https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-152

The gist of beauty: An investigation of aesthetic perception in rapidly presented images

Caitlin Mullin; Massachusetts Institute of Technology, Cambridge, Massachusetts
Gregor Hayn-Leichsenring; Institute of Anatomy I, Jena University Hospital, Friedrich-Schiller-University Jena, Germany
Christoph Redies; Institute of Anatomy I, Jena University Hospital, Friedrich-Schiller-University Jena, Germany
Johan Wagemans; University of Leuven (KU Leuven), Leuven, Belgium

Abstract

While a work of art can evoke an aesthetic experience, the same can apply to a grand ballroom or a sunset. Like fine art, everyday scenes contain aesthetic qualities, with some scenes being preferred over others. The general meaning of a scene, known as scene gist, is extracted rapidly and automatically, with just a brief glance, mainly from the low spatial frequency information in the image. We asked whether such rapid and coarse overall representation also allows for a stable aesthetic impression. In a series of experiments, we investigated to what extent intrinsic (image type) and extrinsic (temporal and spatial resolution) factors affect the aesthetic response to real-world images. We varied these factors in different groups of participants who rated sets of images for aesthetic pleasure. Here we show that aesthetic responses are extracted rapidly, consistently and automatically with just a glance at scene. Moreover, participants preferred natural scenes over urban or indoor scenes, at both rapid and unlimited exposures. This pattern of preference interacted significantly with self-similarity and anisotropy, two image statistics previously shown to correlate with aesthetic response to artworks. The results presented here allow for a deeper understanding of the aesthetic response to our every-day surroundings.

Introduction

Aesthetic experiences are not restricted to viewing art works exposed in museums and art galleries. Whether created by nature or by human hands, there is a design to the world around us and every time we open our eyes, we engage in an aesthetic evaluation of our environment which can subsequently influence our behaviour. For instance, when enter your favorite restaurant, you might prefer a table in the corner overlooking the whole room, or a table by the window to look outside. Compared to other areas of scene perception research, relatively little investigation has been put forth in understanding what drives environmental scene preferences and how they are formed in our mind.

The paradoxical nature of scene perception makes it an interesting field of study. On the one hand, real-world scenes are incredibly complex with seemingly limitless variations. On the other hand, they show consistent regularities that allow for rapid extraction of meaning from a scene with only a glance (*gist*). The process by which the visual system extracts meaningful information from the environment has been studied extensively. Factors such as spatial and temporal resolution, the attentional requirements and the inherent statistical regularities of scene information allow the visual system to organize the incoming stream of chaotic information into coherent and meaningful representations in just an instant. Given the proposed ubiquity of aesthetic evaluations to our visual environments, this paper will contrast the aspects of scene gist in relation to how we form aesthetic evaluations of real-world scene images.

Scene gist: Spatial/temporal resolution

The speed at which we can meaningfully recognize different visual environments has been well documented [1, 2, 3, 4]. Evidence from both behavioral and neurophysiological sources suggest that we accomplish this rapid recognition through an initial feedforward sweep of cortical activation carrying primarily the coarse, low spatial frequency information of the image [5, 6]. This is likely because the gist of a scene is meant to gather as much information as possible on that first glance.

Bar and colleagues [7, 8] proposed that coarse, low-resolution information is initially extracted in order to facilitate visual object recognition by triggering top-down information regarding the object context. This idea was also part of the reverse-hierarchy theory [9], suggesting that coarse, low-spatial frequency information is first extracted during the rapid feed-forward cortical sweep to activate a global scene template, followed by recurrent feed-back signals required to extract the fine details of a scene. As a result, this initial, rapidly produced *scene gist,* is an imprecise representation that maintains the global relations between elements but loses the identity of the local details [10, 11, 1, 12, 13, 14] See Figure 1.

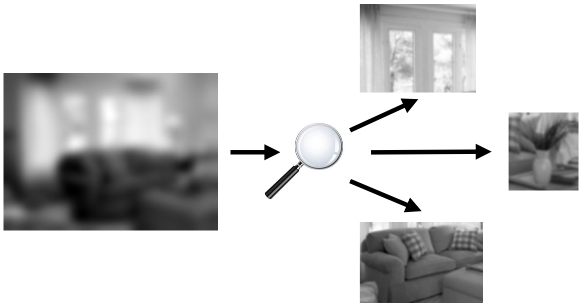

Figure 1. Left image simulates the gist of a scene extracted from the coarse feedforward sweep of information through the hierarchy of the occipito-temporal cortex. Right three images simulate details extracted after feedback for vision with scrutiny [see 9].

Given that meaningful scene information is extracted extremely rapidly and likely at an overall coarse spatial resolution, in Study 1 we asked whether consistent aesthetic evaluations could be elicited from the same kind of impoverished information with extremely short display durations and low resolution images.

Scene gist: Automatic and obligatory

The rapid nature of the way we extract scene category information has led investigators to question whether this process is automatic, meaning that it occurs without directed attention. The role of attention in scene categorization has been controversial.

Some research groups argue that attention is necessary [15], while others suggest that categorization can occur without attention [16]. However, recent findings using a Stroop-like paradigm [17] where scene images themselves were task-irrelevant, revealed that visual categorization is an automatic and obligatory process, meaning that observers cannot help but categorize a scene as soon as they have seen it.

Given that meaningful scene category information is extracted without conscious attention, in Study 2, we asked whether aesthetic evaluations to real-world images also take place automatically during a non-aesthetically related task.

Rapid and automatic aesthetic responses to artistic stimuli

Investigations into the formation of an aesthetic experience from artworks have focused on higher cognitive factors such as successful interpretation and understanding of the artists' intentions [18]. However, recent findings using works of art [19] have found that aesthetic judgments can be extracted rapidly (presentation time of less than 50ms) and automatically, as assessed through implicit tests. Every-day, real-world scenes do not usually have the same picturesque nature as works of art, however they do contain similar low-level statistical regularities [20]. While there are many studies investigating the nature of the aesthetic experience in response to works of art, fewer articles address the aesthetic preference for real-world scenes.

Given their ability to explain aesthetic evaluations of artistic stimuli, in Study 3 we asked whether investigating specific image statistics could explain some of the natural preferences for real-world scene categories.

Study 1: The spatial/temporal nature of the aesthetic evaluations of scenes

In order to determine whether the gist of a scene also contains and aesthetic evaluations, we investigated whether consistent scene preferences could be elicited under gist-like viewing conditions (low resolution and rapid display).

Method

Participants

Three separate groups of 20 participants were recruited to take part in this study. Group 1: 7 men, 13 women, mean age of 24.80 years, SD = 7.16. Group 2: 6 men, 14 women, mean age of 20.85 years, SD = 3.61. Group 3: 7 men, 13 women, mean age of 21.50 years, SD = 4.90. The study was conducted in accordance with the ethical principles regarding research with human participants as specified in The Code of Ethics of the World Medical Association (Declaration of Helsinki). All studies in this paper were approved by the Ethical Committee of the Faculty of Psychology and Educational Sciences (EC FPPW) of the University of Leuven (KU Leuven), and the participants provided informed consent before beginning the experiment.

Stimuli and Design

Stimuli consisted of 75 images selected from the Scene UNderstanding (SUN) database [21]. These images corresponded to a diverse set of naturalistic environments (36 indoor, 39 outdoor images). Given that color is such a strong cue to both scene classification [22] and aesthetic response [23], we selected to greyscale and resize the images (1200 x 900 pixels) in order to emphasize spatial structures. The experimental design consisted of the same stimuli viewed under three different conditions; Full Resolution with Unlimited display duration (FR-U), Low Resolution with Unlimited display duration (LR-U), and Full Resolution with Rapid display duration (FR-R). In a between-groups design, participants rated each image on an aesthetic scale from 1-10 (1 being extremely unattractive or displeasing and 10 being extremely attractive or pleasing) under the different viewing conditions. In all conditions images were presented in the center of the screen at a viewing distance of 60cm from the screen (apparent size 15.18° x 10.94° visual angle). In the FR-U condition, images were presented one at a time and remained on the screen until participants entered an evaluation response.

For the LR-U condition images were filtered using the Gaussian pyramid techniques [24]. In this procedure, the resolution and sample density of an original image were reduced. Repeating the procedure on the output image and reducing it further resulted in a series of quasi-bandpassed images. These images were then expanded by interpolating values between sample points. The selected low resolution images had a spatial frequency of 4.6 cycles per image and were tested in a separate experiment for successful basic level categorization [25]. Images remained on the screen until participant entered a response.

In the FR-R conditions, stimuli were presented for 50ms followed by a static noise mask created by pixel scrambling the original image. The mask was displayed for 1 second followed by the scale where participants were prompted to enter their response.

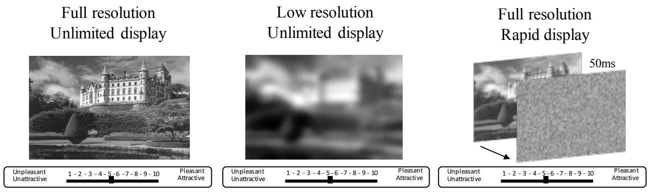

Figure 2. Examples of each viewing condition from Study 1.

Results

A Pearson product-moment correlation coefficient was computed to assess the relationship between the aesthetics score of the three conditions (corrected). Comparing the FR-U condition with the LR-U condition yielded no significant relationship ($r = 0.091$, $p = 0.437$ - see Figure 3A) suggesting that reducing the resolution of real-world scene images alters the aesthetic evaluation. However, comparing the FR-U condition with the FR-R condition yielded a significant positive correlation ($r = 0.76$, $p < 0.01$ – see Figure 3C) suggesting that much of the information needed to arrive at a consistent aesthetic evaluation is available within the timing of scene gist and extended viewing time does not substantially increase or decrease preference of the stimuli. No significant correlation was observed between the LR-U condition and the FR-R condition ($r = 0.173$, $p = 0.136$ – see Figure 3B) suggesting that while meaningful category information is available in both of these viewing conditions their aesthetic evaluation is not equivalent.

Discussion

In their model of aesthetic perception, Leder and colleagues argued that a true aesthetic response takes several stages of processing as well as cognitive mastery of the subject [18]. While this multistage realization of aesthetic beauty may seem to be in contrast to the current findings, his model of aesthetic perception speaks specifically to artistic stimuli that naturally elicit an aesthetic response. The context of which is starkly different from the experience of the participants in the current study.

More recently Redies proposed a model of aesthetic perception that involves the separation of the aesthetic reaction to the perceptual form of an image from its greater cultural and artistic context [26]. The purely perceptual form of beauty is thought to be extracted rapidly (perhaps even unconsciously), with formal properties being universally accepted as pleasing. While this model was also conceptualized to explain aesthetic responses specific to artworks, the channel of the model dedicated to sensory perception can also explain the current results for photographs of scenes. Redies suggests that the intrinsic nature of the image, if found to be beautiful, will activate a beauty-response mechanism in the brain. This mechanism is purely perceptual and universal amongst humans. This theory would account for the extremely high correlation from one subject group to the next [27] as well as the rapid time-course with which participants were able to form their aesthetic evaluations.

The perceptual channel of this model emphasizes the global form or "composition" of the stimuli as paramount for aesthetic appreciation. However, in the LR-U condition, the global aspects of the image were not disrupted, leaving the relationships of the local elements intact. What could account for the lack of significant correlation between this viewing condition and the Full Resolution conditions?

One possible explanation for this is that while reducing the spatial frequency of scene images retains its gist information (basic meaning), fine detail aspects of its content are lost. Content processing makes up the other branch of the Redies model and has a strong effect on the aesthetic response. The loss of content information in the LR-U condition may account for the lack of relationship to unimpaired viewing.

Another possible explanation surrounds the notion of image quality. Photographs may be affected by undesirable blur artifacts from motion of the subject or motion of the camera. These artifacts are commonly thought to be less visually pleasing and some empirical work has demonstrated that reduction in image quality has a strong negative effect on perceived aesthetic evaluation [28]. Therefore, the intentional reduction of image resolution for the LR-U image set may have in fact prematurely reduced the potential for positive aesthetic evaluation in our participants.

Study 2: The automaticity of aesthetic evaluations

Given that the results of Study 1, in which a reliable aesthetic evaluation could be achieved with only a 50ms masked display, we asked whether this was achieved automatically, meaning the reaction would occur even when the observer was not directed to make an aesthetic evaluation. To investigate this further, we employed the Implicit Associations Test (IAT), a method widely used to assess implicit processing on several topics [29].

Method

Participants

Twenty-two participants, took part in the study (8 men, 14 women, mean age of 22.90 years, SD = 4.57).

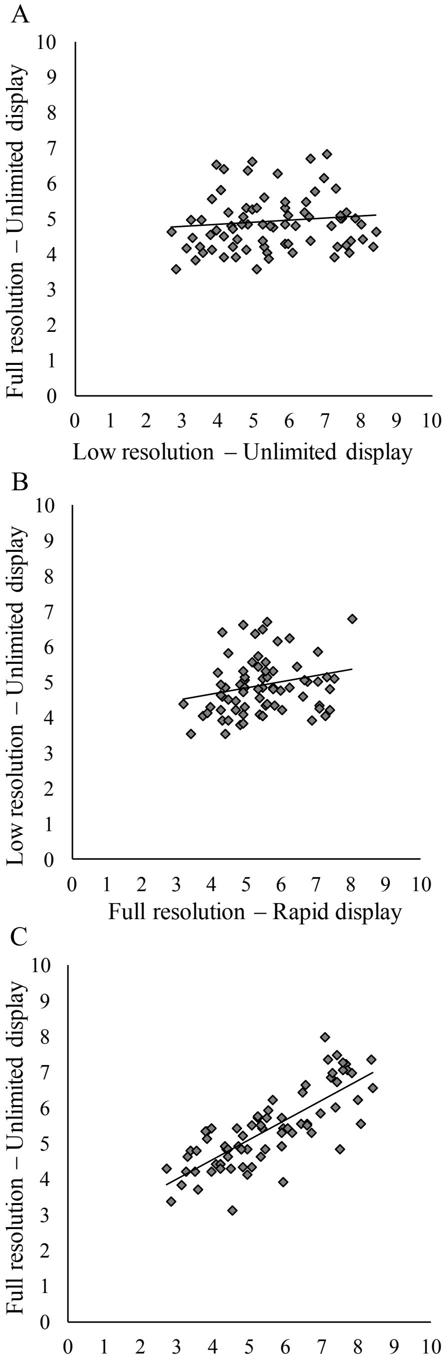

Figure 3. Scatterplot of the relationship between aesthetic evaluation scores of the three viewing conditions of Study 1.

Stimuli and Design

Stimuli were taken from the database of 75 images used in Study 1. The 10 most aesthetically pleasing images and 10 least aesthetically pleasing images were selected from the Full-Resolution-Unlimited viewing condition (see Figure 4). These images revealed some consistencies regarding aesthetic preference for scenes. The most pleasing images were all outdoors and the least pleasing images were all indoors (see below, for further discussion of outdoor/indoor distinction). This complementary category proved useful in designing the IAT study. Participants were told to focus on the indoor/outdoor concept and were unaware of the underlying aesthetic difference between image groups. This allowed us to target the automatic nature of aesthetic processing more directly.

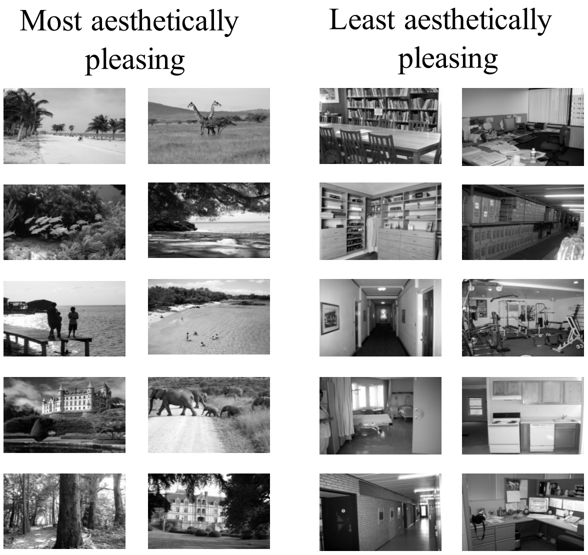

Figure 4. The 10 highest and 10 lowest scoring images from the full resolution-unlimited display duration condition from Study 1.

The IAT is a measurement tool designed to detect the strength of an automatic association between certain concepts and attributes (Greenwald et al., 1998). The current design has a slight variation on the traditional IAT such that the concept of interest was the automatic association of *aesthetically pleasant* vs. *aesthetically unpleasant* scenes, however participants were instructed that the target concept was the indoor/outdoor distinction.

Participants completed the standard IAT design which consists of seven blocks of trials [29]. The test was designed using PsychoPy [30]. Attribute categories were labeled *Pleasant* and *Unpleasant*, and target concepts were labeled *Outdoor* and *Indoor*.

The procedure began with the target concept categorization. In this first block participants categorized the images as indoor or outdoor. This and later categorizations were performed by pressing the left response key for one target concept (e.g. outdoor) and the right response key for the other (e.g. indoor). In the second block the attribute dimension was introduced, also in the form of a two-category differentiation, and participants categorized aesthetically pleasant versus unpleasant words (see Table 1 for the list of words). In the third block, these two categorization tasks were combined to form a dual-categorization task, in which both stimuli for target (indoor/outdoor) and attribute dimension (aesthetically pleasant/unpleasant words) appear. Subsequently, participants learn a reversal of response assignments for the target concepts categorization, and finally the attribute categorization (not changed in response assignments) is combined with this reversed target categorization in the second dual-categorization task. If the target concepts are differentially associated with the attribute dimension, the participant should find one of the combined dual-categorization tasks to be significantly easier than the other. Specifically, if aesthetic processing is indeed automatic, participants should be faster and more accurate in congruent pairings, when outdoor (aesthetically pleasant) images are paired with pleasing words and indoor (aesthetically unpleasant) images are paired with aesthetically unpleasant words, compared to incongruent pairings, pleasant images with unpleasant words. The measure of this difficulty difference provides the measure of implicit association between the target concepts and attribute dimension.

10 pleasant words	**10 unpleasant words**
Beautiful	Ugly
Wonderful	Cluttered
Impressive	Boring
Breathtaking	Repellant
Gorgeous	Hideous
Pleasing	Dreadful
Attractive	Terrible
Enjoyable	Unpleasant
Splendid	Horrible
Marvelous	Repulsive

Table 1. The attribute categories of pleasant and unpleasant words to be associated with the target concepts of high aesthetically pleasing (outdoor) and low aesthetically pleasing (indoor) images in the IAT.

Results & Discussion

We computed the averaged response accuracies and latencies from the main dual-categorization blocks and compared them. The results confirmed our predictions. Participants were significantly more accurate in the congruent block [pleasant image + pleasant word/unpleasant image + unpleasant word] compared with the incongruent block [pleasant image + unpleasant word/unpleasant image + pleasant word] ($t_{21} = 2.24$, $p = 0.03$, one tailed) (Figure 5). Participants also showed a trend towards shorter reaction times for the congruent compared to incongruent condition, however this did not reach statistical significance ($t_{21} = 1.72$, $p=0.06$, one tailed). Taken together, these results suggest that participants could more easily make the congruent association -the outdoor images with pleasant words and the indoor images with unpleasant words, than associate the incongruent pairings together- outdoor images with unpleasant words and indoor images with pleasant words, despite not being directed to making aesthetic evaluations or being aware of any underlying aesthetic variable.

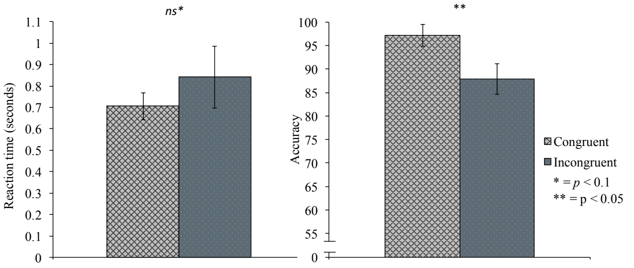

Figure 5. Results of the IAT. Left panel shows trend of participants responding more slowly to the incongruent associations than to the congruent association. The right panel shows that participants were significantly more accurate on the congruent pairs than the incongruent pairs.

The current study attempted to uncover whether aesthetic evaluations to real-world scenes occur automatically by using a task in which aesthetics were irrelevant for the goal of the participants. Studies using the IAT to determine the automaticity of the aesthetic response to artwork [31], designed object [32] or symmetrical patterns [19], have revealed that these evaluations appear to take place rapidly and implicitly. The current results support this finding for real-world scenes, demonstrating that when we look at a scene, we engage a process of aesthetic evaluation that is outside of intention, awareness and control

While the design of this IAT experiment does not allow for the separation of the aesthetic reactions to the stimuli and the overall concept of *outdoor = pleasant* and *indoor = unpleasant*, the relationship between these concepts is clearly one of automatic association. Those studies that have directly addressed aesthetic evaluations of natural scenes and landscapes consistently find observers prefer outdoor scenes with trees and vegetation over urban or indoor scenes [33, 34, 35, 36, 37]. Why we prefer more natural visual environments and the possible underlying features at the root of this preference are explored in Study 3.

Study 3: Investigating the indoor/outdoor aesthetic response

Results of Study 1 revealed that the most aesthetically pleasing images were mainly outdoor and the most aesthetically displeasing images were mainly indoor. This preference for outdoor images has previously been established and widely supported (38, 39, 40, 41, 42, 43). Several theories as to why this preference may exist have been put forward, some evolutionary, some cognitive, and some perceptual.

Kaplan and Kaplan have argued that evolution has left its mark on contemporary humans in the form of an innate preference for natural environments with deflecting vistas [44, 45]. Other investigations have concurred, suggesting that this preference is the result of unlearned factors of evolutionary origin. They assume that because humans evolved over a long period in natural environments, they are to some extent biologically adapted to natural as opposed to built visual content [46, 47, 48]. In addition, natural views tend to be therapeutic compared to urban or indoor scenes in terms of reducing stress or anxiety, and improving post-surgery recovery time (49, 50, 51), suggesting an association with this type of visual information and stress/pain reducing receptors in the brain that may have evolved over time [43].

Several cognitive factors have shown a salient difference between these image types. Natural scenes have shown greater sustained attention and interest much more effectively than the urban views [37]. In addition, outdoor landscapes tend to be less memorable than urban or indoor scenes [52].

From a perceptual perspective, indoor and outdoor scenes differ in their visual properties [53, 54, 55]. Accessing these properties allows observers to quickly gain relevant information about the visual environment and likely drive the evolutionarily adaptive and cognitive responses we have towards these stimulus types. Studies of rapid visual processing demonstrate that we can categorize outdoor images more accurately at brief exposure durations than indoor images [1]. Additionally, investigations of certain statistical features of scene images has revealed that these image statistics are exploited by the visual system in order to estimate the degree of scene naturalness [56].

In an effort to better understand aesthetic preferences, image statistics have also been useful in predicting aesthetic evaluations in artwork and other visual stimuli. In these studies, computer-assisted algorithms are used to extract features that characterize aesthetic images [57, 58, 59, 60, 61]. Image features that can be extracted from the Pyramid Histogram of Oriented Gradients (PHOG), such as Self-similarity, Complexity and Anisotropy have been especially useful in understanding the statistical relationship between artwork and real-world images as well as predicting aesthetic evaluation. Below we briefly introduce these measures.

The property of Self-similarity implies that an image as a whole has an appearance similar to its parts. Closely related concepts are scale-invariance and fractality [62]. Scale invariance is found in many categories of natural images [63], with outdoor foliage and vegetation based images showing some of the highest values. For example, the structure of a tree has the same structure as one of its branches, which has the same structure as the veins of its leaves. Self-similarity is a feature of several different styles of artwork, including but not limited to landscapes as well as portraits (58, 64, 62). Investigations into the role of self-similarity in aesthetic evaluation has revealed that many image categories, which are not aesthetically pleasing, show less self-similarity [60]. Given these findings, we predict a positive correlation between self-similarity as measured by the PHOG method and aesthetic evaluation scores.

There are several objective and subjective methods to measure Complexity. Berlyne [66] described complexity as the regularity of the pattern, the number of elements that form the scene, their heterogeneity, or the irregularity of the forms. He suggested that a high aesthetic appeal is associated with an intermediate level of complexity in an inverted U-shaped function [for a review 66]. Complexity in natural scenes, as measured by the number of independently perceived elements, has shown a moderate to high aesthetic appeal [32, 67]. No significant relationship was observed between Complexity and perceived beauty in abstract artworks [61]. The variation in the relationship between aesthetic evaluation and Complexity in the literature is not be surprising given the different classes of stimuli tested. Therefore, following the literature that has utilized real world images, we predict a moderate to high level of Complexity will be preferred.

Anisotropy is the property of directionally dependent orientations. Natural scenes tend to be less anisotropic (i.e. more isotropic), having edges randomly distributed in various directions, whereas human-made structures such as rooms or cityscapes usually have strong vertical and horizontal edges, making their directions less uniform or more anisotropic [55]. Subsets of Western artworks show a relatively uniform orientation distribution which is similar to natural scenes [68]. However, the contribution of low anisotropy to aesthetic perception remains unspecified. Given that natural images tend to be more isotropic, we predict that the values of anisotropy will be negatively correlated with aesthetic evaluations to the current image set.

We asked whether such image features may be useful in understanding the aesthetic preference for the outdoor scenes observed in Study 1 and others. We set out to replicate the significant findings of Study 1 with a larger dataset that was balanced for indoor, outdoor-urban and outdoor-natural environments. We then used the PHOG method to extract the relevant image statistics from this new dataset for comparison with the aesthetic evaluations from a new pool of participants.

Method

Participants

Two groups of twenty participants took part in the study. Group 1: 5 men, 15 women, mean age of 25.21 years, SD = 10.14. Group 2: 8 men, 12 women, mean age of 24.90 years, SD = 7.94.

Stimuli and Procedure

Stimuli consisted of 300 images selected from the SUN database [21]. These images were balanced for environment type (100 indoor, 100 outdoor-urban and 100 outdoor natural). Images were greyscaled and resized to 1200 x 900 pixels. Given the lack of significant relationship between low-resolution images and our other two viewing conditions from Study 1, this study was restricted to the full-resolution conditions. The experimental design consisted of the same stimuli viewed under two different conditions; Full Resolution with Unlimited display duration (FR-U), and Full Resolution with Rapid display (FR-R). In a between-groups design, participants rated each image on an aesthetic scale from 1-10 (1 being extremely unattractive or displeasing and 10 being extremely attractive or pleasing) under the different viewing conditions. In all conditions images were presented in the center of the screen at a viewing distance of 60cm (apparent size 15.18° x 10.94° visual angle). In the FR-U condition, images were presented one at a time and remained on the screen until participants entered a response. In the FR-R condition, images were displayed for 50ms and masked by a static noise pattern created by pixel scrambling the original image. The mask was displayed for 1 second followed by the scale where participants were prompted to enter their response.

Image statistics

Anisotropy, Complexity and Self-similarity were calculated based on histograms of oriented gradients (HOGs), following approaches that were originally introduced by Dalal and Triggs [69] and Bosch et al. [70].

HOG Anisotropy is a measure for the heterogeneity of luminance gradient strength across orientations in an image. In order to obtain this measure, luminance gradients are binned in 16 equally sized orientation bins that cover the full 360 degrees of orientations in the image [69]. Then, the strengths of the gradients in the different orientation bins are compared. High values of HOG Anisotropy indicate that, for some orientation bins, gradient strength is stronger than for others, while a uniform distribution of luminance gradients across all orientations leads to values close to zero [see 68].

We defined **HOG Complexity** as the total strength of all luminance gradients in an image. Therefore, the sum of the strengths of all oriented gradients, which correspond to the edges or lines with different orientations in the image, was used as a measure of HOG Complexity [68]. Higher values indicate more complex images.

The calculation of **PHOG Self-similarity** is based on a pyramid approach [70]. First, the histogram of oriented gradients for the entire image is calculated (level 0). In the second step, the image is divided into 4 rectangles of the same size and the HOG features are calculated for each rectangle (level 1). Then, each of the 4 subimages are again divided into equal rectangles and the HOG features is calculated for the resulting 16 subimages, as well (level 2). We took this approach up to level 3. In order to calculate PHOG Self-similarity, we compared the HOG features on level 3 (64 subimages) with the HOG features on level 0 by using the Histogram Intersection Kernel [71]. The obtained measure has a range from 0 to 1, with higher values indicating more self-similar images.

The analysis was carried out using MATLAB 2008A [for a more detailed account of the calculations, see the appendix to 72].

Results

A Pearson product-moment correlation coefficient was computed to assess the relationship between the aesthetics score of the two conditions. Comparing the unlimited viewing condition with the rapid viewing condition again yielded a strong significant relationship (r = 0.79, p < 0.001 - see Figure 6) suggesting that most of the information needed to make an aesthetic evaluation is attained in the first glance of a scene.

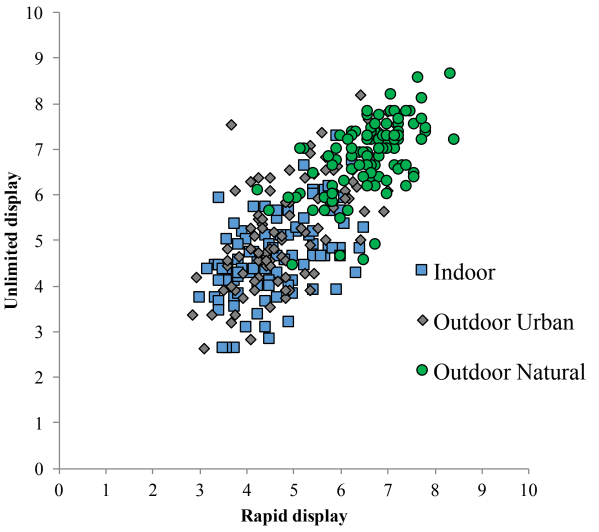

Figure 6. Scatterplot of the relationship between aesthetic evaluation scores for unlimited and rapidly presented displays of scenes.

Breaking down this relationship into the different image classes revealed that outdoor-natural images were preferred over indoor images, replicating the findings from Study1. In view of the finding that observers prefer natural scenes, it is not surprising that the presence of man-made features in natural settings, such as in the outdoor-urban set, elicited more negative preference [73, 74].

Comparing the aesthetic evaluations of these two image classes in the rapid presentation condition with their extracted image features revealed no significant relationship between the aesthetic evaluation score and Anisotropy (r = 0.13, p > 0.05), Complexity (r = 0.03, p > 0.05) or Self-similarity (r = 0.04, p > 0.05). However, when we focused on the more extreme aesthetic dichotomy of outdoor-natural and indoor images an interesting inverse relationship was revealed.

Examining the image feature of Anisotropy (see Figure 7A; for outdoor images, we observed a significant positive correlation between the anisotropy measures and the aesthetic evaluation scores (r = 0.31 p < 0.01). Participants thus preferred non-uniform over uniform orientation distributions. Conversely, for indoor images, we observed a significant negative correlation between the property of Anisotropy and the aesthetic evaluation score (r = -0.22, p = 0.02) meaning that participants preferred uniform over non-uniform orientations in indoor environments.

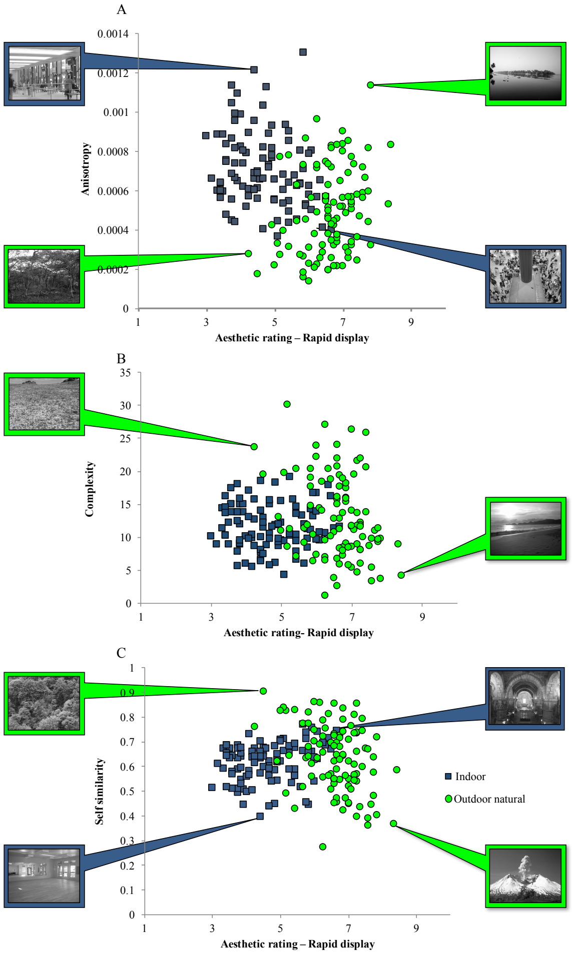

Figure 7. Pearson r correlation between aesthetic evaluations for indoor and outdoor-natural scenes with PHOG measures of (A) Anisotropy, (B) Complexity and (C) Self-similarity.

Examining the image feature of Complexity, we observed a significant negative correlation between the Complexity measure and the aesthetic evaluation score for outdoor images ($r = 0.30$ $p < 0.01$), i.e., participants preferred scenes that were composed of fewer elements (see Figure 7B). No significant relationship between complexity and aesthetic evaluation score was observed for indoor scenes ($r = 0.2$, $p = 0.78$).

Finally, for the image feature of Self-similarity, we observed a significant negative correlation between the Self-similarity score and aesthetic evaluation for outdoor-natural images ($r = -0.35$, $p < 0.01$), meaning that despite the abundance of Self-similar patterns existing in nature, participants preferred less self-similar natural images (see Figure 7C). Conversely, for indoor images, we observed a significant positive correlation between the Self-similarity measure and the aesthetic evaluation ($r = 0.25$, $p = 0.01$) meaning that participants preferred more self-similar and fractal-type indoor images.

Discussion

Given the established relationships between these images features and aesthetic evaluations, the current results were slightly surprising. Although we predicted that participants would prefer images with more uniformly distributed orientations (lower anisotropy), this was only true for indoor images, while less uniformly distributed orientations were preferred for outdoor images. In addition, given that natural images tended to be more self-similar, we predicted a positive relationship with aesthetic evaluation and PHOG Self-similarity. However this relationship was only observed for indoor images and natural images showed the inverse relationship, with less Self-similar images tending to be preferred. A recent study examining clustering algorithms for aesthetic perception revealed that the relationship between complexity and aesthetics was not a single relationship but rather a composite effect of a positive and negative relation subgroup [75]. From the results of Study 2 and others, we case safely assume that indoor images correspond to a negative subgroup and natural-outdoor scene to a positive subgroup. Therefore, closely examining the extremes of the images in Figure 7 might explain these unexpected findings by clarifying the reciprocal nature of the inverse relationship.

Predicting aesthetic evaluations from image statistic for real-world scenes may depend on the original expectation, or template, of that image category and the desire for balance between the extremes of the image property. For instance, the same type of images continues to be preferred in both image categories. For indoor scenes: high self-similarity and low anisotropy, meaning the indoor image tends to have the statistics that are more commonly associated with nature. This may not be surprising as people often tend to decorate their indoor environments with aspects of nature such as plants and other vegetation. The expected uniform orientations, with smooth flat surfaces that often accompany indoor environments may be seen as drab or mundane. Therefore, if the indoor image tends to share image statistics that are more associated with nature images, this may push the aesthetic preference towards a more positive score, even if the associated statistic is not based on 'natural' elements.

When examining natural outdoor scenes, the reverse pattern is observed. Despite the literature suggesting that observers prefer more evenly distributed orientations, moderate to high complex element arrays and more natural self-similar, we observed the opposite pattern for each of our image features. However, this may again stem from the notion that moderation is key. When examining some of the examples from Figure 7, it is clear that extreme values in these features can make the images look more like textures. Previous work on aesthetics for natural images have emphasized a clear focal point as a strong predictor [76]. In these highly self-similar and complex texture-like images the focal point is lost. In addition, the depth plane is restricted, giving the scene a more enclosed perception, more akin to an indoor or built environment [77].

While we replicated the finding that observers prefer natural outdoor scenes to indoor or urban scene presented rapidly, the nature of the image statistics exert their own influence on the results. Future work will more deeply examine the relationship of these image categories to their underlying statistics by artificially manipulating these statistics and observing their influence on aesthetic perception.

Summary

The central findings of these three studies highlight a fundamental aspect of human vision. Within each glance at our environment, we assigning an automatic (possibly unconscious) preference to our surroundings based partly on the arrangement and distribution of low-level visual features.

References

[1] Fei-Fei, Li, et al. "What do we perceive in a glance of a real-world scene?." *Journal of vision* 7.1 (2007): 10-10.

[2] McCotter, M., Gosselin, F., Sowden, P., & Schyns, P. (2005). The use of visual information in natural scenes. *Visual Cognition*, *12*(6), 938-953.

[3] Oliva, Aude. "Gist of the scene." *Neurobiology of attention* 696.64 (2005): 251-258.

[4] Oliva, Aude, and Antonio Torralba. "Building the gist of a scene: The role of global image features in recognition." *Progress in brain research* 155 (2006): 23-36.

[5] Kaplan, Ehud. "The M, P, and K pathways of the primate visual system." *The visual neurosciences* 1 (2004): 481-493.

[6] Xu, Xiangmin, et al. "A comparison of koniocellular, magnocellular and parvocellular receptive field properties in the lateral geniculate nucleus of the owl monkey (Aotus trivirgatus)." *The Journal of physiology* 531.1 (2001): 203-218.

[7] Bar, Moshe, et al. "Top-down facilitation of visual recognition." *Proceedings of the National Academy of Sciences of the United States of America* 103.2 (2006): 449-454.

[8] Kveraga, Kestutis, Avniel S. Ghuman, and Moshe Bar. "Top-down predictions in the cognitive brain." *Brain and cognition* 65.2 (2007): 145-168.

[9] Hochstein, Shaul, and Merav Ahissar. "View from the top: Hierarchies and reverse hierarchies in the visual system." *Neuron* 36.5 (2002): 791-804.

[10] Schyns, Philippe G., and Aude Oliva. "From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition." *Psychological science* 5.4 (1994): 195-200.

[11] Morrison, Donald J., and Philippe G. Schyns. "Usage of spatial scales for the categorization of faces, objects, and scenes." *Psychonomic Bulletin & Review* 8.3 (2001): 454-469.

[12] Sampanes, Anthony Chad, Philip Tseng, and Bruce Bridgeman. "The role of gist in scene recognition." *Vision research* 48.21 (2008): 2275-2283.

[13] Greene, Michelle R., and Aude Oliva. "Recognition of natural scenes from global properties: Seeing the forest without representing the trees." *Cognitive psychology* 58.2 (2009): 137-176.

[14] Rayner, Keith, et al. "Eye movements and visual encoding during scene perception." *Psychological science* 20.1 (2009): 6-10.

[15] Cohen, Michael A., George A. Alvarez, and Ken Nakayama. "Natural-scene perception requires attention." *Psychological science* (2011).

[16] Li, Fei Fei, et al. "Rapid natural scene categorization in the near absence of attention." *Proceedings of the National Academy of Sciences* 99.14 (2002): 9596-9601.

[17] Greene, Michelle R., and Li Fei-Fei. "Visual categorization is automatic and obligatory: Evidence from Stroop-like paradigm." *Journal of vision* 14.1 (2014): 14-14.

[18] Leder, Helmut, et al. "A model of aesthetic appreciation and aesthetic judgments." *British journal of psychology* 95.4 (2004): 489-508.

[19] Bertamini, Marco, Alexis Makin, and Giulia Rampone. "Implicit association of symmetry with positive valence, high arousal and simplicity." *i-Perception* 4.5 (2013): 317-327.

[20] Graham, Daniel J., and Christoph Redies. "Statistical regularities in art: Relations with visual coding and perception." *Vision research* 50.16 (2010): 1503-1509.

[21] Xiao, Jianxiong, et al. "Sun database: Large-scale scene recognition from abbey to zoo." *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*. IEEE, 2010.

[22] Oliva, Aude, and Philippe G. Schyns. "Diagnostic colors mediate scene recognition." *Cognitive psychology* 41.2 (2000): 176-210.

[23] Schloss, Karen B., and Stephen E. Palmer. "Aesthetic response to color combinations: preference, harmony, and similarity." *Attention, Perception, & Psychophysics* 73.2 (2011): 551-571.

[24] Burt, Peter, and Edward Adelson. "The Laplacian pyramid as a compact image code." *IEEE Transactions on communications* 31.4 (1983): 532-540.

[25] Mullin, Caitlin, et al., ‘Perception of real world scenes at multiple spatial scales.” *Journal of Vision.* 2014; 14(10):865-865. doi: 10.1167/14.10.865

[26] Redies, Christoph. "Combining universal beauty and cultural context in a unifying model of visual aesthetic experience." *Frontiers in human neuroscience* 9 (2015): 218.

[27] Vessel, Edward A., and Nava Rubin. "Beauty and the beholder: highly individual taste for abstract, but not real-world images." *Journal of vision* 10.2 (2010): 18-18.

[28] Tinio, Pablo PL, and Helmut Leder. "Natural scenes are indeed preferred, but image quality might have the last word." *Psychology of Aesthetics, Creativity, and the Arts* 3.1 (2009): 52.

[29] Greenwald, Anthony G., et al. "Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity." *Journal of personality and social psychology* 97.1 (2009): 17.

[30] Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1), 8-13.

[31] Pavlovic, Masa Dragoljub. "Automatic processes in aesthetic judgment: Insights from the implicit association test." *Psihologija* 45.4 (2013).

[32] Mastandrea, Stefano, and Fridanna Maricchiolo. "Implicit and explicit aesthetic evaluation of design objects." *Art & Perception* 2.1-2 (2014): 141-162.

[33] Biederman, Irving, and Edward Vessel. "Perceptual Pleasure and the Brain A novel theory explains why the brain craves information and seeks it through the senses." *American scientist* 94.3 (2006): 247-253.

[34] Kaplan, Stephen. "Environmental preference in a knowledge-seeking, knowledge-using organism." (1992).

[35] Kaplan, Rachel, and Stephen Kaplan. *The experience of nature: A psychological perspective*. CUP Archive, 1989.

[36] Kaplan, Stephen, Rachel Kaplan, and John S. Wendt. "Rated preference and complexity for natural and urban visual material." *Perception & Psychophysics* 12.4 (1972): 354-356.

[37] Ulrich, Roger S. "Natural versus urban scenes some psychophysiological effects." *Environment and behavior* 13.5 (1981): 523-556.

[38] Kaplan, Stephen, Rachel Kaplan, and John S. Wendt. "Rated preference and complexity for natural and urban visual material." *Perception & Psychophysics* 12.4 (1972): 354-356.

[39] Zube, Ervin H., Robert O. Brush, and Julius Gy Fabos, eds. “*Landscape assessment: values, perceptions and resources*.” Halsted press, 1975.

[40] Wohlwill, Joachim F. "Environmental aesthetics: The environment as a source of affect." *Human behavior and environment*. Springer US, 1976. 37-86.

[41] Palmer, James F. "An investigation of the conceptual classification of landscapes and its application to landscape planning issues." *Priorities for Environmental Design Research, Part I* (1978): 92-103.

[42] Bernaldez, F. González, and F. Parra. "Dimensions of landscape preferences from pairwise comparisons." (1979).

[43] Biederman, I., & Vessel, E. (2006). Perceptual Pleasure and the Brain A novel theory explains why the brain craves information and seeks it through the senses. *American scientist*, *94*(3), 247-253.

[44] Kaplan, Rachel. "Some psychological benefits of gardening." *Environment and behavior* 5.2 (1973): 145.

[45] Kaplan, Stephen. "Aesthetics, affect, and cognition environmental preference from an evolutionary perspective." *Environment and behavior* 19.1 (1987): 3-32.

[46] Driver, B. L., and Peter Greene. "Man's nature: innate determinants of response to natural environments." (1977).

[47] Iltis, Hugh H., Orie L. Loucks, and Peter Andrews. "Criteria for an optimum human environment." *Bulletin of the Atomic Scientists* 26.1 (1970): 2-6.

[48] Stainbrook, Edward. "Human needs and the natural environment." *Man and Nature in the City* (1968): 1-9.

[49] Ulrich, Roger S., et al. "Stress recovery during exposure to natural and urban environments." *Journal of environmental psychology* 11.3 (1991): 201-230.

[50] Ulrich, Roger S. "Human responses to vegetation and landscapes." *Landscape and urban planning* 13 (1986): 29-44.

[51] Ulrich, Roger. "View through a window may influence recovery." *Science* 224.4647 (1984): 224-225.

[52] Isola, Phillip, et al. "What makes an image memorable?." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.

[53] Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive psychology*, *41*(2), 176-210.

[54] Torralba, Antonio, and Aude Oliva. "Statistics of natural image categories." *Network: computation in neural systems* 14.3 (2003): 391-412.

[55] Vailaya, Aditya, Anil Jain, and Hong Jiang Zhang. "On image classification: City images vs. landscapes." *Pattern Recognition* 31.12 (1998): 1921-1935.

[56] Groen, Iris IA, et al. "From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category." *The Journal of Neuroscience* 33.48 (2013): 18814-18824.

[57] Datta, Ritendra, et al. "Studying aesthetics in photographic images using a computational approach." *European Conference on Computer Vision*. Springer Berlin Heidelberg, 2006.

[58] Graham, Daniel J., and David J. Field. "Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities." *Spatial vision* 21.1 (2007): 149-164.

[59] Redies, Christoph. "A universal model of esthetic perception based on the sensory coding of natural stimuli." *Spatial vision* 21.1 (2007): 97-117.

[60] Amirshahi, S. A., Koch, M., Denzler, J., & Redies, C. (2012). PHOG analysis of self-similarity in esthetic images. *Proc. SPIE (Human Vision and Electronic Imaging XVII)*, *8291*, 82911J.

[61] Mallon, Birgit, Christoph Redies, and Gregor Uwe Hayn-Leichsenring. "Beauty in abstract paintings: perceptual contrast and statistical properties." *Frontiers in human neuroscience* 8 (2014): 161.

[62] Taylor, Richard P., et al. "Perceptual and physiological responses to Jackson Pollock's fractals." *Brain and Art* (2014): 43.

[63] Ruderman, D. L. (1997). Origins of scaling in natural images. *Vision research*, *37*(23), 3385-3398.

[64] Redies, Christoph, et al. "Artists portray human faces with the Fourier statistics of complex natural scenes." *Network: Computation in Neural Systems* 18.3 (2007): 235-248.

[65] Berlyne, Daniel E. Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation. Hemisphere, 1974.

[66] Nadal, M. "Complexity and aesthetic preference for diverse visual stimuli." *PhD, Departament de Psicologia, Universitat de les Illes Balears* (2007).

[67] Ulrich, Roger S. "Visual landscape preference: A model and application." *Man-Environment Systems* (1977).

[68] Redies, Christoph, et al. "PHOG-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects." *European Conference on Computer Vision*. Springer Berlin Heidelberg, 2012.

[69] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1. IEEE, 2005.

[70] Bosch, Anna, Andrew Zisserman, and Xavier Munoz. "Representing shape with a spatial pyramid kernel." *Proceedings of the 6th ACM international conference on Image and video retrieval*. ACM, 2007.

[71] Barla, Annalisa, et al. "Image kernels." *Pattern Recognition with Support Vector Machines*. Springer Berlin Heidelberg, 2002. 83-96.

[72] Braun, Julia, et al. "Statistical image properties of print advertisements, visual artworks and images of architecture." (2013). https://doi.org/10.3389/fpsyg.2013.00808

[73] Clamp, Peter. "Evaluating English landscapes—some recent developments." *Environment and Planning A* 8.1 (1976): 79-92.

[74] Evans, Gary W., and Kenneth W. Wood. "Assessment of environmental aesthetics in scenic highway corridors." *Environment and Behavior* 12.2 (1980): 255-273.

[75] Güçlütürk, Yağmur, Richard HAH Jacobs, and Rob van Lier. "Liking versus complexity: Decomposing the inverted U-curve." *Frontiers in human neuroscience* 10 (2016).

[76] Ulrich, Roger S. "Aesthetic and affective response to natural environment." *Behavior and the natural environment*. Springer US, 1983. 85-125.

[77] Küller, Rikard. "A semantic model for describing perceived environment." (1972).