, A., Oliva, A., & Isola, P. (2019). GANalyze: Toward Visual Definitions of Cognitive Image Properties. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 10. https://doi.org/10.1109/ICCV.2019.00584

Goetschalckx, L., Andonian, A., & Wagemans, J. (2021). Generative adversarial networks unlock new methods for cognitive science. *Trends in Cognitive Sciences*, *25*(9), 788–801.

Goldstein, E. B. (2009). Perceiving objects and scenes: The gestalt approach to object perception. *Goldstein EB. Sensation and Perception. 8th ed. Belmont, CA: Wadsworth Cengage Learning*, *1*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48. https://doi.org/10.1016/j.neucom.2015.09.116

Hoenig, F. (2005). Proceedings of the first eurographics conference on computational aesthetics in graphics, visualization, and imaging.

Hume, D. (1740). *A treatise of human nature*. Courier Corporation.

Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and predicting image memorability at a large scale. *Proceedings of the IEEE international conference on computer vision*, 2390–2398.

Kong, S., Shen, X., Lin, Z., Mech, R., & Fowlkes, C. (2016). Photo Aesthetics Ranking Network with Attributes and Content Adaptation. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016* (pp. 662–679). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_40

Linares, D., & López-Moliner, J. (2016). Quickpsy: An r package to fit psychometric functions for multiple groups. *The R Journal, 2016, vol. 8, num. 1, p. 122-131.*

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? *Personality and Social Psychology Review*, *8*(4), 364–382. https://doi.org/10.1207/s15327957pspr0804_3

Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling. *Journal of statistical software*, *48*, 1–36.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sartwell, C. (2017). *Entanglements: A system of philosophy*. SUNY Press. https://books.google.be/books?id=qDUrvgAACAAJ

Sartwell, C. (2022). Beauty. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2022). Metaphysics Research Lab, Stanford University.

Sejnowski, T. J. (2018). *The Deep Learning Revolution*. MIT Press.

Valenzise, G., Kang, C., & Dufaux, F. (2022). Advances and Challenges in Computational Image Aesthetics. In B. Ionescu, W. A. Bainbridge, & N. Murray (Eds.), *Human Perception of Visual Information: Psychological and Computational Perspectives* (pp. 133–181). Springer International Publishing. https://doi.org/10.1007/978-3-030-81465-6_6

Van Geert, E., & Wagemans, J. (2021). Order, complexity, and aesthetic preferences for neatly organized compositions. *Psychology of Aesthetics, Creativity, and the Arts*, *15*(3), 484.

van Rossum, G., & Team, P. D. (2016). *Python 3.6 language reference*. Samurai Media Limited.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wanzer, D. L., Finley, K. P., Zarian, S., & Cortez, N. (2020). Experiencing flow while viewing art: Development of the Aesthetic Experience Questionnaire. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(1), 113–124. https://doi.org/10.1037/aca0000203

Wölfflin, H. (1932). Principles of art history, trans. *MD Hottinger (New York, 1932)*, *229*.

Appendix

Appendix A: Shortened Aesthetic Experience Questionnaire

In general, when I view art...

1. I experience a wide range of emotions. (emotional)
2. I compare the past culture of the art with present-day culture. (cultural)
3. The composition of a work of art is important to me. (perceptual)
4. I try to understand the work completely. (understanding)
5. I have a clear idea of what to look for when viewing the work of art. (flow-proximal)
6. I lose track of time when I view the work of art. (flow-experience)

Appendix B: Task Instructions