Title
Using Gibbs Sampling with People to characterize perceptual and aesthetic evaluations in multidimensional visual stimulus space

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Van Geert, Eline
Jacoby, Nori

Publication Date
2024

Using Gibbs Sampling with People to characterize perceptual and aesthetic evaluations in multidimensional visual stimulus space

Eline Van Geert[1,2] (eline.vangeert@kuleuven.be)
Nori Jacoby[1] (nori.jacoby@ae.mpg.de)
[1]Computational Auditory Perception Group, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[2]Laboratory of Experimental Psychology, Department of Brain and Cognition,
Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium

Abstract

Aesthetic appreciation is inherently multidimensional: many different stimulus dimensions (e.g., colors, shapes, sizes) contribute to our aesthetic experience. However, most studies in empirical aesthetics used either non-parametrically controlled multidimensional or parametrically controlled unidimensional stimuli, preventing insight into the relative contribution of each stimulus dimension or any potential interactions between them to perceptual and aesthetic evaluations. To adress this gap we combined two recent developments: the Order & Complexity Toolbox for Aesthetics (Van Geert, Bossens, & Wagemans, 2023) for generating multidimensional parametrically controlled stimuli, and Gibbs Sampling with People (Harrison et al., 2020) for efficiently characterizing subjective evaluations in multidimensional stimulus space. We show the advantages of this new approach by estimating multidimensional probability distributions for both aesthetic (pleasure and interest) and perceptual evaluations (order and complexity) in two visual multidimensional parametric stimulus spaces, and we compare our results with findings from earlier studies that used either non-parametric or unidimensional stimuli.

Keywords: empirical aesthetics; order; complexity; multidimensionality; aesthetic appreciation; OCTA toolbox; Gibbs Sampling with People; transmission chain experiments

Introduction

Every day we evaluate the world around us in aesthetic terms. This feature of our cognition in turn influences designers, artists, architects, and advertisers, whose work shapes our environment. A key challenge to studying appreciation is its multidimensional nature: many different stimulus dimensions contribute to our aesthetic experience, and they may do so to a different extent and in different ways (Van Geert, Warny, & Wagemans, 2024). Most previous research either manipulated a single stimulus dimension (e.g., Spehar, Walker, & Taylor, 2016; Wilson & Chatterjee, 2005; Sun & Firestone, 2022), or used non-controlled stimulus sets in which differences on multiple stimulus dimensions cannot be clearly separated (e.g., natural images, paintings, or photographs; Braun, Amirshahi, Denzler, & Redies, 2013; Graf & Landwehr, 2017; Van Geert & Wagemans, 2021). Both approaches prevent insight in the relative importance of each of the single dimensions or any interactions between them in a multidimensional context. Even when studies did include multidimensional parametric stimuli, some of the dimensions were discretized rather than studied in a continuous fashion (e.g., Jacobsen & Höfel, 2002; Gartus & Leder, 2013; Palmer & Schloss, 2010).

Why are full-fledged multidimensional investigations of parametric stimulus spaces lacking so far? First, an easy tool was lacking to generate such multidimensional parametrically controlled stimuli. Second, using traditional research methods, multidimensionality comes with a cost, sometimes called the 'curse' of dimensionality, namely leading to an exponential increase in the required number of trials. Only recently, new methods have been developed that can efficiently sample high multidimensional spaces without leading to exponential increases in the number of trials required (Martin, Griffiths, & Sanborn, 2012; Sanborn, Griffiths, & Shiffrin, 2010; Harrison et al., 2020; van Rijn et al., 2021, 2022).

The current project enabled the characterization of subjective evaluations for fine-grained, high-dimensional parametric stimulus spaces by combining the Order & Complexity Toolbox for Aesthetics (OCTA; Van Geert et al., 2023) for stimulus creation and Gibbs Sampling with People (GSP; Harrison et al., 2020) for efficiently sampling the high-dimensional space. We show how this new approach can further our understanding of the relations between commonly studied aspects of perceptual appearance (perceived order and complexity; e.g., Van Geert & Wagemans, 2020) and aesthetic appreciation (pleasure and interest; e.g., Graf & Landwehr, 2017) in two visual stimulus spaces.

We hypothesized distinct contributions of each stimulus dimension to the subjective evaluations as well as distinct multidimensional landscapes for each of the evaluation criteria. We expected to confirm the correlations between evaluation criteria found in previous research using non-parametric or unidimensional stimuli (e.g., Van Geert & Wagemans, 2021), but also to bring new insights into how different stimulus dimensions interact in contributing to our perceptual and aesthetic evaluations.

Background

The Order & Complexity Toolbox for Aesthetics

Recently the Order & Complexity Toolbox for Aesthetics (OCTA; Van Geert et al., 2023) was created. OCTA is available as a Python toolbox as well as an online point-and-click application and allows researchers to create reproducible multidimensional, parametric stimulus sets. The focus of OCTA is on the creation of multi-element displays varying qualitatively (i.e., different types) and quantitatively (i.e., different levels) in order and complexity, based on regularity and variety along multiple element features (e.g., shape, size, color, orientation). Recent studies using OCTA have manipulated color, shape, and size complexity and found that

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

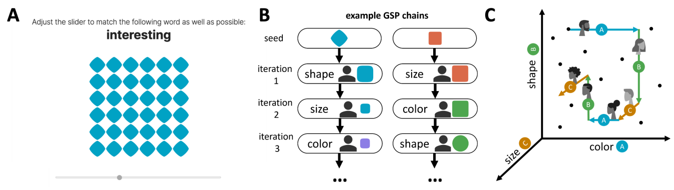

Figure 1: Illustration of the Gibbs Sampling with People (GSP) method like in Experiment 1. (A) Interface for a participant. (B) Illustration of the GSP procedure. (C) Illustration of how GSP efficiently samples the multidimensional space.

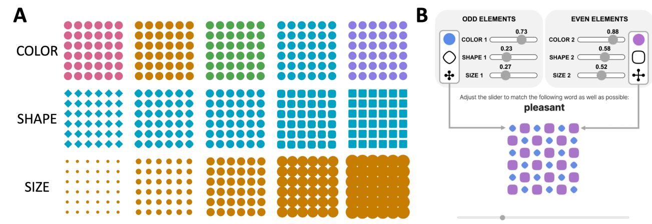

Figure 2: Illustration of the parameters included in (A) Experiment 1 and (B) Experiment 2.

color complexity was more positively appreciated than shape and size complexity (Van Geert, Warny, & Wagemans, 2024; Van Geert, Hofmann, & Wagemans, 2024). In addition, appreciation for complexity decreased when the level of order in the pattern decreased (operationalized as the number of element position switches in the pattern; Van Geert, Warny, & Wagemans, 2024), for each of the tested complexity manipulations. Although these studies varied the *number* of different colors, shapes, and sizes in the stimulus, they fixed the *size* of the difference between these feature values per dimension.

Efficiently estimating subjective probability distributions in multidimensional space

Gibbs Sampling with People (GSP; Harrison et al., 2020) is an efficient way to sample internal representations from a multidimensional latent space. In the paradigm, participants explore a single continuous stimulus dimension per trial with a slider (cf. Figure 1A) and select the value on the stimulus dimension that maximizes the given criterion (e.g., interesting). The stimulus in the next iteration will have the parameter value selected in the previous trial as a fixed characteristic, and another continuous stimulus dimension is then explored (cf. Figure 1B-C). As the slider dimension that participants adjust varies across trials, GSP allows to explore the full multidimensional space over the course of the trials. It can be shown that in GSP parameter combinations are sampled proportionally to their exponentiated subjective utility (Harrison et al., 2020), which allows to characterize the latent distribution of criterion values in the multidimensional space. Some first studies using GSP show the method's promise in both visual and auditory stimulus domains (Harrison et al., 2020; Kumar et al., 2022; van Rijn et al., 2021, 2022, 2024).

In the two experiments reported here, we combined OCTA for multidimensional parametric stimulus generation and GSP for efficiently sampling the multidimensional space. We demonstrate the fruitfulness of combining these techniques to study visual empirical aesthetics in a systematic, multidimensional manner. In a first experiment, we investigated the influence of absolute color, shape, and size on how ordered, complex, pleasant, and interesting an image is evaluated to be. In a second study, we introduced additional complexity by allowing participants to vary the absolute color, shape, and size of odd and even elements in the image independently.

Methods

Participants

Participants were recruited online via Prolific and provided consent in accordance with an approved protocol (Max Planck Ethics Council #2021_42). We recruited native English speakers, born in and currently residing in the United Kingdom. Participants who failed a brief color blindness test were excluded from participation. In the first GSP experiment, 98 participants (36 female, 61 male, 1 other; M_{age}=39.4 years, SD_{age}=11.8 years) took part. In Experiment 2, GSP data was collected from 195 participants (74 female, 119 male, 2 other; M_{age}=39.0 years, SD_{age}=11.7 years). Participants received £3.50 for their participation in the study, which on average took 20-25 minutes to complete.

Stimuli

Stimuli were vector images created using OCTA. For Experiment 1, 6-by-6 grid stimuli of identical elements were generated in which the color, shape, and size of the elements could vary (cf. Figure 2A). These stimulus dimensions were chosen because they are core visual dimensions that have shown a relation to perceptual appearance and aesthetics in previous research (e.g., Palmer, Schloss, & Sammartino, 2013; Schloss & Palmer, 2011; Smart & Szafir, 2019). In the color domain, stimuli varied in hue. The colors were generated in Oklch space (Ottosson, 2020), with a fixed lightness and chroma (0.65 and 0.15, respectively) and then transformed into their respective RGB color. Shapes were squircles, the most common specific case of a larger class of supershapes (Gielis, 2003): $r=\left(\left|\frac{cos(\theta)}{a}\right|^n+\left|\frac{sin(\theta)}{a}\right|^n\right)^{-1/n}$, with a being the semi-diameter of the shape and the n parameter varying between 1 (diamond) and 10 (approximate square), with 2 (circle) as the midpoint. The size of the shapes varied between 30% and 130% of the available element spacing, simultaneously in the horizontal and the vertical direction. In Experiment 1, all elements in the grid had identical feature values, leading to a three-dimensional (color x shape x size) stimulus space (cf. Figure 2A). In Experiment 2, the same feature dimensions were used, but now applied in a checkerboard pattern: odd and even elements could change color, shape, and size independently, leading to a six-dimensional stimulus space (cf. Figure 2B).

GSP task

In the GSP task, participants were asked to adjust a slider to make an image as ordered, complex, pleasant, or interesting as possible (cf. Figure 1A). Experiment 1 contained 200 chains of 24 iterations (i.e., 8 cycles through all 3 parameters per chain; each chain starting from a different random parameter combination; 50 chains per evaluation criterion). Each chain included responses from several participants, and each participant contributed to multiple chains. Depending on the trial, the slider changed the elements' color, shape, or size. Experiment 2 contained 200 across-participant chains of 48 iterations. Depending on the trial, the slider changed the odd or even elements' color, shape, or size (cf. Figure 2B).

Procedure

In both experiments, participants went through the same procedure. After providing informed consent, they conducted a color blindness test consisting of six Ishihara plates (Clark, 1924; Harrison et al., 2020). To determine the scaling factor for fixing the stimulus presentation size regardless of screen resolution, participants resized a rectangle on the screen to match the size of a physical object with known size (i.e., student card, identity card, or bank card).

Participants were instructed to sit at an arm's length distance directly in front of the screen, and to keep that distance as constant as possible throughout the experiment. After providing some demographic information (i.e., their gender, age, education level, mother tongue, country of birth, and country of residence), participants received instructions concerning the GSP task and conducted maximally 50 GSP trials. Afterwards, participants completed two brief questionnaires: the Personal Need for Structure (PNS; Neuberg & Newsom, 1993) and the Art Interest scale of the Vienna Art Interest and Art Knowledge questionnaire (VAIAK; Specker et al., 2020). Both experiments were implemented with PsyNet (Harrison et al., 2020), a platform for designing and running complex online experiments (`www.psynet.dev`).

Results

Experiment 1

Experiment 1 examined to what extent certain color, shape, and size combinations are perceived as ordered, complex, pleasant, or interesting in 6-by-6 multi-element displays with identical elements. We estimated the multidimensional probability distribution for each evaluation criterion by fitting a three-dimensional kernel density estimate (KDE) to the GSP samples.[1] From these three-dimensional KDEs, also the one- and two-dimensional KDEs can be calculated. We assessed split-half reliability of the reported three-dimensional KDEs for each evaluation criterion via bootstrapping 1000 times (95% highest density continuous interval [HDCI] around the mean reliability estimate). The lower boundary of this reliability interval exceeded .70 for each of the four evaluation criteria, which assures the KDE results can be reliably interpreted (r_{ordered} = .90 [.85, .94]; r_{complex} = .84 [.78, .89]; r_{pleasant} = .79 [.74, .84]; $r_{\text{interesting}}$ = .82 [.77, .87]).

[1] For the KDEs of Experiment 1 and Experiment 2, we used a kernel width of .07 and .10 respectively. In the analyses, the KDE is calculated based on a grid size of .05. In some of the Figures, we represent the KDE with a grid size of .01.

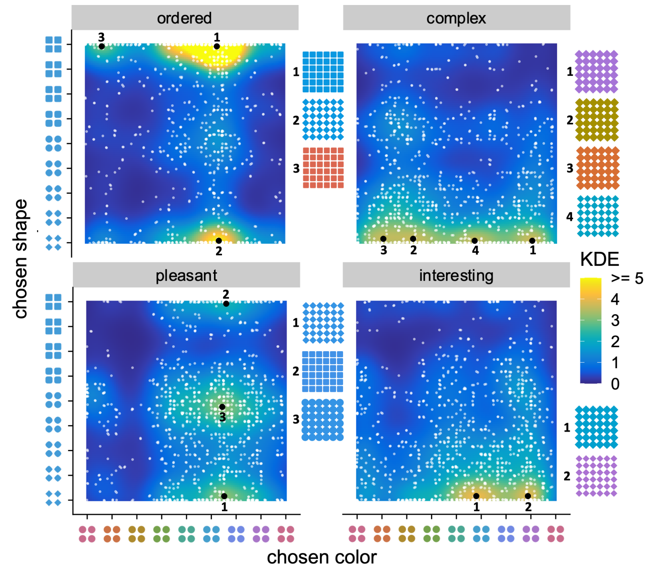

Figure 3: Two-dimensional KDEs for the color and shape choices in Experiment 1, with density expressed relative to a uniform distribution. The local maxima with a density higher than 2.5 are marked by black dots and visualized on the side.

Figure 3 shows the two-dimensional KDEs for the chosen colors and shapes in Experiment 1, separately for each evaluation criterion. Numerically, density is expressed relative to a uniform distribution, with values above one indicating that the feature value combination was more often chosen than expected under a uniform distribution. Although the KDEs for the four evaluations show some overlap, they also differ in clear and interpretable ways. For example, blue squares and blue diamonds were evaluated as particularly ordered, and all blue-colored familiar shapes (i.e., diamonds, circles, and squares) were often chosen as pleasant. For complexity and interest evaluations, diamond shapes were very dominant.

Figure 4 shows the one-dimensional KDEs for the chosen colors, shapes, and sizes in Experiment 1, which emphasize the peaks visible also in Figure 3. The stability of the one-dimensional local maxima (peaks) for each evaluation criterion was calculated using 1000 bootstraps (sampled with replacement). We report the peaks' mean density relative to a uniform distribution, with a bootstrapped 95% HDCI.

The peaks we find are interpetable and consistent with previous literature. In particular, on the color dimension (cf. Figure 4A), a maximum occurred for blue-colored shapes for order (M=2.50 [2.27, 2.71]) and pleasantness (M=1.87 [1.70, 2.07]), and a minimum occurred at the color between orange and green (M_{ordered}=0.44 [0.34, 0.57]; M_{pleasant}=0.23 [0.17, 0.30]). These results align with earlier work on hue prefer-

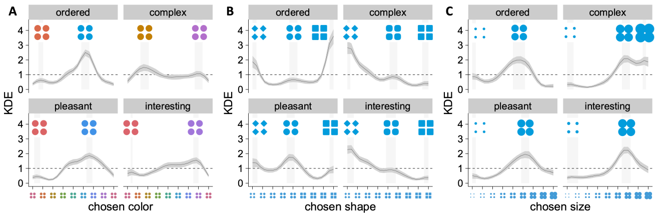

Figure 4: One-dimensional KDEs for (A) color, (B) shape, and (C) size in Experiment 1. Local maxima that appeared in $\geq$ 80% of all bootstraps are indicated by shaded areas.

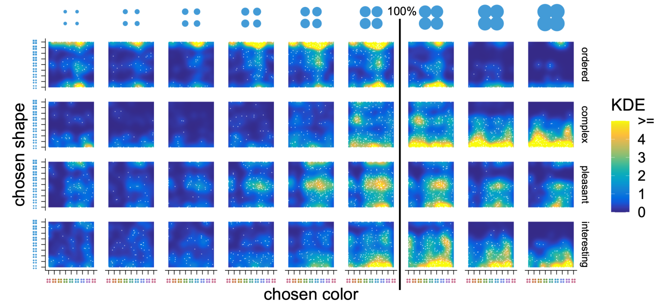

Figure 5: Three-dimensional KDEs for color, shape, and size in Experiment 1.

ences that showed a general preference for cool (green, cyan, blue) over warm colors (red, orange, yellow), with a maximum at blue and a minimum around yellow to yellow-green (Palmer & Schloss, 2010). This minimum for the color between orange and green also appeared for the interestingness evaluation (M=0.52 [0.41, 0.61]). In the complexity condition, this color resulted in a maximum (M=1.50 [1.30, 1.73]). For the evaluation of interestingness, most density occurred for purple-colored shapes (M=1.56 [1.37, 1.75]).

For both order and pleasantness, local maxima arose around the included familiar shapes (i.e., circles, squares, and diamonds; cf. Figure 4B), in line with earlier research that found preferences for object shapes to the extent that they conform to categorical prototypes (Rosch, 1975; Palmer et al., 2013; Martindale, Moore, & West, 1988). However, squares (M=3.61 [3.08, 4.08]) and diamonds (M=1.88 [1.48, 2.28]) showed much higher density for being ordered than circles (M=0.67 [0.51, 0.81], $ps<.0001$), whereas circles were relatively more probable when pleasantness was evaluated (M_{circles}=1.77 [1.54, 1.99]; M_{squares}=0.94 [0.68, 1.17], $p_{\text{diff}} <.0001$; M_{diamonds}=1.40 [1.08, 1.69], p_{diff}=.03). This preference for circles above diamonds and squares is in line with earlier work on preferences for objects with curved rather than sharp contours (e.g., Silvia & Barona, 2009; Bar & Neta, 2006). In case participants were optimizing for complexity or interestingness, the diamond shapes were preferred (M_{complex}=2.84 [2.41, 3.27]; $M_{\text{interesting}}$=2.31 [2.06, 2.59]).

When inspecting the one-dimensional peaks for size (cf. Figure 4C), we see similar distributions for order and pleasant, with the largest peak for *nearly* touching shapes (M_{ordered}=2.03 [1.80, 2.25]; M_{pleasant}=1.96 [1.73, 2.17]). For complexity and interest, a peak occurs for slightly overlapping shapes (M_{complex}=2.11 [1.87, 2.34]; $M_{\text{interesting}}$=2.25 [2.03, 2.48]), and for complexity there is an additional peak for strongly overlapping shapes (M=1.97 [1.70, 2.23]).

Figure 5 shows the three-dimensional KDEs for the chosen color, shape, and size combinations for each evaluation criterion in Experiment 1, which allows to see interactions between the dimensions. For example, in the results for order, there is an interaction visible between size and shape: smaller squares (size $\leq$ 100%) and touching diamonds (size $\approx$ 100%) were seen as most ordered. This could be related to a prototypical proportion of filled versus blank space (i.e., density in terms of how the figure looks like): for the same bounding box, squares take up much more space than diamonds. As a consequence, for the same size (in these Experiments defined based on the bounding box), squares will have a larger spatial density. A similar interaction occurs for pleasantness, where the distribution for circles peaks at a smaller size (around 100%) than the distribution for diamonds (above 100%). For complexity and interest, the highest peaks arose for shapes that fill more than 100% or approximately 100%.

Figure 7B shows the overall correlations between the three-dimensional probability distributions for the four evaluations. Similar to previous research with non-parametrically controlled stimuli (cf. Figure 7A; Van Geert & Wagemans, 2021), we found a strong positive Pearson's product-moment correlation between complexity and interest, and a positive correlation between order and pleasantness. In contrast to this earlier study, order did not have a strong positive relation with interest, and complexity showed a positive relation with pleasantness rather than a negative one. This difference in results could be due to the stimulus space we chose for Experiment 1: all stimuli in the space were relatively simple (low number of elements) and ordered (no deviations from the homogeneous pattern of colors, shapes, and sizes). In such a simple and ordered setting, complexity could get more room to be appreciated (in line with classical theories on optimal arousal level, or an optimal level of complexity; Berlyne, 1960, 1971; Van Geert & Wagemans, 2020). To assess to what extent the marginal probabilities for each of the feature dimensions (i.e., color, shape, size) contributed to the three-dimensional KDEs, we calculated the conditional Shannon entropy (Shannon, 1948) for each of the feature dimensions given that the values for the two other feature dimensions are known. The conditional entropies for color (4.06-4.29), shape (3.80-4.17), and size (3.80-4.09) were relatively equal, indicating that the dimensions contributed similarly to the overall entropy. In addition, we assessed the extent to which interactions between the feature dimensions contributed to the three-dimensional KDEs. The sum of the conditional entropies almost equalled the total entropy in the three-dimensional KDEs, indicating that the mutual information between the three feature dimensions (i.e., the total entropy minus the

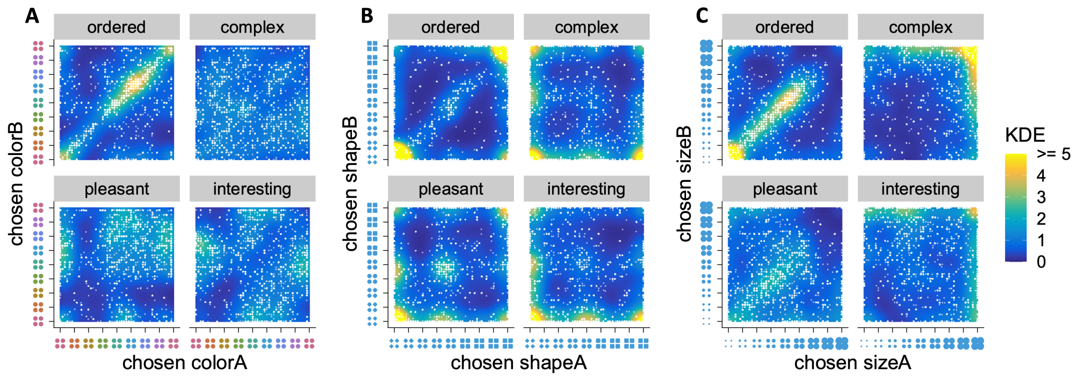

Figure 6: Two-dimensional KDEs for the chosen combination of **(A)** colors, **(B)** shapes, and **(C)** sizes for each evaluation criterion in Experiment 2, with density expressed relative to a uniform distribution.

three conditional entropies) contributed less than 2% to the overall entropy (1.01-1.53%). Given the small contribution of the interactions, this finding indicates that the feature dimensions acted as largely independent.

Experiment 2

Experiment 2 examined to what extent *pairs* of colors, shapes, and sizes are perceived as ordered, complex, pleasant, or interesting in 6-by-6 multi-element displays with different feature values for odd and even elements. Although participants could only manipulate the color, shape, or size of either the odd or even elements within a specific trial, from these choices we could also derive probability distributions for feature *difference*, which can be seen as a stimulus-based, "objective" measure of complexity.

We estimated the multidimensional probability distributions for each evaluation criterion (a) by fitting two-dimensional KDEs to the GSP samples for each feature separately; (b) by fitting three-dimensional KDEs to the GSP samples aggregated across the two dimensions per feature (e.g., representing choices for a color regardless of which other color was present) and thus affording a direct comparison with Experiment 1; and (c) by fitting three-dimensional KDEs to the GSP samples using the absolute difference between the two chosen values per feature (i.e., representing choices for a particular level of difference in color, shape, or size). Compared with Experiment 1, the 95% HDCIs for the split-half reliabilities in Experiment 2 were wider. Although the reliabilities were above .70 for most evaluations, the 95% HDCI did include values below .70 for order when looking at the absolute values (M=.77 [.66,.86]) and for complexity (M=.80 [.68,.89]), pleasantness (M=.68 [.47,.83]), and interest (M=.78 [.55,.91]) when looking at the difference values. To simplify the interpretation, we focus on the two-dimensional KDE results per feature, which are in direct relation to the six-dimensional probability distribution, and combine those with descriptions of the local maxima in the one-dimensional KDEs for both combined absolute values and difference values.

Figure 6 shows the two-dimensional KDEs for the chosen combination of colors, shapes, and sizes for each evaluation criterion in Experiment 2. Again, we see complex but interpretable structure. For order, the highest density was located around the diagonal for all three feature dimensions, indicating small differences in color, shape, and size to be evaluated as most ordered (M_{color}=2.05 [1.96, 2.14]; M_{shape}=2.04 [1.90, 2.18]; M_{size}=2.16 [2.01, 2.29]). However, for shapes there was a high density concentration around identical diamonds (M=2.49 [1.99, 3.00]) and identical squares (M=1.91 [1.56, 2.25]), and for sizes the density around the diagonal decreased once shapes start to overlap (i.e., around a size of +- 100% of the available element spacing). For absolute color values, we see the same local minimum for order as in Experiment 1, around the color between orange and green.

For the evaluation of complexity, overlapping shapes were most prominent, as were combinations of diamonds and squares. For pleasantness, a clear local minimum was present on the color dimension around the color between orange and green that was also perceived as unpleasant in Experiment 1. For shape, combinations of familiar shapes were most prominent for pleasantness. For pleasantness in the size domain, a local maximum arose for images with similarly-sized, non-overlapping odd and even elements. Color combinations chosen to be interesting were located off-diagonal, indicating a choice for a maximal difference in color between the odd and even elements. In the shape domain, combinations of familiar shapes including a diamond as one of the shapes were most prominent in the interesting evaluation. For size combinations, images including large size differences between the odd and even elements as well as images including large overlapping elements showed a local maximum for being interesting.

Figure 7 shows the overall correlations between the (C) combined absolute and (D) difference three-dimensional probability distributions in Experiment 2 for the four evaluations. Similar to previous research and Experiment 1, we found strong positive correlations between complexity and interest, order and pleasantness, and pleasantness and interest. When considering the absolute values, order did have a

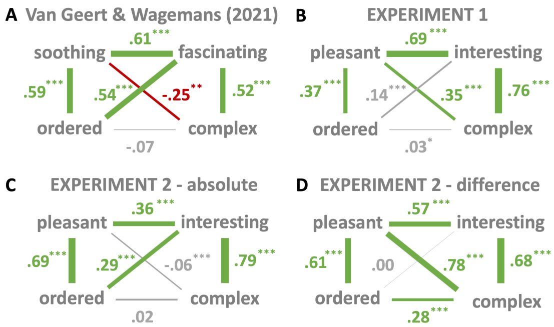

Figure 7: Correlations between order, complexity, and appreciation in (A) Van Geert and Wagemans (2021) and (B-D) the current experiments. For (B-D), correlations are calculated between the three-dimensional KDEs for each pair of evaluation criteria. *Note.* * $p < .01$, ** $p < .001$, *** $p < .0001$.

positive relation with interest, but this positive relation was absent when considering the differences between feature values. This may indicate for example that whereas absolute shapes that are evaluated as ordered were also perceived as interesting, shape differences that were perceived as ordered (i.e., small differences) were not necessarily perceived as interesting. In contrast to our expectations, complexity showed a positive relation with pleasantness rather than a negative one when looking at the difference values, and a non-significant one when looking at the combined absolute values. Also in this case, this difference in correlation could indicate that for example color differences that were evaluated as complex (i.e., large differences) were also often evaluated as highly pleasant, whereas absolute colors that were perceived as complex were not necessarily experienced as pleasant.

To assess to what extent interactions between features were important for the overall probability distributions, we calculated the conditional Shannon entropy for a feature combination given that the values for one of the four other dimensions was known. Given that the conditional entropies for the color combinations, shape combinations, and size combinations were all very similar to their respective overall entropies (Δ_{ordered} = 0.32-0.89%; Δ_{complex} = 0.22-0.64%; Δ_{pleasant} = 0.20-0.51%; $\Delta_{\text{interesting}}$ = 0.19-0.46% difference compared to total entropy), we can conclude that choices for each feature dimension (e.g., color, shape, or size) were made relatively independent of the two other feature dimensions. In addition, we assessed the extent to which each color, shape, or size combination could be predicted based on the average of the feature values in the display, or based on the signed or absolute difference between the two feature values. To do so, we calculated the conditional entropy of the full two-dimensional KDEs given the average, the signed difference, or the absolute difference and compared it to the total entropy for the two-dimensional distribution. Both *absolute* feature values (53.89-67.18%) as well as *differences* between absolute features (Δ_{signed}=48.71-61.93%; Δ_{abs}=38.41-50.56%) contributed in determining the probability distributions for each feature, with some overlap in explained entropy (14.23-24.32%). This confirms the idea that it is important to take both absolute and difference aspects of the stimulus into account when investigating appreciation.

Discussion

In this work, we demonstrate a new approach to efficiently characterize perceptual and aesthetic evaluations in multidimensional parametric stimulus space. To do so, we combine OCTA for the generation of multidimensional parametric stimuli and GSP to efficiently sample evaluations in the multidimensional space. This enabled a fine-grained, high-dimensional yet controlled characterization of the interrelations between different stimulus dimensions, perceived order and complexity, and appreciation.

In contrast to unidimensional or non-parametric approaches, our multidimensional parametric approach enabled us (a) to provide richer results and interpretations, (b) to assess the relative contributions of different stimulus aspects, and (c) to verify whether interactions between stimulus dimensions mattered. In these particular cases the different feature dimensions acted relatively independent. However, we found that both absolute values and difference values need to be taken into account when studying appreciation: a mere focus on either absolute values or objective complexity (here defined as the difference between the two absolute values present) could only partially explain the results.

Of course there are also some limitations to the work presented here. First, GSP is a powerful exploratory tool, but the assumptions underlying the efficient sampling (e.g., that trials are relatively independent and only weakly affected by previous stimuli) may induce biases. Therefore, findings need to be validated with confirmatory experiments that do not rely on generative sampling (Harrison et al., 2020). Second, we defined two specific, relatively simple and ordered stimulus spaces including a particular set of parametric stimulus dimensions. In future studies, we will apply the same approach to diverse stimulus spaces including more complex and unordered sets and different parametric stimulus dimensions. Third, the current experiments assumed general population preferences in a Western sample and did not take individual or cultural differences into account. In future studies, we will run within-participant GSP chains to assess individual preferences (Harrison et al., 2020) and compare multidimensional characterizations of perception and appreciation across cultures (Jacoby et al., 2024; Jakubowski, Polak, Rocamora, Jure, & Jacoby, 2022; McPherson et al., 2020).

Overall, this work provides a novel methodological approach for efficiently characterizing subjective evaluations in multidimensional visual stimulus space, thereby enabling new theoretical insights in the complex interplay between diverse stimulus dimensions in shaping our perceptual and aesthetic evaluations.

Acknowledgments

This work has been supported by a travel grant for a long stay abroad from the Research Foundation – Flanders (FWO) awarded to Eline Van Geert (Grant V432023N) and by long-term structural funding from the Flemish Government awarded to Johan Wagemans (METH/21/02). We thank Peter Harrison, Pol van Rijn, and Frank Höger for their support and advice when further developing PsyNet to match the projects' needs.

References

Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, *17*(8), 645–648. doi: 10.1111/j.1467-9280.2006.01759.x

Berlyne, D. (1960). *Conflict, arousal and curiosity*. McGraw-Hill.

Berlyne, D. (1971). *Aesthetics and psychobiology*. New York, NY: Appleton-Century-Crofts.

Braun, J., Amirshahi, S. A., Denzler, J., & Redies, C. (2013). Statistical image properties of print advertisements, visual artworks and images of architecture. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00808

Clark, J. H. (1924). The Ishihara Test for color blindness. *American Journal of Physiological Optics*, *5*, 269–276.

Gartus, A., & Leder, H. (2013). The small step toward asymmetry: Aesthetic judgment of broken symmetries. *i-Perception*, *4*(5), 361–364. doi: 10.1068/i0588sas

Gielis, J. (2003). A generic geometric transformation that unifies a wide range of natural and abstract shapes. *American Journal of Botany*, *90*(3), 333–338. doi: 10.3732/ajb.90.3.333

Graf, L. K. M., & Landwehr, J. R. (2017). Aesthetic pleasure versus aesthetic interest: The two routes to aesthetic liking. *Frontiers in Psychology*, *8*. doi: 10.3389/fpsyg.2017.00015

Harrison, P., Marjieh, R., Adolfi, F., van Rijn, P., Anglada-Tort, M., Tchernichovski, O., … Jacoby, N. (2020). Gibbs sampling with people. In *Advances in Neural Information Processing Systems* (Vol. 33). doi: 10.48550/arXiv.2008.02595

Jacobsen, T., & Höfel, L. (2002). Aesthetic judgments of novel graphic patterns: Analyses of individual judgments. *Perceptual and Motor Skills*, *95*(3), 755-766. doi: 10.2466/pms.2002.95.3.755

Jacoby, N., Polak, R., Grahn, J. A., Cameron, D. J., Lee, K. M., Godoy, R., … others (2024). Commonality and variation in mental representations of music revealed by a cross-cultural comparison of rhythm priors in 15 countries. *Nature Human Behaviour*, 1–32.

Jakubowski, K., Polak, R., Rocamora, M., Jure, L., & Jacoby, N. (2022). Aesthetics of musical timing: Culture and expertise affect preferences for isochrony but not synchrony. *Cognition*, *227*, 105205.

Kumar, S., Correa, C. G., Dasgupta, I., Marjieh, R., Hu, M., Hawkins, R. D., … Griffiths, T. L. (2022). Using natural language and program abstractions to instill human inductive biases in machines. In A. H. Oh, A. Agarwal, D. Belgrave, & K. Cho (Eds.), *Advances in Neural Information Processing Systems.*

Martin, J. B., Griffiths, T. L., & Sanborn, A. N. (2012). Testing the efficiency of Markov chain Monte Carlo with people using facial affect categories. *Cognitive Science*, *36*(1), 150–162.

Martindale, C., Moore, K., & West, A. (1988). Relationship of preference judgments to typicality, novelty, and mere exposure. *Empirical Studies of the Arts*, *6*(1), 79–96. doi: 10.2190/MCAJ-0GQT-DJTL-LNQD

McPherson, M. J., Dolan, S. E., Durango, A., Ossandon, T., Valdés, J., Undurraga, E. A., … McDermott, J. H. (2020). Perceptual fusion of musical notes by native amazonians suggests universal representations of musical intervals. *Nature communications*, *11*(1), 2786.

Neuberg, S. L., & Newsom, J. T. (1993). Personal need for structure: Individual differences in the desire for simple structure. *Journal of Personality and Social Psychology*, *65*(1), 113–131. doi: 10.1037/0022-3514.65.1.113

Ottosson, B. (2020). *A perceptual color space for image processing.* https://bottosson.github.io/posts/oklab/.

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, *107*(19), 8877-8882. doi: 10.1073/pnas.0906172107

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology*, *64*(1), 77–107. doi: 10.1146/annurev-psych-120710-100504

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192.

Sanborn, A. N., Griffiths, T. L., & Shiffrin, R. M. (2010). Uncovering mental representations with Markov chain Monte Carlo. *Cognitive Psychology*, *60*(2), 63–106.

Schloss, K. B., & Palmer, S. E. (2011, September). The Role of Spatial Organization in Preference for Color Pairs. *Perception*, *40*(9), 1063–1080. doi: 10.1068/p6992

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.

Silvia, P. J., & Barona, C. M. (2009). Do people prefer curved objects? Angularity, expertise, and aesthetic preference. *Empirical Studies of the Arts*, *27*(1), 25–42. doi: 10.2190/EM.27.1.b

Smart, S., & Szafir, D. A. (2019, May). Measuring the Separability of Shape, Size, and Color in Scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). Glasgow Scotland Uk: ACM. doi: 10.1145/3290605.3300899

Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H. (2020). The Vienna Art Interest and Art Knowledge questionnaire (VAIAK): A unified and validated measure of art interest and art knowledge. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(2), 172 – 185. doi: 10.1037/aca0000205

Spehar, B., Walker, N., & Taylor, R. P. (2016). Taxonomy of individual variations in aesthetic responses to fractal patterns. *Frontiers in Human Neuroscience*, *10*. doi: 10.3389/fnhum.2016.00350

Sun, Z., & Firestone, C. (2022). Beautiful on the inside: Aesthetic preferences and the skeletal complexity of shapes. *Perception*, *51*(12), 904-918. doi: 10.1177/03010066221124872

Van Geert, E., Bossens, C., & Wagemans, J. (2023). The Order & Complexity Toolbox for Aesthetics (OCTA): A systematic approach to study the relations between order, complexity, and aesthetic appreciation. *Behavior Research Methods*, *55*, 2423–2446. doi: 10.3758/s13428-022-01900-w

Van Geert, E., Hofmann, D., & Wagemans, J. (2024). *The perception and appreciation of order and complexity.*

Van Geert, E., & Wagemans, J. (2020). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(2), 135–154. doi: 10.1037/aca0000224

Van Geert, E., & Wagemans, J. (2021). Order, complexity, and aesthetic preferences for neatly organized compositions. *Psychology of Aesthetics, Creativity, and the Arts*, *15*(3), 484–504. doi: 10.1037/aca0000276

Van Geert, E., Warny, A., & Wagemans, J. (2024). *A systematic approach to study preferences for complexity at different levels of order.*

van Rijn, P., Mertes, S., Janowski, K., Weitz, K., Jacoby, N., & André, E. (2024). Giving robots a voice: Human-in-the-loop voice creation and open-ended labeling. *arXiv preprint arXiv:2402.05206*.

van Rijn, P., Mertes, S., Schiller, D., Dura, P., Siuzdak, H., Harrison, P., ... Jacoby, N. (2022). VoiceMe: Personalized voice generation in TTS. *arXiv*. doi: 10.48550/arXiv.2203.15379

van Rijn, P., Mertes, S., Schiller, D., Harrison, P. M., Larrouy-Maestri, P., André, E., & Jacoby, N. (2021). Exploring emotional prototypes in a high dimensional TTS latent space. In *Proc. Interspeech 2021.* doi: 10.21437/Interspeech.2021-1538

Wilson, A., & Chatterjee, A. (2005). The assessment of preference for balance: Introducing a new test. *Empirical Studies of the Arts*, *23*(2), 165–180. doi: 10.2190/B1LR-MVF3-F36X-XR64