EMPIRICAL INSIGHTS INTO ARCHITECTURAL AESTHETICS: A NEUROSCIENTIFIC PERSPECTIVE

ELISSA HARTANTO[1] , ASHLEY CHEN[2] AND IMMANUEL KOH[3]
[1,2,3]Singapore University of Technology and Design.
[1]elissa_hartanto@sutd.edu.sg, 0000-0001-7487-6978
[2]ashley_chen@sutd.edu.sg, 0009-0008-3010-8131
[3]immanuel_koh@sutd.edu.sg, 0000-0002-1181-1082

Abstract. What makes a design beautiful? Design styles developed during past eras such as byzantine, classical, gothic, renaissance and baroque are universally admired as being cultural icons and are widely appreciated by people from all walks of life. Throughout the years, many philosophers, architects and physicists have come up with theories and frameworks to measure the subjective topic of aesthetics, but none stood out like Birkhoff's aesthetic measure which used a mathematical approach to quantifying beauty. In this paper, we investigate the aesthetic appeal of generative AI model outputs, trained on datasets recognized for their aesthetic quality, and by employing biometric data analysis to cross-reference these results with Birkhoff's aesthetic measurement framework. Stemming from Neuroarchitecture, wearable technologies offer an insight into the correlation between spatial qualities and human perception that can be extended into aiding us, architects in designing better for the built environment. In our experiment, we generated a set of interior images in assorted styles following current interior design trends. The generated outputs are first scored based on Birkhoff's measurements of aesthetics and cross referenced with data obtained from wearable technologies such as an eye tracker and electroencephalogram (EEG) headset. Eye tracking glasses can detect fixations, saccade patterns, and pupil dilation, which can reflect subconscious thoughts from the user. The EEG is also utilised to complement the eye tracking data as a means to reflect on positive or negative impressions towards a particular subject. Overall, this innovative approach adapts Birkhoff's aesthetic measurement in a human-centric and evidence-based way, providing architects with a framework to systematically evaluate design. It merges Birkhoff's theorem with unbiased subconscious metrics to compare current and historical aesthetic trends, and behavioural research to pinpoint common aesthetic preferences. This method also leverages biometric data to align architectural design more closely with user perspectives, breaking down traditional communication barriers and offering clearer insights into client preferences.

Keywords. Neuroaesthetics, Generative Design, Eye-tracking, ML

ACCELERATED DESIGN, Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2024, Volume 3, 69-78.

1. Introduction

1.1. MEASUREMENT OF AESTHETICS ACCORDING TO: BIRKHOFF

What makes a design beautiful? Design styles developed during past eras such as byzantine, classical, gothic, renaissance and baroque are universally admired as being cultural icons and are widely appreciated by people from all walks of life. Throughout the years, many philosophers, architects, and physicists have come up with theories and frameworks to measure the subjective topic of aesthetics. Out of these theories, Birkhoff's Theorem of Aesthetics (Douchová, 2016), is one of the many aesthetic measuring tools that are still relevant to this day, garnering attention and interest for appealing to those beyond the design field by bringing a mathematical perspective to qualifying art and various types of artistic work (Hubner, 2023), suggesting that the aesthetic appeal of an artwork is related to the balance between its order and complexity. Aesthetic measure increases when the perceived order is high relative to complexity.

1.2. GENERATIVE AI

The decision to assess AI-generated interiors stemmed from the notion that Generative AI models possesses a level of aesthetic comprehension, allowing us to peer into the collective view and understanding of aesthetics throughout the ages (Hullman, 2023). These models have been trained on datasets renowned for being “universally” appealing, capturing the evolving essence of beauty as perceived across different epochs. The process by which AI perceives aesthetics involves several key steps. Firstly, the data collection for AI systems requires large datasets of images or other visual data to learn and understand patterns. When applying a criterion for the definition of what is considered as Aesthetic, developers of the AI system define specific aesthetic criteria or parameters that the model should consider when evaluating aesthetics (Epstein & Hertzmann, 2023). There is an algorithmic evaluation that processes new, unseen data based on the learned patterns and aesthetic criteria. It then assigns aesthetic scores or rankings to the input based on how well it aligns with the learned aesthetic patterns through feedback loops, used to refine and improve the model over time (Gillis, 2023). AI's perception of aesthetics is based on the patterns it has learned from the training data.

1.3. MEASUREMENT OF AESTHETICS ACCORDING TO: WEARABLE TECHNOLOGY

Based on recent advancements in wearable technologies such as Neuralink (Capoot, 2022), there is an anticipation that wearable devices become a common public data collection for the in the future. This research paper employs the use of a screen-based eye tracker and an EEG headset to obtain perceptions on aesthetics. Eye trackers can produce saturations in heatmaps developed from the recorded data such as pupil size to indicate levels of interest or arousal, measuring a physiological level of aesthetic response. The patterns and duration of fixations and saccades can reveal how viewers engage in an image, and how the overall composition affects their viewing experience. Data obtained from the EEG correlates to emotional response and acts as another measure to validate the eye tracking results. Frontal asymmetry in EEG studies measures the difference in activity between the left and right frontal brain regions, often used as an indicator of emotional processing, where greater left-side activation suggests approach-related emotions and right-side activation indicates withdrawal-related emotions (iMotions, 2022). Ultimately, incorporating EEG and eye-tracking allows for a multidimensional approach to understanding aesthetics, combining the objectivity of computational analysis with the subjectivity of human perception and neurophysiological response. This can lead to new insights, such as identifying which aspects of complexity are most engaging or which elements of order are most harmonious to the human eye.

2. RESEARCH QUESTION

In this paper, we examine the aesthetic appeal of generative AI model outputs, trained on datasets recognized for their aesthetic quality, by employing evidence-based human data analysis to cross-reference these results with Birkoff's aesthetic measurement framework. Our hypothesis posits that if artificial intelligence can determine the most aesthetically pleasing images based on Birkoff's theorem, participants will exhibit similar responses to these images.

3. METHODOLOGY

3.1. BIRKHOFF'S AESTHETIC MEASUREMENT

A series of interiors were generated using a text-to-image generative model following current design trends. Nine images, depicting the living room, dining area, and bedroom, were created as universally familiar spaces between the participants. This approach ensures these environments, devoid of specific context, can be perceived as existing anywhere, effectively eliminating preconceived notions or biasness about its location. The nine interiors were then evaluated with Birkoff's formula. Using ChatGPT4 by Open AI, Birkhoff's aesthetic measure M is a concept from the field of mathematics and aesthetics, proposed by Birkhoff.

Birkhoff's theorem suggests that the aesthetic value M of a work of art is directly proportional to its order O, which is a measure of symmetry or balance, and inversely proportional to its complexity C, which represents the number of elements or the intricacy of the design (Douchová,2016). Mathematically, it can be expressed as:

$$M = \frac{O}{C}$$

$$M = \frac{Symmetry\ Score + Vantage\ Point\ Score}{Number\ of\ Segments + Ratio\ of\ SD\ for\ Color\ Histogram}$$

Fig 1: Birkhoff's Aesthetic Measure, Fig 2: Newly interpreted M for generated interior images

M is reinterpreted for image analysis. M is a ratio of the summation of symmetry and vantage scores which represent O, over the summation of number of segments and ratio of standard deviation (SD) for colour histogram which represent C (figure 2). Symmetry score indicates the balance and harmony in an image while vantage point score encapsulates elements of balance and focus from the perspective of the observer (figure 3). Vantage point assessment involves examining factors such as composition and perspective within the images, which includes how elements are arranged and presented from the viewer's perspective. The use of an AI model also considered the spatial relationships among different objects or elements, and how they contribute to the overall aesthetic appeal. For instance, in interior design, a room that exhibits a clear focal point, balanced furniture arrangement, and harmonious spatial relationships might achieve a high vantage point score.

A higher value of symmetry and vantage points score pertains to the readability and comfort in perceiving the space. Number of segments is an indication of the distinct areas or elements in the image which could represent a level of clutter and disorganisation. The elements in the generated images were distinguished using a segmentation model. The total number of elements was divided by eighty-three, which is the maximum number of elements among the nine generated images (figure 3). Lastly, the SD in the colour histogram was obtained from photoshop and divided by 127.5 to obtain the ratio of SD (figure 3). In an 8-bit grayscale image, the maximum SD would occur when half the pixels are black (0) and half are white (255), which would give a SD of 127.5. A higher SD in the colour histogram implies more complexity due to a broader range of colours.

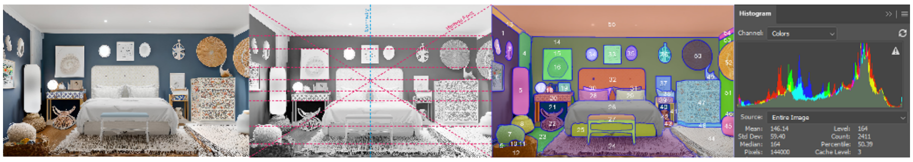

Fig 3: Generated interior of Bedroom 2, Symmetry, and vantage point lines, numbered segmented image, colour value histogram from photoshop (Left-Right)

For the initial aesthetic assessment with Birkhoff's Aesthetic Measurement, ChatGPT 4 (OpenAI) was used to calculate the above equations and based on the nine images that was uploaded, the AI model returned the following values:

Name	Order	Normalised Complexity	Aesthetic Measure (Normalised)
Living room 1	0.233111	5.030817	0.046337
Bedroom 2	0.327606	4.491548	0.072938
Living room 2	0.382875	4.970591	0.077028
Bedroom 3	0.426224	4.249052	0.100310
Dining room 3	0.364057	3.629249	0.100312fa
Dining room 2	0.500562	4.530557	0.110486
Dining room 1	0.554269	4.557095	0.121628
Living room 3	0.564235	4.373664	0.129007
Bedroom 1	0.728882	3.768575	0.193410

Table 1: The 'Aesthetic Measure (Normalised)

Bedroom 1 (Table 1) achieved the highest aesthetic measure score (0.193410), suggesting that it has the most favourable balance between order (which includes symmetry and vantage point) and complexity (which considers the number of segments and colour distribution) among the nine images. According to the mathematical model we used, this would make it the most aesthetically pleasing image. Dining room 5 is considered to have a median aesthetic value. It neither scored the highest nor the lowest on the aesthetic scale, indicating that its balance of order and complexity is moderate compared to the others. Living room 1 has the lowest aesthetic measure score (0.046337). This means that, within the context of the specific mathematical model applied, it has the least favourable balance between order and complexity. It is important to note that while we use terms like "most" and "least" aesthetic based on the scores, these rankings are based on a theoretical model and specific criteria we've programmed into the calculation.

Fig 4: 9 generated interior images arranged from lowest to highest M (left to right)

3.2. WEARABLE TECHNOLOGY SET UP

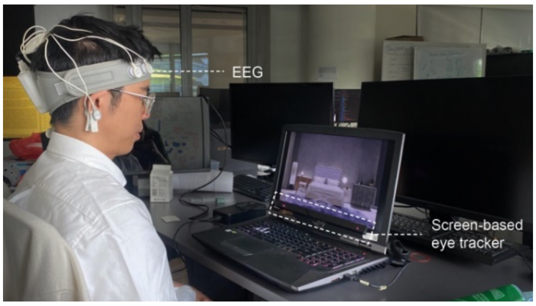

Fig 5: Conducting study on generated image, equipped with Tobii X3-120Hz and Enobio 8.

Initially, a screen-based eye tracker (Tobii Pro X3-120Hz) was affixed to the base of the computer screen, ensuring precise gaze tracking. Participants were also provided with an EEG (Enobio-8) headband, to measure frontal asymmetry. Electrodes were positioned at nodes F7 and F8 to facilitate accurate micro voltage values in neural data collection, which will then be calculated for frontal alpha asymmetry. The core of the experiment involved the presentation of a video compilation consisting of the nine generated interior images, each displayed for a standardised duration of 10 seconds. This setup allowed us to investigate the nuanced relationship between visual attention and neural processes as participants interacted with the images (Albright et al., 2020).

4. RESULTS

4.1. EYE TRACKING

Fig 6: 9 generated interior images re-arranged from lowest to highest FAA (left to right) overlaid with gaze patterns.

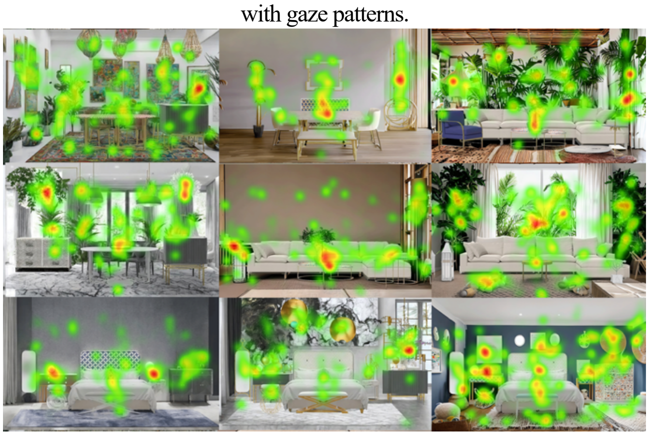

Fig 7: 9 generated interior images re-arranged from lowest to highest FAA (left to right) overlaid with eye tracking heatmaps.

It can be observed that images with higher FAA resulted in more orderly and directed gaze patterns (Figure 6). Interiors with distinct and symmetrically arranged furniture guides the participants' eyes to wander towards these focal points, which in turn creates lines that proposes a high symmetry score. However, when it comes to a more "cluttered" interior, the patterns are haphazard, thus further confirming the lack of order. The heatmaps (Figure 7) are a direct reflection of the intensity of the fixation of the gazes. The higher the concentration, the longer the duration of the fixation. Like the gaze patterns, a room like Dining room 1 (top left) is represented as a heatmap covered in many smaller specks of green over the entirety of the image. T is a result of high number of gazes spread throughout the image but with short fixation timing, indicating disinterest. The intensity of the last image (bottom-right) represents the high fixation duration of the image, from larger bright red to clusters around the image. The user has looked intently within the concentrated areas, which are the furniture pieces.

4.2. FRONTAL ASYMMETRY RESULTS

Interior	Frontal Asymmetry
Dining room 1	0.118
Dining room 3	0.163
Living room 2	0.309
Dining room 2	0.323
Living room 3	0.400
Living room 1	0.411
Bedroom 1	0.560
Bedroom 3	0.627
Bedroom 2	0.719

Table 2: Frontal Asymmetry scores.

The frontal alpha asymmetry (FAA) scores presented (Table 2) are positive numbers, which in the context of EEG research, typically indicate greater relative left frontal activity. Left frontal activation has been associated with approach-related emotions, which are as positive or indicative of engagement. This observation assumes that the stimuli are designed to evoke an emotional response and that the participants' frontal EEG asymmetry reflects their engagement with these stimuli. It also underscores the importance of contextual information for accurate interpretation (Imotions, 2022).

The observed FAA scores reflect a predominance of left frontal cortical activity across a variety of interior stimuli. Notably, the highest positive values, seen in Bedroom 2 (0.719) and Bedroom 3 (0.627), suggest a strong approach-oriented emotional engagement, potentially indicating that the stimuli presented in these conditions were the most positively received. None of the stimuli were associated with withdrawal-related responses, which would be indicated by negative FAA scores.

The variability in the magnitude of these scores reflect differences in the intensity of the positive emotional response elicited by each stimulus. For instance, while Dining Room 1 shows the lowest FAA score (0.118), it still indicates a bias towards approach-related emotional processing, albeit less pronounced than the higher scores.

It is important to note that while higher positive FAA scores correlate with increased approach-related emotional responses, these interpretations are contingent upon the assumption that the stimuli are emotionally or cognitively engaging.

5. DISCUSSION

5.1.1. Bedroom

Based on the aesthetic scores, Bedroom 1 achieves the highest score. This is attributed to the well-defined positions of objects within the room, contributing to positive symmetry scores and a more defined vantage point. The orderly arrangement and the minimal use of colours reduce the complexity score. From an AI standpoint, interiors that display uniform positioning and colour, along with well-defined furniture placement, are considered more aesthetic. However, extremely minimalistic rooms do not necessarily equate to the highest aesthetic appeal. A lack of furniture and colour variety might not be perceived as highly positive.

The wearable technology data, comprising eye-tracking and EEG results, mirror these sentiments. Participants showed a marked preference for Bedroom 1, as evidenced by the high saturation in the focal points of the room, particularly the furniture positions. The gaze patterns align with the interior's symmetry, corroborating the Birkhoff aesthetic measure findings. The saturation in Bedroom 1 illustrates effective fixations on key furniture, by deep reddish colours and large clusters.

It is noteworthy, that while Bedroom 1 scored highly in both aesthetic and wearable technology measures, the other bedrooms (2 and 3) did not exhibit a similar alignment. Despite lower aesthetic scores, these rooms ranked in the top three based on the wearable technology results. The larger clusters in these rooms suggest a more systematic experience, with focal points being the primary furniture.

5.1.2. Living room

Living Room 3 achieved the second-highest score in Birkhoff's aesthetic measure. However, the EEG results placed it in a more neutral position compared to other interior styles. Living Room 1, which scored lowest in the aesthetic measure, showed equivalent results to Living Room 3 in eye-tracking and EEG data, with both rooms ranking in the middle for frontal asymmetry. Living Room 2's results aligned closely with its aesthetic scoring. The heatmap data for these living rooms indicated an equal distribution of scattered gazes and fixations, with focal points around the centre of the images and on specific features within the interior spaces.

5.1.3. Dining Room

The Dining Room images were ranked in the middle to lower end, with Dining Room 1 being the lowest. This contrasts with the EEG results, which placed these rooms in higher positions, except for Dining Room 1, is still viewed as the least fixated. The primary reasons include its broad spectrum of colours and a lack of order due to numerous scattered decorations. EEG data revealed low fixation levels and a high number of scattered gazes showing up as smaller green specks, suggesting a lack of focused interest, due to the overwhelming number of elements in the space. The positioning of Dining Room 3 is at the lower end of the FAA scoring, despite a more neutral aesthetic scoring, indicates that a minimalistic approach in interior design might not always be stimulating, as evidenced by the low directional gaze patterns.

6. LIMITATIONS AND FUTURE WORK

While AI-generated images have improved, achieving high-resolution images with intricate details can still be challenging. Minute details may be blurred or misrepresented. Therefore, ChatGPT4 may have been affected when identifying key perspective and symmetry lines to find the score for the O values. Furthermore, the "warping" effect in these images, as typically produced in ai generated images, may have contributed to the slight discrepancies in the object count and C score. Therefore, training an AI model will produce more accurate images to provide the participants with a realistic environment. In terms of the AI model in ChatGPT4, it still requires further refining and is unable to perform high computational tasks and take in large datasets. Furthermore, it is not evident that GPT4 is entirely correct. The system makes multiple mistakes; therefore their responses still must be thoroughly checked.

Next, generated images are limited to a two-dimensional representation, which inherently lacks the capacity to offer the complete sensory experience that one encounters when physically experiencing a space. Furthermore, there's the potential distraction caused by the contrasting black border around the image, which is a result of resolution limitations. Leveraging the immersive capabilities of virtual reality (VR) could overcome these limitations. Utilising spaces generated by methods like Neural Radiance Fields (NeRF) within VR environments can provide a more immersive and comprehensive platform for exploring and evaluating architectural aesthetics.

Lastly, aesthetic experience involves more than just a balance between order and complexity. It includes cultural influences, personal tastes, and other factors that are not captured by EEG or eye tracking. Further research could benefit from integrating these FAA scores with other behavioural and self-report measures to enrich the understanding of how these interior stimuli affect emotional states.

7. CONCLUSION

Having distinct positioning and geometry of objects is ideal for Birkhoff's Aesthetic Measurement for high symmetrical value, whereas if objects are unbalanced on one side or scattered throughout an interior image, it will be deemed as unorderly. For example, there will be increasing difficulty in quantifying M, as future architectural designs become more complex and irregularly shaped as the industry progresses with computational design and building technologies. This equation was a broader trend in the early 20th century that sought to find mathematical and scientific explanations for concepts traditionally considered subjective, influenced by the rapid advancements in technology and a general belief in progress and rationality. Architects of this time include Le Corbusier, Frank Lloyd Wright, Mies van der Rohe and Walter Gropius who works reflected the technologies, materials, and societal changes, often with a focus on functionalism, simplicity, and a break from traditional forms.

Similarly, artificial intelligence and wearable technology can be widely adopted in the industry and potentially redefine a new design pedagogy. The methodologies written in this paper can account for a new measurement of aesthetics, taking regards the current appreciation and expectation of today's generation. This multidimensional approach of computational analysis, human perception, and neurophysiological response, offers a holistic understanding of aesthetics.

With Virtual Reality/Mixed Reality, simulations can be created based on cognitive data to produce a “perfect environment” that can be directly translated into architectural design. This will introduce a new mode of communication, where designers are able to translate their design languages as an experience for the clients, while clients are able to translate their sentiments to the designers.

References

Albright, T. D., Gepshtein, S., & Macagno, E. (2020). Visual neuroscience for architecture: Seeking a new evidence-based approach to design. Architectural Design, 90(6), 110–117. https://doi.org/10.1002/ad.2639

Capoot, A. (2023, January 25). Precision Neuroscience, co-founded by Neuralink alum, is creating a brain implant thinner than a human hair. CNBC. https://www.cnbc.com/2023/01/25/capoot-precision-neuroscience-12523.html

Carbon, C.-C. (2011). Cognitive mechanisms for explaining dynamics of Aesthetic Appreciation. I-Perception, 2(7), 708–719. https://doi.org/10.1068/i0463aap

Douchová, V. (2016). Birkhoff's aesthetic measure. AUC PHILOSOPHICA ET HISTORICA, 2015(1), 39–53. https://doi.org/10.14712/24647055.2016.8

Epstein, Z., Hertzmann, A., Akten, M., Farid, H., Fjeld, J., Frank, M. R., Groh, M., Herman, L., Leach, N., Mahari, R., Pentland, A. “Sandy,” Russakovsky, O., Schroeder, H., & Smith, A. (2023). Art and the science of Generative AI. Science, 380(6650), 1110–1111. https://doi.org/10.1126/science.adh4451

Gillis, A. S., Burns, E., & Brush, K. (2023, July 27). What is deep learning and how does it work?: Definition from TechTarget. Enterprise AI. https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network

Harmon-Jones, E., & Gable, P. A. (2017). On the role of asymmetric frontal cortical activity in approach and withdrawal motivation: An updated review of the evidence. Psychophysiology, 55(1). https://doi.org/10.1111/psyp.12879

Hu, M., & Roberts, J. (2020). Built environment evaluation in virtual reality environments—a cognitive neuroscience approach. Urban Science, 4(4), 48. https://doi.org/10.3390/urbansci4040048

Hübner, R., & Ufken, E. S. (2023a). On the beauty of vases: Birkhoff's aesthetic measure versus Hogarth's line of beauty. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1114793

Hullman, J., Holtzman, A., & Gelman, A. (2023, August 21). Artificial Intelligence and aesthetic judgement. arXiv.org. https://doi.org/10.48550/arXiv.2309.12338

iMotions. https://imotions.com/blog/learning/best-practice/frontal-asymmetry-101-get-insights-motivation-emotions-eeg/

Jahanian, A. (2016). Quantifying aesthetics of visual design applied to automatic design. Springer Theses. https://doi.org/10.1007/978-3-319-31486-0

Kaplan, S. (1987). Aesthetics, Affect, and Cognition: Environmental Preference from an Evolutionary Perspective. Environment and Behavior, 19(1), 3–32. https://doi.org/10.1177/0013916587191001

Megahed, Y & S.Gabhr, H. (2010). Quantitative architectural aesthetic assessment.

Pedersen, M., Masulli, P., & Bülow, P. (2022, November 21). Frontal asymmetry 101 - how to get insights on motivation and emotions from EEG. Reznik, S. J., & Allen, J. J. (2017). Frontal asymmetry as a mediator and moderator of emotion: An updated review. Psychophysiology, 55(1). https://doi.org/10.1111/psyp.12965

Salingaros, N. A. (1995). The laws of architecture from a physicist's perspective. Physics Essays, 8(4), 638–643. https://doi.org/10.4006/1.3029208