On visual texture preference: Can an ecological model explain why people like some textures

more than others?

Kyle D. Stephens & Donald D. Hoffman

University of California, Irvine

Author Note

All authors, Department of Cognitive Sciences, University of California, Irvine. Correspondence concerning this article should be addressed to Kyle D. Stephens, 493 Social Science Labs, University of California, Irvine, CA 92697. E-mail: kdstephe@uci.edu

Abstract

What visual textures do people like and why? Here we test whether Palmer and Schloss' (2010) *ecological valence theory* can predict people's preferences for visual texture. According to the theory, people should like visual textures associated with positive objects or entities and dislike visual textures associated with negative objects or entities. We compare the results for the ecological model to a more traditional texture-preference model based on computational features – namely, the model of Thumfart et al. (2011) – and find that the ecological model performs reasonably well considering its lower complexity, explaining 63% of the variance in the human preference data.

Keywords: visual texture, aesthetic preference, ecological valence theory

On visual texture preference: Can an ecological model explain why people like some textures more than others?

Introduction

For members of the species *H. sapiens,* textures are ubiquitous. Every day we are surrounded by objects and surfaces that give us unique feelings when we touch them – fur feels soft, tree bark feels rough, and silk feels smooth. These different feelings arise from variations on the surfaces of objects which also lead to variations in the pattern of light that reaches our eyes – fur *looks* soft, tree bark *looks* rough, and silk *looks* smooth. We call this visual aspect of surface variation *visual texture*[1]. Artists and designers have long used visual texture as a tool to evoke emotions and set moods (see, e.g., Brodatz, 1966; Gatto, Porter, & Selleck, 1999). In the current paper, we investigate the aesthetics of visual texture more formally by empirically exploring the question: What visual textures do people like and why? We examine whether the ecological valence theory proposed by Palmer and Schloss (2010) for color preferences – and later extended to odor preferences (Schloss, Goldberger, Palmer, & Levitan, 2015) – can also be used to explain preferences for visual texture. But, first we review previous work on visual texture and the aesthetics of images.

There are many models that predict perceptual properties of visual textures. For example, there are models that predict how rough (Ho, Landy, & Maloney, 2006; Tamura, Mori, & Yamawaki, 1978), glossy (Anderson & Kim, 2009; Kim & Anderson, 2010; Montoyishi, Nishida, Sharan, & Adelson, 2007), or complex (Amadasun & King, 1989; Tamura et al., 1978)

---

[1] More formally, if we think of an image as a 2-D array of pixels and we think of each pixel as a realization of a random variable on intensity, then visual textures are those images that have variation in pixel intensity at the limit of resolution but whose spatial covariance is relatively homogenous (i.e., moving a small window around the image does not significantly alter the statistics within the window; Haindl & Filip, 2013; Julesz, 1962; Portilla & Simoncelli, 2000), although other constraints are often employed and precise definitions vary by application (Haindl & Filip, 2013; Sebe & Lew, 2001; Tuceryan & Jain, 1998). Also note that visual texture can arise from variations that are not tactile (e.g., a birch table looks different from a mahogany table but they might feel identical).

they look, as well as how they are segmented (Ben-Shahar, 2006; Julesz, 1962, 1981; Landy & Bergen, 1991; Malik & Perona, 1990; Portilla & Simoncelli, 2000; Rosenholtz, 2000; Tyler, 2004; Victor, 1988) or classified (e.g., Dong, Tao, Li, Ma, & Pu, 2015; Guo, Zhang, & Zhang, 2010; Haralik, Shanmugam, & Dinstein, 1973; Randen & Husy, 1999; Varma & Zisserman, 2005) by human observers (for reviews, see Bergen, 1991; Landy & Graham, 2004; Rosenholtz, 2015).

There are also many models that predict aesthetics of images, including how much observers like the images (e.g., Chen, Sobue, & Huang, 2008; Kawamoto & Soen, 1993; S. Kim, E. Kim, Jeong, & J. Kim, 2006; Lee & Park, 2011; Um, Eum & Lee, 2002; Zhang et al., 2011; for reviews, see Joshi et al., 2011; Palmer, Schloss, & Sammartino, 2013). But these models are typically applications-focused – they were not designed to predict preferences for visual textures generally, nor do they necessarily extend well to the general space of visual textures (Thumfart et al., 2011).

There is, however, an extant model that predicts preferences for visual textures specifically, the architecture for which has been proposed by two different research groups.

Current Computational Model of Visual Texture Aesthetics

Thumfart et al. (2011) and Liu, Lughofer, and Zeng (2015) have both recently published a parametric model that predicts human ratings on several aesthetic properties[2] for visual textures (Figure 1). Given a visual texture, this model makes predictions using a combination of low-level computational features measured from that texture and human ratings on other aesthetic properties for that texture. For instance, Thumfart et al. found that how "elegant" a texture looks was best-predicted by a combination of eight computational features measured from that texture

---

[2] "Aesthetic properties" are operationalized as antonym pairs such as Warm-Cold, Hard-Soft, or Artificial-Natural. Thus, preference is defined by the antonym pair Dislike-Like.

(top three: the response energy of a particular oriented Gabor filter applied to the image, the contribution of blue to the image color, and the mean variance of image intensities) and human ratings on four other aesthetic properties for that texture (ratings of how "glossy," "silky," "woodlike," and "grainy" it looks). In general, the computational features are derived from well-known image-processing features – like the Fourier power spectrum of an image or the gay-level co-occurrence matrix for an image – and the aesthetic properties include primary/physical descriptions (e.g., Cold-Warm, Rough-Smooth), more abstract descriptions (e.g., Simple-Complex, Inelegant-Elegant), and preference (Dislike-Like).

This model is hierarchical in that some aesthetic properties are predicted using only computational features, whereas others are predicted using a combination of computational features and other aesthetic properties (see Figure 1). Both Thumfart et al. and Liu et al. consider preference (i.e., ratings of Dislike-Like) to be the highest-level aesthetic property and thus their models predict preference using both computational features and human ratings on all other aesthetic properties.[3] For instance, Thumfart et al. predicted preference using linear regression with 214 potential regressors: 188 computational features and ratings for 26 other aesthetic properties (see Figure 1).

---

[3] Because of the hierarchical feed-forward nature of these models, preference predictions could eventually be reduced to only computational features. For example, if ratings of Cold-Warm turned out to be important for predicting preference, then, in future iterations of the model, human ratings of Cold-Warm could be replaced by measurements of the computational features that best predict ratings of Cold-Warm (see Figure 1).
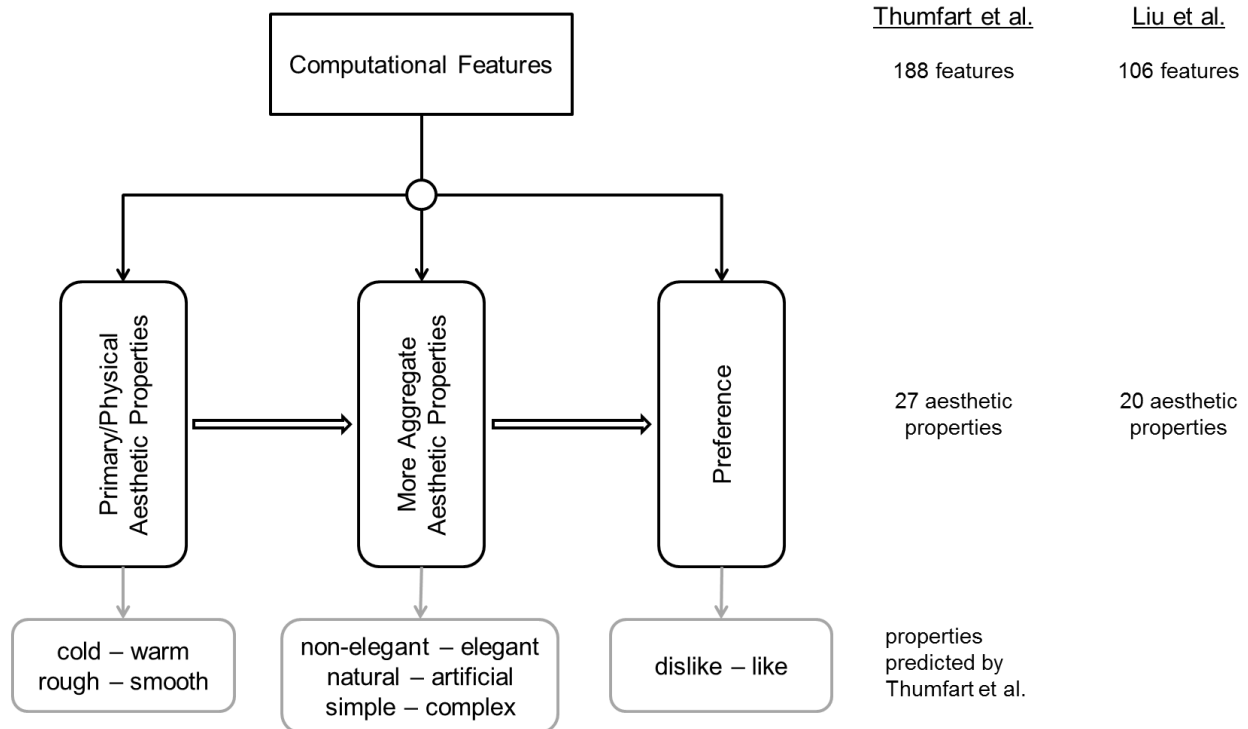
Figure 1. Structure of the hierarchical feed-forward model of texture aesthetics used by Thumfart et al. (2011) and Liu et al. (2015) (adapted from Thumfart et al.'s representation of the model). The model predicts human ratings of aesthetic properties using measurements of computational features and ratings of other aesthetic properties. In the current paper, we are interested in preference, or Dislike-Like.

This model has been able to predict human preference (i.e., Dislike-Like) for visual texture with high accuracy (in Liu et al.'s instantiation of the model, they claim to account for as much as 99% of the variance in human preference ratings). However, the model employs a "see what sticks" approach which emphasizes model prediction over model understanding. The numerous features and aesthetic properties constitute a veritable grab-bag from the literature on computer vison, image processing and aesthetics. But, to the credit of Thumfart et al. and Liu et al., their models are more interpretable than black-box models such as neural networks, and Thumfart et al. increased interpretability by using an optimization measure that punishes complexity (defined by number of regression terms).

Still, these models were not constrained by any overarching theory dictating which features should be included and which should not, and it's unclear why certain features are so important for predicting preference while others are not. For instance, Thumfart et al. found that how "premium", "sophisticated" and "woodlike" a texture looks were all important for predicting how much someone likes that texture, whereas how "elegant" and "natural" it looks were not. Likewise, they found that the skewness of the energy contained in 12 circular ring segments of the Fourier power spectrum of a texture was important for predicting how much someone likes that texture, whereas the kurtosis or standard deviation of the energy contained in those same 12 ring segments was not (see Table 1). While these results are interesting and useful, they are not theoretically motivated and would not have been predicted by any current theory.

In other words, this model is good at predicting *what* visual textures people like, but does not give the best sense of *why*.

Ecological Valence Theory

In contrast to the "see what sticks" computational model described above, Palmer and Schloss' (2010) ecological valence theory (EVT) is based on a simple premise: People's preferences for basic stimuli (e.g., colors, odors) are derived from how they feel about the objects or entities associated with those stimuli (note that the theory was initially only proposed for colors and only later extended to other types of stimuli/modalities). So, for example, a student might have a particular affinity for the color blue because they like blueberries and/or attend the University of California, Irvine (whose official color is blue), whereas another student might have an affinity for the color red because they like cherries and/or attend the University of Southern California (whose official color is red). Indeed, there is evidence that people like the

colors of their own social group more than members of a rival social group (Schloss & Palmer, 2014; Schloss, Poggesi & Palmer, 2011).

But, Palmer and Schloss take this basic premise a step further and claim that people's preferences are dictated by a summary statistic that accounts for the valence across *all* objects associated with a particular stimulus. That is, how much a person might like a given stimulus can be explained by how positive or negative people feel, on average, about all objects and entities associated with that stimulus. The motivation for this reasoning is evolutionary: Assuming that people have positive feelings about objects that lead to beneficial outcomes, then stimuli that are typically associated with those objects serve as indicators to the beneficial outcomes. Thus, such preferences should steer organisms to approach objects that lead to beneficial outcomes and avoid those that lead to harmful outcomes, an evolutionarily advantageous strategy (Palmer & Schloss, 2010; Schloss et al., 2015). This is the ecological valence theory.

Palmer, Schloss and colleagues have already found support for this theory for colors and odors: Average color preferences are well predicted by how people feel about all objects associated with the colors (Palmer & Schloss, 2010; Taylor & Franklin, 2012), and average odor preferences are well predicted by how people feel about all objects associated with the odors (Schloss et al., 2015). Furthermore, the theory predicts that preferences across cultures or social groups should vary only to the degree that these groups associate different objects with the same stimuli or have different feelings about the same objects, and there is empirical support that this is the case with colors (Schloss, Hawthorne-Madell, & Palmer, in press; Yokosawa, Yano, Schloss, Prado-Leon, & Palmer, 2010).

Thus far, Palmer, Schloss and colleagues have only tested the ecological valence theory with colors and odors, but the theory should be generally applicable to different kinds of stimuli in any modality (see, e.g., Schloss et al., 2015). The goal of the present research was to test how well the ecological valence theory accounts for preferences for visual texture: Can people's average preferences for visual textures be predicted by how positive or negative they feel about all objects associated with the textures?

Present Study

Testing the ecological valence hypothesis. To test whether preference for a given visual texture can be predicted by the average valence for all objects or entities associated with that visual texture, we adapted the procedure of Palmer and Schloss (2010) to visual textures. This procedure required data from four different groups of participants.

The first group gave the to-be-predicted preference ratings. These participants were simply asked to rate how much they liked each visual texture on a line-mark ratings scale coded from -100 (dislike) to $+100$ (like).

The next three groups of participants were needed to calculate the predicted preference[4] for each texture: one group typed out descriptions of objects they associated with each visual texture (the object-description group); another group rated their emotional response (positive/negative) to a condensed set of these object descriptions (the description-valence group); and the final group rated how well each visual texture matched all of the descriptions ascribed to that texture (the texture-description match group). The prediction for any given visual texture was then given by the sum of the average valence ratings for each object associated with the texture, weighted by how well those objects match the texture:

---

[4] Palmer and Schloss (2010) and Schloss et al. (2015) refer to their preference predictions as weighted affective valence estimates, or WAVEs.

$$P_t = \frac{1}{n_t} \sum_{d \in D_t} w_{td} v_d \tag{1}$$

where, P_t is the predicted preference for texture t (-100 to 100), n_t is the number of descriptions ascribed to texture t, D_t is the set of all descriptions ascribed to texture t, w_{td} is the average match weighting between texture t and description d (0 to 1), and v_d is the average valence rating for description d (-100 to 100). Using this formulation, Palmer and Schloss (2010) accounted for 80% of the variance in average color preferences, and Schloss et al. (2015) account for 76% of the variance in average odor preferences.

Testing other hypotheses. One might (reasonably) question whether applying this formulation to visual textures is sensible. After all, isn't it generally obvious what object a (natural) visual texture corresponds to? In this case, preference for a visual texture would reduce to how people feel about the single associated object, an uninteresting result. Schloss et al. (2015) faced a similar problem when applying the ecological valence theory to odors, since they used odor pens that were designed to smell like a particular object (e.g., *honey, apple, leather*). But, they found that the associated objects were more ambiguous than one would think and that odor preference was better explained by the valence for all objects associated with the odor pens, rather than just the object the pen was designed to smell like. We hypothesized a similar outcome for visual textures.

To test this outcome, we examined two additional hypotheses (following Schloss et al., 2015): the *single-associate hypothesis* and the *namesake hypothesis*. The single-associate hypothesis states that preference for a visual texture is best predicted by the valence of the single object that the texture is most associated with. The namesake hypothesis states that preference for a visual texture is best predicted by the valence of the namesake object that produced the

visual texture. So, for example, one of the visual textures used was a close-up of a strawberry (see Figure 2). The ecological valence theory predicts that preference for this texture is best explained by the weighted average valence ratings for all descriptions ascribed to this texture (e.g., 'beans', 'flowers', 'eggs', 'peacock tail', 'seeds', 'strawberry', 'insect eyes', etc.); the single-associate hypothesis predicts that preference for this texture is best explained by the average valence of the description 'eggs,' since this was the description most often associated with the texture; and the namesake hypothesis predicts that preference for this texture is best explained by the average valence of the description 'strawberry,' since this is the namesake object that the texture was produced from.

Comparing to the computational models. After testing the single-associate and namesake hypotheses, we also compared the results of the ecological valence model to the computational model of Thumfart et al. (2011). First, we compared the error of the ecological model to the error of Thumfart et al.'s model using an error term that punishes the model for complexity (see Equation 2 and Appendix B). Then, we tested Thumfart et al.'s model more directly.

Thumfart et al. found that preference for a given visual texture was best predicted by a linear equation with six factors: (1) how "premium" it looks, (2) how "sophisticated" it looks, (3) how "rough" it looks[5], (4) how "woodlike" it looks, (5) the skewness of the distribution of Fourier energy in concentric rings in the Fourier power spectrum of the texture (*cSkew*), and (6) the *strength* of the texture, a computational feature derived from the neighborhood graytone difference matrix for the texture. Thus, we had participants rate textures on line-mark rating

---

[5] Thumfart et al. actually used a computational feature termed *roughness* by Tamura et al. (1978). However, Tamura et al. designed the feature *roughness* to match human ratings of Rough-Smooth as closely as possible. Furthermore, we found that using the feature *roughness* in lieu of ratings of Rough-Smooth did not change our results (see Results and Discussion section).

scales for each of the antonym pairs described above (NotPremium-Premium, NotSophisticated-Sophisticated, Rough-Smooth, and NotWood-Wood), and we calculated their *cSkew* and their *strength*. We recruited two groups of participants: one group rated the textures on the scales for premium, sophisticated, and rough, and the other group rated the textures on the scale for woodlike.

Review. Table 1 gives an overview of the different models and hypotheses tested in this paper.

Table 1. Tested Models and Hypotheses

Model/Hypothesis	Preference Prediction Equation	Variables Explained
Full EVT model	$P_t = \dfrac{1}{n_t} \displaystyle\sum_{d \in D_t} w_{td} v_d$	P_t: predicted preference for texture t. n_t: number of descriptions ascribed to texture t. D_t: set of all descriptions ascribed to texture t. w_{td}: average match weight between texture t and description d. v_d: average valence rating for description d.
Single-Associate hypothesis (restricted EVT model)	$P_t = v_{d'(t)}$	$v_{d'(t)}$: average valence rating for description $d'(t)$. $d'(t)$: description most frequently ascribed to texture t.
Namesake hypothesis (restricted EVT model)	$P_t = v_{d^*(t)}$	$v_{d^*(t)}$: average valence rating for description $d^*(t)$. $d^*(t)$: description of what t is actually a texture of.
Thumfart et al. model	$P_t \propto$ "premium" + "sophisticated" + "rough" + $cSkew$ + "wood" + $strength$	"premium", "sophisticated", "rough", "wood": aesthetic properties gathered for texture t via human ratings. $cSkew$, $strength$: computational features derived from the Fourier power spectrum and the neighborhood gray tone difference matrix for texture t respectively.

Materials and Methods

Participants

There were six separate groups of participants, all of whom were undergraduate students at the University of California, Irvine and received course credit for participation. No participant

ever participated in more than one experimental task (see Table 2 for an overview of the six groups of participants and their different tasks).

36 participants (19 male) took part in the texture-preference rating task, rating how much they liked each visual texture. The mean age was 20.8 years (range 18-28 years).

32 participants (15 male) took part in the object-description task, describing objects associated with each visual texture. Once all of the descriptions were consolidated (see Procedure section), one of the authors (K. Stephens) rated how well he thought the descriptions matched the visual textures they were ascribed to on a 0-100 line-mark scale. We ended up excluding one female participant because 38 of her descriptions were given a 0 match-rating by this author (this was 4.5 standard deviations above the mean number of 0 match-ratings for the rest of the group). After excluding this participant, the mean age was 21.6 years (range 18-37 years).

37 participants (18 males) took part in the description-valence rating task, rating how positive or negative they felt about each object description. The mean age was 20.7 years (range 18-32 years).

27 (14 male) participants took part in the texture-description match rating task, rating how well each description matched the texture it was ascribed to. We excluded one male participant because he did not finish the experiment (we lost connection to our Matlab license three-quarters of the way through). After excluding this participant, the mean age was 22.5 years (range 18-40 years).

24 (11 male) participants took part in the first aesthetic-property rating task, rating how "premium", "sophisticated", and "rough" the textures looked. The mean age was 21.1 years (range 18-36 years).

Finally, 22 (6 male) participants took part in the second aesthetic-property rating task, rating how "woodlike" the textures looked. We excluded one female participant because her mean reaction time was more than 2 standard deviations faster than the mean for the group. After excluding this participant, the mean age was 20.7 years (range 18-34 years).

Stimuli

Texture-preference rating, object description, and aesthetic-property rating. For these tasks the stimuli consisted of 62 different visual textures. 52 of these textures were from Brodatz's (1966) album, and 10 were from shutterstock.com. We used the same Brodatz textures as those used by Rao and Lohse (1993, 1996), excluding D43, D75, D102, and D108 (see tables in the Appendix for a complete list). Of the Shutterstock images, eight were chosen specifically to have either high valence (chocolate, fresh lettuce, fresh strawberry, sea/ocean) or low valence (mud, mold, rotten strawberry, infected skin); the remaining two images were Gaussian blurred versions of the chocolate and the mud images respectively. All images were grayscale and viewed at 256-by-256 pixels (see Figure 2 for the Shutterstock images; see tables in the appendix for a complete list of all visual textures. See Brodatz, 1966 or Rao & Lohse, 1993 for pictures of the Brodatz textures).

Texture-description match rating. Because of the large number of descriptions for each texture (even after consolidation; see Procedure section), it would have taken too long to show participants all texture-description pairs for all 62 textures in the texture-description match rating task. Thus, we only included 47 of the original 62 textures in this task. We included all 10 of the shutterstock.com images, but only 37 of the original 52 Brodatz textures. We chose which 15 Brodatz textures to exclude (D11, D25, D29, D31, D39, D40, D42, D55, D63, D80, D82, D83,

D89, D94, D101) based on their similarity to other textures already included (see table A1 in the

appendix for a complete list of all visual textures used in this experiment).

Description-valence rating. For the description-valence rating task, the stimuli consisted

of black text presented in 28-point Arial font.
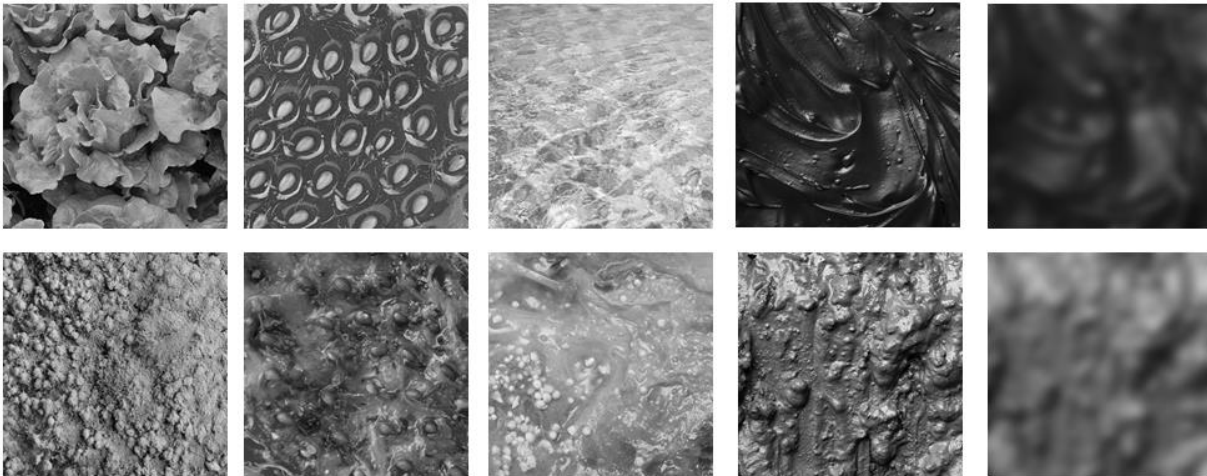
Figure 2. The ten visual textures from Shutterstock used for all tasks. The rightmost pictures are Gaussian-blurred versions of the preceding pictures. From top left to bottom right: lettuce, strawberry, tropical ocean, chocolate, Gaussian-blurred chocolate, mold, rotted strawberry, skin infection, mud, Gaussian-blurred mud.

Displays

Monitor for visual textures. For any task displaying visual textures (i.e., the texture-

preference rating, object description, texture-object match rating, and aesthetic-property rating

tasks), stimuli were presented using an 27-inch Apple iMac display set at a resolution of

2560x1440 and a refresh rate of 60Hz. All stimuli were presented against a neutral gray

background with luminance 28.8 cd/m^2 (as measured using Photo Research PR-670

spectroradiometer). The monitor was calibrated using an X-Rite ColorMunki Photo

spectrophotometer. Participants were all run in the same room, one at a time.

Monitor for descriptions. For the task displaying only text (i.e., the description-valence rating task), stimuli were presented using Dell computers attached to 17-inch Dell LCD monitors set at a resolution of 1280x1024 and a refresh rate of 60Hz. The text stimuli were displayed against a white background. Participants were run in groups of one to five in a room containing 8 computers.

Line-mark rating scale. Five of the tasks – the texture-preference rating, description-valence rating, texture-description matching rating, and aesthetic-property rating tasks – gathered ratings using a line-mark rating scale, located at the bottom of the display. This scale consisted of a dark-gray rectangular box demarcated into 10 equally spaced regions (5 below the center point and 5 above), with text above the scale indicating what rating was being performed (e.g., "How much do you like this texture?"), text at the end-points of the scale to indicate the extreme ratings (e.g., "not at all" vs. "very much"), and a black slider bar that participants could slide along the scale with the mouse to indicate their rating. Participants could either click and drag the slider bar or move the bar by pointing and clicking (i.e., without dragging). When displayed on the larger monitor (i.e., for the texture-preference rating, texture-object match rating, and aesthetic-property rating tasks), the top edge of the question text was approximately 247 pixels below the bottom edge of the texture, the rating scale was 768 pixels wide by 72 pixels high, and the scale's top edge was approximately 41 pixels below the bottom edge of the question text. When displayed on the smaller monitor (i.e., for the description-preference rating task), the rating scale was made smaller – 384 pixels wide by 52 pixels high – in proportion to the smaller resolution of the screen. The black slider bar was always 6 pixels wide by the height of the scale (either 72 or 52 pixels), and all text was black, in 24-point Arial font. The scale and the question text were always horizontally centered. The code for all tasks using a line-mark rating scale was written in

Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007).

Description edit boxes. The object-description task gathered descriptions using edit boxes located below the textures. There were five edit boxes in total, where participants could click and enter text. The edit boxes were 200 pixels wide by 50 pixels high and were located approximately 137 pixels below the bottom edge of the texture. They were horizontally centered and evenly spaced approximately 123 pixels apart from one another. Below the edit boxes was a pushbutton labeled 'submit' where participants could click to input their responses and move on to the next trial. This pushbutton was 150 pixels wide by 50 pixels high, horizontally centered, and the top edge was approximately 96 pixels below the bottom edge of the edit boxes. The code for this task was written using Matlab's uicontrol functions.

Procedure

The present study included five between-subject tasks: (i) texture-preference rating, (ii) object description, (iii) description-valence rating, (iv) texture-description match rating, and (v) aesthetic-property rating (see Table 2). For all tasks, there was a blank 500ms ISI in between trials, and participants sat approximately 60cm away from the monitor. At this viewing distance, the visual textures spanned approximately 5.7 degrees of visual angle.

Texture-preference rating task. On each trial, participants viewed one of the 62 visual textures, vertically and horizontally centered on a neutral gray background with the text "How much do you like this texture?" displayed below the texture. Participants indicated their preference using a line-mark rating scale labeled "not at all" on the left end (coded as -100) and "very much" on the right end (coded as +100). Participants used the mouse to indicate their

preference. They pressed the spacebar to input their rating and move on to the next texture. The textures were displayed in random order and all participants gave ratings for all textures.

Object description task. Participants saw each texture individually against a neutral gray background and were instructed to type out as many objects as they could think of (up to five) that might describe or be associated with the given visual texture. They were instructed not to name adjectives (e.g., 'heavy,' 'light,' 'furry'), abstract entities (e.g., 'dreams,' 'hopes,' 'fears'), or objects that would not be known to other people (e.g., 'my aunt's blanket'). They were also told that the experimenters were interested in all associated objects, whether pleasant or unpleasant. They typed their responses into one of five edit boxes located below the texture on the screen and clicked a pushbutton labeled 'submit' to move on to the next texture once they were satisfied with their descriptions. They were not given a time-limit. The textures were displayed in random order and all participants gave descriptions for all 62 textures.

After combining exact repeats, we were left with 1,264 unique descriptions (from an initial total of 4,858). We compiled these descriptions into a single list of 206 items using a procedure similar to that of Palmer and Schloss (2010). First, we discarded items from the list if they: (i) were abstract concepts instead of objects (e.g., 'anger', 'dream', 'joy'); (ii) were adjectives instead of objects (e.g., 'big', 'bumpy', 'blurry', 'curvy', 'dark', 'heavy'); or (iii) were ambiguous or could describe many different things (e.g., 'arrangement', 'game', 'layers', 'material', 'mixture', 'surface', 'painting'). We excluded 366 descriptions this way, leaving 898 items on the list. We then categorized the descriptions to reduce the number that needed to be rated in the description-preference rating and the texture-object match rating tasks. We combined descriptions into a single object category if they seemed to refer to the same object (e.g., 'blanket', 'quilt' and 'piece of a blanket'; 'beehive', 'bee's nest' and 'bee cells'; 'carpet',

'carpret [sic.]' and 'rug') and, in certain cases, we combined descriptions into superordinate

categories if there were several exemplars referring to the same type of object. In the latter case,

exemplars were included in the category description (e.g., 'bedding (covers, sheets, pillows)';

'bricks (brick wall, brick building, brick sidewalk)'; 'reptile skin (snake, crocodile, lizard)').

Descriptions were further excluded if they were given by only one person for only one image and

they didn't fit into any other object categories. After this process, we ended up with the final list

of 206 consolidated object descriptions.

Even after this consolidation, each texture still had, on average, 20.4 descriptions

attributed to it. With this many texture-description pairings, the texture-description matching task

would have been too long (even after pairing the textures down from 62 to 47), so two

independent raters ran a preliminary version of the texture-description match rating task and we

excluded any texture-descriptions pairings that were given by only one participant and assigned a

0 match rating by both independent raters. After this procedure, the textures had an average of

18.4 descriptions each.

Description-valence rating task. Following Palmer and Schloss (2010), participants

were first given a list of 8 sample items (sunset, bananas, diarrhea, sidewalk, boogers/snot, wine

(red), chalkboard, chocolate) to give them an idea of the range of descriptions they would see.

They were then presented with each of the 206 consolidated descriptions in black text,

horizontally and vertically centered against a white background. The procedure was the same as

for the texture-preference rating task, but the instruction text was different. Below each

description was the text, "What's your emotional reaction to the object(s)/thing described

above?" and participants indicated their rating using a line-mark rating scale labeled "negative"

on the left end (coded as -100) and "positive" on the right end (coded as +100). The descriptions were displayed in random order and all participants gave ratings for all descriptions.

Texture-description match rating task. Because of the large number of texture-description pairings, we used only 47 of the original 62 visual textures in this task to keep it from being too long. We chose which textures to exclude based on their similarity to other textures already included (see Stimuli section). On each trial, participants viewed one of the 47 textures along with one of the descriptions ascribed to that texture against a neutral gray background. The description was displayed in black, 28-point Arial font, horizontally centered, approximately 110 pixels above the top edge of the texture. Below each texture-description pairing was a line-mark rating scale with the text "How well does this description match the image?" The scale was labeled "very poorly" on the left end (coded as 0) and "very well" on the right end (coded as 1). Participants saw all possible parings of texture images and descriptions ascribed to those textures. We split the participants for this experiment into two groups: One group saw the same description for all textures given that description before moving on to the next description, and the other group saw the same texture for all descriptions given to that texture before moving on to the next texture. The order of the descriptions and the textures was randomized.

Aesthetic-property rating tasks. Participants viewed each of the 62 visual textures, one at a time, against a neutral gray background with the text "This texture looks:" displayed below the texture. They indicated their rating for the requested aesthetic property using a line-mark rating scale coded from -100 to +100.

For the first aesthetic-property rating task, participants rated the textures on three different antonym pairs: NotPremium-Premium, NotSophisticated-Sophisticated, and Rough-Smooth. The experiment was blocked so that participants rated all textures on one aesthetic

property before moving on to the next, and they were given a short break between blocks. The

order of the aesthetic properties was randomized as was the order of the textures within a block.

The second aesthetic-property rating experiment was identical to the first, except that

participants rated the textures on only one antonym pair: NotWood-Wood.

Review

Table 2 gives an overview the six groups of participants and their experimental tasks.

Table 2. Participants and their Experimental Tasks

Participant Group	Number of Participants *started (finished)*	Task	Description of Task	Data from Task (see variables in Table 1)
1	36	Texture-preference rating	Participants rated how much they like each visual texture on a scale from Dislike (-100) to Like ($+100$)	Baseline ratings for model checking
2	32 (31)	Object description	Participants typed out descriptions of objects that they thought might describe or be associated with each visual texture	n_t and D_t for all textures t
3	37	Description-valence rating	Participants rated their emotional response to each object description (consolidated from Group 2) on a scale from Negative (-100) to Positive ($+100$)	v_d for all descriptions d
4	27 (26)	Texture-description match rating	Participants rated how well the descriptions (consolidated from Group 2) matched the visual textures they were ascribed to on a scale from Very Poorly (0) to Very Well (1)	w_{td} for all textures t and descriptions d
5	24	Aesthetic-property rating	Participants rated visual textures on the scales NotPremium-Premium, NotSophisticated-Sophisticated, and Smooth-Rough (all -100 to $+100$)	Ratings for "premium", "sophisticated" and "rough"
6	22 (21)	Aesthetic-property rating	Partcipants rated visual textures on a scale ranging from NotWood (-100) to Wood ($+100$)	Ratings for "wood"

Results and Discussion

We first describe participants' texture preference ratings for the 47 visual textures used in

all tasks. We then test the ecological valence hypothesis for these textures, namely that texture

preferences can be explained by the combined valence of all objects associated with the textures.

Next, we test this account against the single-associate and namesake hypotheses. Finally, we fit

the model of Thumfart et al. (2011) and compare the results of this model to those of the

ecological valence model.

Texture preferences

We used Cronbach's coefficient alpha (Cronbach, 1951; see also, Ritter, 2010) to assess

interrater reliability for texture-preference ratings. Reliability was high (0.83), so we averaged

across participants to get mean preference ratings. Figure 3 shows the mean preference ratings

for the 47 visual textures used in all experiments (see table A2 in the appendix for preference

ratings for the other 15 visual textures used in this experiment). The most preferred textures

included clouds, lettuce, lace, and tropical sea. The least preferred textures were the two blurred

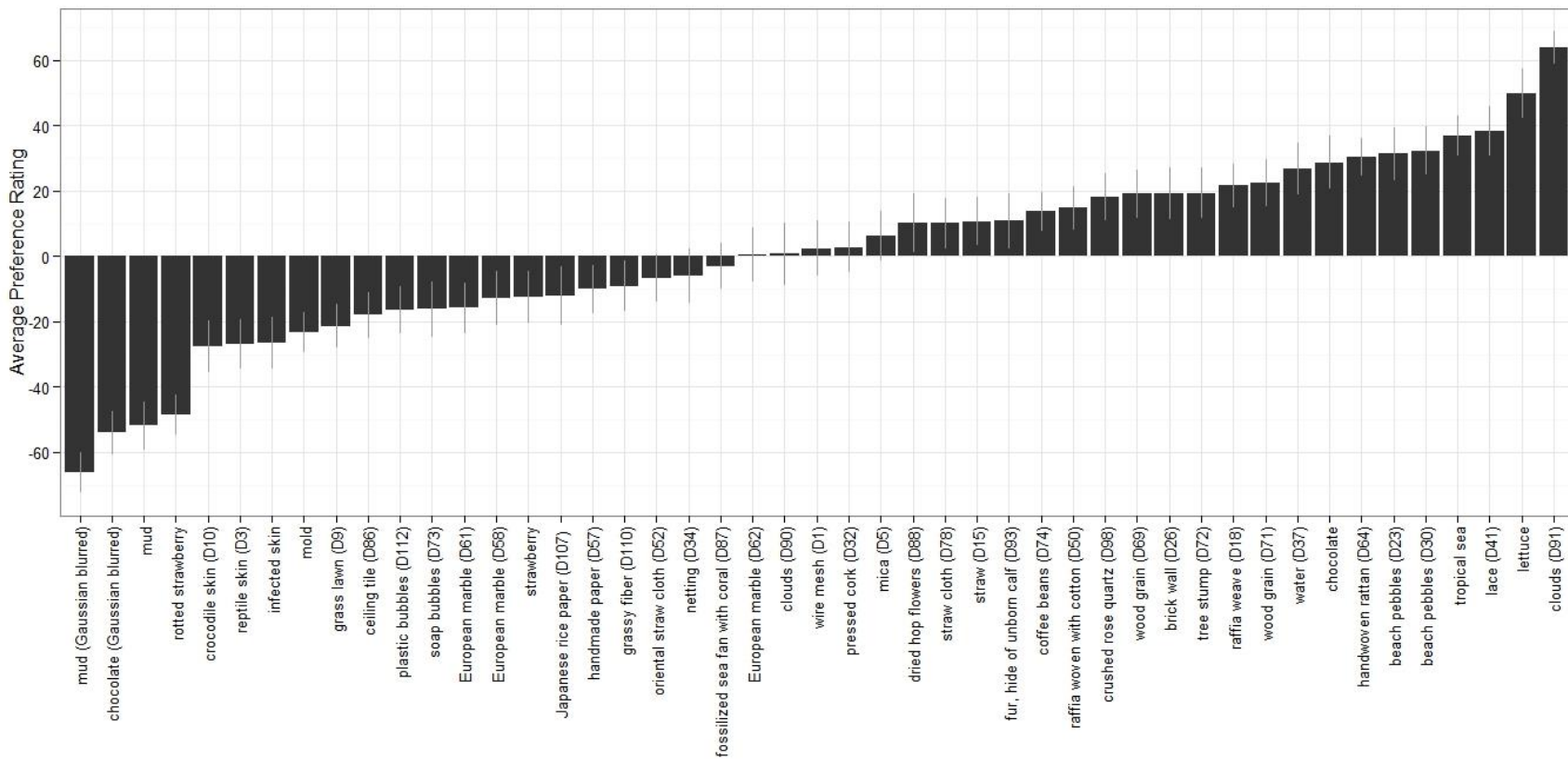images, mud, and rotted strawberry.

Figure 3. Average texture-preference ratings (*y*-axis) for each of 47 visual textures (*x*-axis). For Brodatz textures, Brodatz's (1966) identifier is given in parentheses, and the labels are his descriptions. For the other textures, the labels come from the Shutterstock search terms that produced the images. Error bars represent the standard errors of the means.

**Testing the ecological account of texture preferences**

We used Cronbach's coefficient alpha to assess interrater reliability separately for the groups that rated description-valence and description-texture matching. Reliability was high for these groups $(0.96, 0.99)$, so we averaged across participants to obtain mean valence-ratings for each description (v_d) and mean match-ratings[6] for each texture-description pairing (w_{td}). Using these ratings along with the consolidated descriptions from the object description task, we were able to predict preference ratings for the each texture using Equation 1.

Figure 4 shows the strong positive correlation $(r = 0.79, p < .001)$ between the measured texture preferences, averaged across participants, and the preferences as predicted by Equation 1. This indicates that people like textures that remind them of positive objects and dislike textures that remind them of negative objects, and it supports the ecological valence hypothesis. It is also worth noting that the prediction equation accounted for less variance if we either dropped the normalizing term $1/n_t$ $(r = 0.66)$, or set all of the non-zero weights $(w_{td}$'s$)$ to 1 $(r = .72)$.

---

[6] Recall that we split the participants in the texture-description match rating task into two groups: both groups saw all pairings of textures and descriptions ascribed to those textures, but one group saw the same description for all textures given that description before moving on to the next description, and the other group saw the same texture for all descriptions given to that texture before moving on to the next texture. The latter group went significantly slower (mean difference in reaction time per trial $= 2.05\ sec$, $t(24) = 10.36,\ p < .001$), but, as Cronbach's alpha indicates, the groups' ratings were in strong agreement, and using data from either group led to the same correlation coefficient.
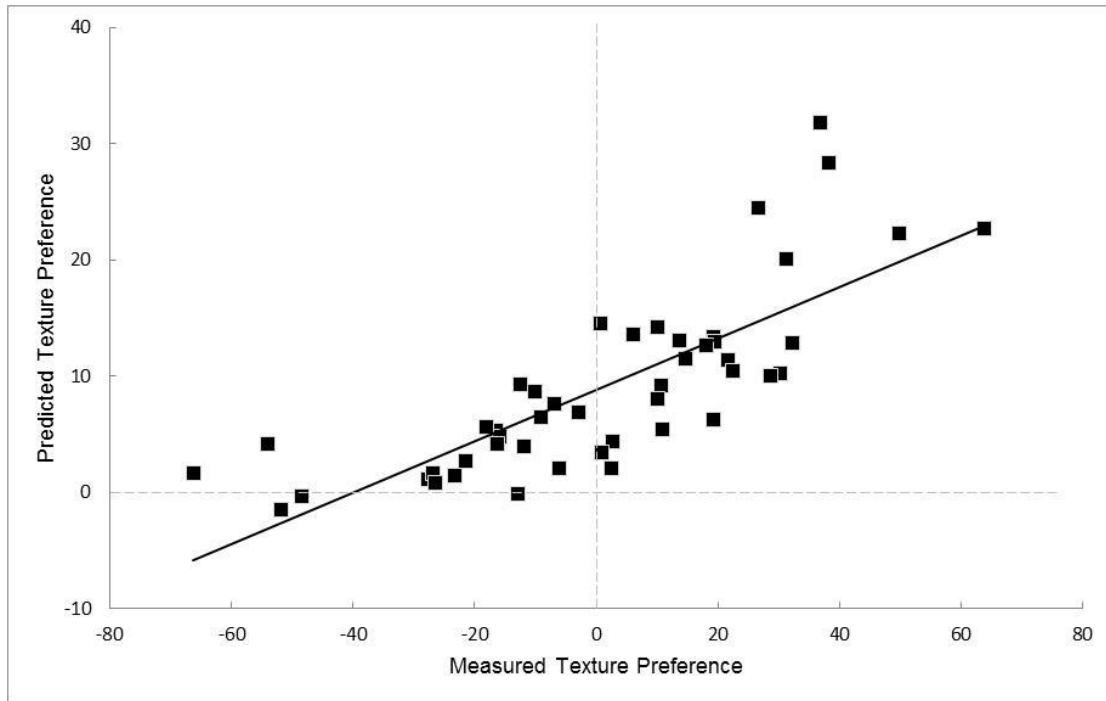
Figure 4. For a set of 47 visual textures, shows the correlation ($r = 0.79, p < .001$) between measured preferences, averaged across participants, and the preferences predicted by Equation 1. Numbers represent ratings on a line-mark rating scale, where -100 indicates that the texture is liked "not at all" and +100 indicates that it is liked "very much."

The main problem with the predicted ratings is that their range is compressed compared to the measured ratings and that most of the predicted ratings are positive while many of the measured ratings are negative. Palmer, Schloss and colleagues reported similar issues when predicting color preferences (Palmer & Schloss, 2010) and odor preferences (Schloss et al., 2015). To explain why the model underpredicts negative ratings, they hypothesized that participants in the object-description group either underreport negative objects because they are too shy to report gross or disgusting things, or they are just biased toward generating/thinking about positive objects. To explain why the ratings for the prediction are compressed relative to the measured ratings, we posit that people likely have weaker feelings about text describing objects than they have about actual images.

Despite these failings, the model does reasonably well: The predictions account for 62.5% of the variance in measured preference with a single predictor.

Evaluating the single-associate hypothesis

We tested the single-associate hypothesis by first identifying the object named most frequently for each visual texture (see table A1 in the Appendix). We then calculated the correlation between average preference ratings for the 47 visual textures and average valence ratings for the most frequently named objects. Although there was a significant positive correlation ($r = 0.47, p < .001$), it was weaker than the relation between measured preferences and preferences predicted by Equation 1 (correlations compared using the Fisher r-to-z transformation: $z = 2.63, p = .01$). Therefore, texture preferences were better predicted by a summary statistic of valence for all objects associated with a texture than by valence for the single object most frequently associated with a texture.

It is also worth noting that, although the valence of the most frequent description did not predict texture preference better than the valence of all associated descriptions, texture preference was negatively correlated with the number of descriptions ($r = -0.67, p < .001$), so that the textures with fewer descriptions were better liked. Taylor and Franklin (2012) found this to be the case with colors as well, but Schloss et al. (in press) did not, nor did Schloss et al. (2015) find this with odors. On average, each texture had 18.4 descriptions associated with it (range = 6-34).

Evaluating the namesake hypothesis

We tested the namesake hypothesis by first identifying the "namesake" object for each visual texture, that is, the object from which the texture image was created. We then calculated

the correlation between average preference ratings for the visual textures and average valence ratings for the namesake objects.

To identify the namesake objects, we chose the consolidated description that was closest either to Brodatz' (1966) description (for the Brodatz textures) or to the Shutterstock search term that produced the image (for the Shutterstock textures). The matches were not always exact, but were reasonably close. For example, Brodatz described texture D37 as "water," but we used "ocean/sea" as the namesake object. Similarly, one of the Shutterstock textures was produced from the search term "rotten strawberry," but we used "rotten/spoiled food" as the namesake object (see table A1 in the appendix for more details). We excluded texture D32 because nothing close to Brodatz' description for the texture (pressed cork/corkboard) was ever given as an object description and thus D32's namesake was never given a valence rating. We also excluded the two Gaussian blurred visual textures because the namesake objects for these textures would be too difficult to visually determine. Thus, overall, the correlation was based on ratings for 44 of the visual textures.

Although there was a significant positive correlation ($r = 0.60, p < .001$), it was weaker than the relation between measured preferences and preferences predicted by Equation 1 (with the 44 textures: $r = 0.82, p < .001$; correlations compared using the Fisher r-to-z transformation: $z = 2.10, p = .04$). Thus texture preferences were better predicted by the average valence over all objects associated with that texture than by the valence of the namesake object that produced the texture.[7]

---

[7] When reading a preliminary version of this manuscript, Charlie Chubb pointed out that the namesake hypothesis still explains a good amount of the variance considering it has a significant additional constraint. The problem with comparing the namesake hypothesis model to the full EVT model is that both models lack parameters so there's no way to quantify this additional constraint (i.e., the models can't be compared with a traditional methodology such as a likelihood ratio). This problem also arises when comparing the single-associate hypothesis model to the full EVT model.

Comparing to the computational model of Thumfart et al.

While the ecological model accounted for 62.5% of the variance in texture preference ratings, the accuracy of this prediction was not nearly as good as that provided by the computational model of Thumfart et al. (2011) and Liu et al. (2015), the latter of which claimed to account for as much as 99% of the variance in human preference ratings. However, the ecological model utilized only a single predictor, whereas the models of Thumfart et al. and Liu et al. utilized 6 and 5 predictors respectively. Since Thumfart et al. optimized their model using an error measure that punished the model for complexity (i.e., number of predictors), we were able to compare the results of the ecological model to the results of Thumfart et al.'s model, taking into account the lower complexity of the ecological model. The punished error measure was a convex combination of the following form (see Appendix B for more details):

$$Error_{punished} = \alpha \times Accuracy + (1 - \alpha) \times Complexity \qquad (1)$$

with α in the range $[0,1]$. Thus, the parameter α controls the ratio of accuracy versus complexity, with lower complexity emphasized for decreasing values of α. Using this formulation, we found that the punished error for the ecological model matched that of Thumfart et al.'s model when $\alpha = 0.35$ (see Appendix B). Thus, although the ecological model is less accurate than Thumfart et al.'s model, it has comparable performance when the modeler values low complexity over high accuracy. We were unable to make such a comparison with Liu et al.'s model since they did not work with a punished error term.

We also tested Thumfart et al.'s model more directly by measuring the six factors in their final texture-preference prediction equation: (1) how "premium" a texture looks, (2) how "sophisticated" a texture looks, (3) how "rough" a texture looks, (4) how "woodlike" a texture looks, (5) the skewness of the distribution of Fourier energy in concentric rings in the Fourier

power spectrum of a texture (*cSkew*), and (6) the *strength*, a computational feature derived from the neighborhood graytone difference matrix of a texture (see Table 1). We gathered ratings for "premium," "sophisticated," "rough" and "woodlike" in the aesthetic-property ratings tasks, and we measured the computational features *cSkew* and *strength* following Thumfart et al. (2011) and Amadasun and King (1988) respectively. We used Cronbach's coefficient alpha to assess interrater reliability separately for ratings of "premium", "sophisticated", "rough", and "woodlike," Reliability was generally high for all rating scales ($0.76, 0.81, 0.75$, and 0.93 respectively), so we averaged across participants and used the average ratings as predictors for each texture.

A multiple linear regression revealed that "premium" was the only significant predictor (Full Model: $F(6,40) = 12.2, p < .001$, multiple-$R^2 = 0.65$; Premium: $t(40) = 7.56, p < .001$). Interestingly though, how "premium" a texture looks accounted for 62.3% of the variance in texture preference ratings ($r = 0.79, p < .001$), almost identical to the variance accounted for by the ecological valence model. Furthermore, "premium" seemed to track better with preference than the ecological predictions since it was not compressed relative to preference, nor was it inflated in the positive direction (see Figure 5). We hypothesize that this is simply because participants in our experiment viewed Dislike-Like as synonymous with NotPremium-Premium.
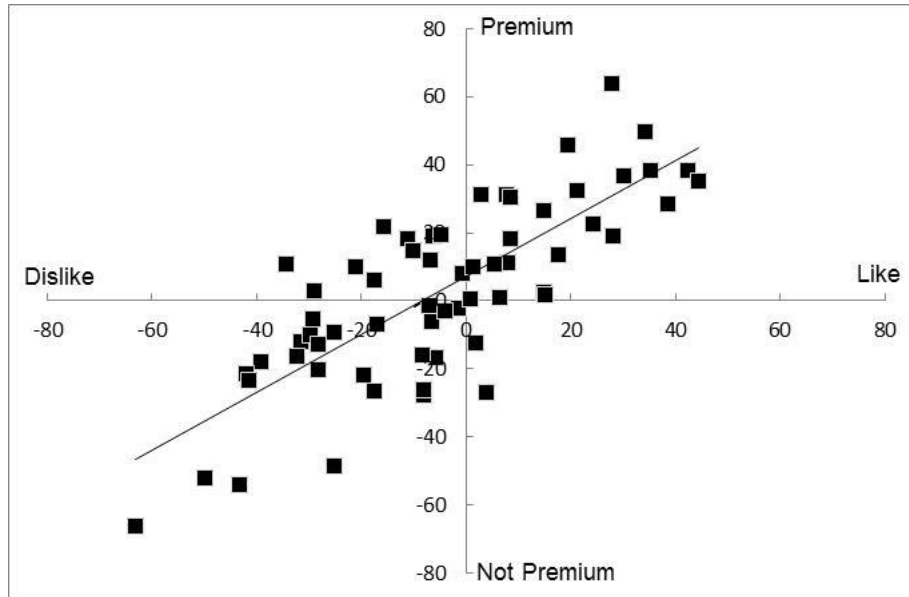
Figure 5. For a set of 47 visual textures, shows the correlation ($r = 0.79, p < .001$) between measured preferences and measured premiumness, averaged across participants. Numbers represent ratings on a line-mark rating scale, where -100 indicates that the texture is liked "not at all" or is "not premium" and +100 indicates that it is liked "very much" or is "premium."

We should also note that this is not necessarily a critique on Thumfart et al.'s model since they did not intend the model to be tested in this way. In personal communication, S. Thumfart indicated that they intended only to investigate the feature set as a whole (i.e., all 188 computational features + 27 aesthetic antonym pairs) and that any particular equation from the paper may not necessarily be applicable to other samples of textures.

Finally, we did not explicitly test Liu et al.'s instantiation of the model, but we note that they found the aesthetic property Mussy-Harmonious to be an important predictor of preference, and we hypothesize that, as with NotPremium-Premium, Mussy-Harmonious may serve as a synonym for Dislike-Like. Indeed, there is evidence that this may be the case since "preference" and "harmony" are often conflated in color research (for discussion, see Palmer et al., 2013).

General Discussion

The goal of the present study was to test whether the ecological valence theory – already used to explain color preferences (Palmer & Schloss, 2010) and odor preferences (Schloss et al., 2015) – could be extend to account for people's preferences for visual textures. The theory posits that preference for a given stimulus is determined by the combined valence of all objects associated with that stimulus, and that those preferences steer organisms to approach beneficial outcomes and avoid harmful ones (Palmer & Schloss, 2010; Schloss et al., 2015). We found support for the theory for visual textures: The combined valence of all objects associated with a set of visual textures accounted for 62.5% of the variance in texture preference ratings, which was more than was explained by the valence of the single most-frequently associated object for each texture or by the valence of the single namesake object that produced each texture.

While this prediction accuracy is reasonable, the model did better explaining average preferences for colors (where it accounted for 80% of the variance; Palmer & Schloss, 2010) and odors (where it accounted for 76% of the variance; Schloss et al., 2015). This may seem curious, especially when compared to the model's performance for colors: Looking at the visual textures in Figure 2, one might reason that it should be more obvious what objects are associated with these textures than with a set of uniformly-colored images, so the model should perform better for visual textures than for colors. However, the objects associated with each visual texture were not as obvious as one would think: Each texture had an average of 18 descriptions associated with it. Furthermore, visual textures are not homogeneous fields like colors, and other factors such as composition, style or context likely play a role in determining preferences for visual textures (for review, see Palmer et al., 2013).

The prediction accuracy of the ecological valence model also pales in comparison to the computational models of Thumfart et al. (2011) and Liu et al. (2015) – the latter of which claims to account for as much as 99% of the variance in texture preference ratings – although we found that the ecological valence model performs similarly to Thumfart et al.'s model when one uses an optimization measure that emphasizes lower complexity (the ecological model has only 1 predictor; Thumfart et al.'s model has 6). One problem with the computational models is their use of some aesthetic properties that participants may interpret as synonymous with preference (e.g., NotPremium-Premium for Thumfart et al., Mussy-Harmonious for Liu et al.). Still, for practical applications (e.g., in marketing or advertising), a computational model like Thumfart et al.'s or Liu et al.'s is preferable not only because of the increase in prediction accuracy, but also because the end result of such a model is able to predict texture preference for novel images based only on computational features measured from the images (i.e., with no human intervention). While the ecological valence model has low complexity in a statistical sense – it only uses one predictor – it has high complexity in a pragmatic sense – computing that single predictor requires a lot of experimentation.

The real benefit of the ecological valence theory is as a theoretical tool. The computational model of Liu et al. (2015) may be able to account for 99% of the variance in human texture-preference, but it utilizes predictors that have no theoretical significance. For example three of the predictors in their equation for texture preference come from a set of 20 features which were generated from a larger set of 106 computational features whose dimensionality was reduced using stochastic neighbor embedding. Because of the nature of this reduction technique, these features don't even have a clear physical interpretation (unlike the starting 106 computational features) let alone a compelling theoretical one. The ecological

valence theory, on the other hand, stems from a basic, evolutionarily-inspired premise: An observer's preferences for low-level stimuli should be driven by the real-world objects those stimuli are most associated with, and how harmful or beneficial those objects are to the observer. This premise leads to an entire theory of preferences (i.e., the ecological valence theory) and generates testable predictions.

The main limitation of the present study was the range of visual textures used. We used only 63 visual textures (only 47 for prediction), most of which were Brodatz textures, and all of which were naturalistic and grayscale. This set of visual textures does not span the range of possibility, and it could be the case that the ecological valence model performs poorly for abstract, non-natural textures. Furthermore, it is known that adding color to a visual texture affects a person's emotional response to that texture (e.g., Lucassen, Gevers, & Gijsenij, 2010). Any future study should utilize a more complete set of textures, perhaps sampling from a range of texture databases (as in, e.g., Thumfart, Heidl, Scharinger, & Eitzinger, 2009).

In sum, we conclude that – despite the limitations and although there are likely other mechanisms at play – it is reasonable to think of a person's preference for a visual texture as a summary statistic of how they feel about all the objects associated with that texture.

References

Amadasun, M., & King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man and Cybernetics, 19*(5), 1264-1274.

Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of Vision, 9*(11), 10.1-17.

Ben-Shahar, O. (2006). Visual saliency and texture segregation without feature gradient. *Proceedings of the National Academy of Sciences of the USA, 103*(42), 15704-15709.

Bergen, J. R. (1991). Thoeries of visual texture perception. In D. Regan (Ed.), *Vision and visual dysfunction, vol. 10B* (pp. 114-134). New York: Macmillan.

Brodatz, P. (1966). *Textures: a photographic album for artists and designers.* New York: Dover.

Chen, Y., Sobue, S., & Huang, X. (2008). Mapping functions of color image features and human KANSEI. *IIHMSP '08 International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (pp. 725-728). Harbin: IEEE.

Chronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 197-334.

Dong, Y., Tao, D., Li, X., Ma, J., & Pu, J. (2015). Texture classification and retrieval using shearlets and linear regression. *IEEE Transactions on Cybernetics, 45*(3), 358-369.

Gatto, J. A., Porter, W. A., & Selleck, J. (2010). *Exploring visual design: The elements and principles* (4 ed.). Davis Publications.

Guo, Z., Zhang, L., & Zhang, D. (2010). Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern Recognition, 43*(3), 706-719.

Haindl, M., & Filip, J. (2013). *Visual texture: accurate material appearance measurement, representation and modeling.* London: Springer-Verlag.

Haralik, R. M., Shanmugam, K., & Dinstein, I. (1973). Texture features for image classification. *IEEE Transactions on Systems, Man and Cybernetics, 3*(6), 610-621.

Ho, Y. X., Landy, M. S., & Maloney, L. T. (2006). How direction of illumination affects visually perceived surface roughness. *Journal of Vision, 6*(5), 634-648.

Joshi, D., Datta, R., Fedorovskaya, E., Luong, Q., Wang, J. Z., Li, J., & Luo, J. (2011).

      Aesthetics and emotions in images. *IEEE Signal Processing Magazine, 28*(5), 94-115.

Julesz, B. (1962). Visual pattern discrimination. *IEEE Transactions on Information Theory, 8*(2),

      84-92.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature,*

      290, 91-97.

Kawamoto, N., & Soen, T. (1993). Objective evaluation of color design II. *Color Research and*

      Application, 18(4), 260-266.

Kim, J., & Anderson, B. L. (2010). Image statistics and the perception of surface gloss and

      lightness. *Journal of Vision, 10*(9), 3.

Kim, S., Kim, E. Y., Jeong, K., & Kim, J. (2006). Emotion-based textile indexing using colors,

      texture and patterns. In *Advances in Visual Computing* (pp. 9-18). Berlin: Springer.

Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision*

      Research, 31(4), 679-691.

Landy, M. S., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa, & J. S.

      Werner (Eds.), *The Visual Neurosciences* (pp. 1107-1118). Cambridge, MA: MIT Press.

Lee, J., & Park, E. (2011). Fuzzy similarity-based emotional classification of color images. *IEEE*

      Transactions on Multimedia, 13(5), 1031-1039.

Liu, J., Lughofer, E., & Zeng, X. (2015). Could linear model bridge the gap between low-level

      statistical features and aesthetic emotions of visual textures? *Neurocomputing, 168*, 947-

      960.

Lucassen, M. P., Gevers, T., & Gijsenij, A. (2011). Texture affects color emotion. *Color*

      Research and Application, 36, 426-436.

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A, 7*(5), 923-932.

Montoyishi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature, 447*, 206-209.

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences of the USA, 107*, 8877-8882.

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology, 64*, 77-107.

Portilla, J., & Simoncelli, E. (2000). A parametric texture model based on joint statistics of complex wavelent coefficients. *International Journal of Computer Vision, 40*(1), 49-71.

Randen, T., & Husy, J. H. (1999). Filtering for texture classification: a comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*(4), 291-310.

Rao, A. R., & Lohs, G. (1993). *Towards a texture naming system: Identifying relevant dimensions of texture perception.* IBM Research Technical Report, No. RC19140.

Rao, A. R., & Lohse, G. L. (1996). Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Research, 36*(11), 1649-1669.

Ritter, N. L. (2010). *Understanding a widely misunderstood statistic: Chronbach's alpha.* New Orleans, LA: Paper presented at Southwestern Educational Research Association Converence.

Rosenholtz, R. (2000). Significantly different textures: a computational model of pre-attentive texture segmentation. In V. David, *Computer Vision -- ECCV 2000* (pp. 197-2111). Berlin: Springer.

Rosenholtz, R. (2015). Texture perception. In J. Wagemans (Ed.), *Oxford handbook of perceptual organization.* Oxford, UK: Oxford University Press.

Schloss, K. B., & Palmer, S. E. (2014). The politics of color: preference for republican-red vs. democtratic-blue. *Psychonomic Bulletin and Reivew, 21*, 1481-1488.

Schloss, K. B., Goldberger, C. S., Palmer, S. E., & Levitan, C. A. (2015). What's that smell? An ecological approach to understanding preferences for familiar odors. *Perception, 44*, 23-38.

Schloss, K. B., Hawthorne-Madell, D., & Palmer, S. E. (in press). Ecological influences on individual differences in color preference. *Attention, Perception, and Psychophysics*.

Schloss, K. B., Poggesi, R. M., & Palmer, S. E. (2011). Effects of university affiliation and "school spirit" on color preferences: Berkeley versus Stanford. *Psychonomic Bulletin and Review, 18*, 498-504.

Sebe, N., & Lew, M. S. (2001). Texture features for content-based retrieval. In *Principles of visual information retrieval* (pp. 51-85). London: Springer-Verlag.

Strauss, E. D., Schloss, K. B., & Palmer, S. E. (2013). Color preferences change after experience with liked/disliked colored objects. *Psychonimic Bulletin and Review, 20*, 935-943.

Tamura, H., Mori, S., & Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics, 8*(6), 460-473.

Taylor, C., & Franklin, A. (2012). The relationship between color-object associations and color preference: further investigation of the ecological valence theory. *Psychonomic Bulletin and Review, 19*, 190-197.

Thumfart, S., Heidl, W., Scharinger, J., & Eitzinger, C. (2009). A quantitative evaluation of texture feature robustness and interpolation behaviour. In X. Jaing, & N. Petkov (Eds.),

Computer analysis of images and patters: Lecture notes in computer science (Vol. 5702, pp. 1154-1161). Berlin: Springer.

Thumfart, S., Jacobs, R. H., Lughofer, E., Eitzinger, C., Cornelissen, F. W., Groissboeck, W., & Richter, R. (2011). Modeling human aesthetic perception of visual textures. *ACM Transactions on Applied Perception, 8*(4), Article 27.

Tuceryan, M., & Jain, A. (1998). Texture analysis. In C. H. Chen, L. F. pau, & P. S. Wang (Eds.), *The Handbook of Pattern Recognition and Computer Vision* (2nd ed., pp. 207-248). New Jersey: World Scientific Publishing Co.

Tyler, C. W. (2004). Theory of texture discrimination of based on higher-order perturbations in individual texture samples. *Vision Research, 44*(18), 2179-2186.

Um, J., Eum, K., & Lee, J. (2002). A study of the emotional evaluation models of color patterns based on the adaptive fuzzy system and the neural network. *Color Research and Application, 27*(3), 208-216.

Varma, M., & Zisserman, A. (2005). A statistical approach to texture classification from single images. *International Journal of Computer Vision, 62*, 61-81.

Victor, J. D. (1988). Models for preattentive texture discrimination: Fourier analysis and local feature processing in a unified framework. *Spatial Vision, 3*(4), 263-280.

Yokosawa, K., Yano, N., Schloss, K. B., Prado-Leon, L. R., & Palmer, S. E. (2010). Cross-culteral studies of color preferences: US, Japan, and Mexico [abstract]. *Journal of Vision, 10*(7), 408.

Zhang, H., Augilius, E., Honkela, T., Laaksonen, J., Gamper, H., & Alene, H. (2011). Aanalyzing emotional semantics of abstract art using low-level image features.

Proceddings of the 10th International Conference on Advances in Intelligent Data

Analysis (pp. 413-423). Berlin: Springer-Verlag.

Appendix A

Table A1shows all of the information needed to run the calculations in this paper for the

47 visual textures that we predicted preferences for. Table A2 shows information for the

remaining 15 textures whose preferences we did not predict. The title (first column for both

tables) of each texture is either the description given by Brodatz (1966), with his numbering in

parentheses, or the search term that produced the image on Shutterstock.

Table A1. For the 47 visual textures whose preference was predicted, shows the average preference rating, rating predicted by the ecological valence theory, number of descriptions, average premium rating, most frequently associated object description with its average valence rating, and namesake description with its average valence rating. All rating scales vary from -100 to +100. Note: there is no valence rating for pressed cork because it was not listed during the object description task, and the blurred chocolate and mud were not given namesake descriptions.

Visual Texture	Preference	Ecological Prediction	nDescriptions	Premium	Most Frequent Description (MFD)	Valence of MFD	Namesake Description (ND)	Valence of ND
wire mesh (D1)	2.46	2.05	17	14.77	fence / gate	8.67	wire mesh	-13.33
crocodile skin (D10)	-27.65	1.10	29	-7.98	reptile skin (snake, crocodile, lizard)	-25.66	reptile skin (snake, crocodile, lizard)	-25.66
Japanese rice paper (D107)	-11.96	3.99	24	-31.49	branches / roots (of a tree)	15.86	paper	25.66
grassy fiber (D110)	-9.16	6.44	16	-25.24	grass	31.57	grass	31.57
plastic bubbles (D112)	-16.52	5.31	24	-5.72	bee hive	-29.03	bubbles	35.6
straw (D15)	10.7	9.24	14	-34.35	grass	31.57	straw	10.94
raffia weave (D18)	21.64	11.44	21	-15.65	woven basket	20.31	woven basket	20.31
beach pebbles (D23)	31.31	20.13	12	7.75	rocks / pebbles / stones	22.35	rocks/pebbles/stones	22.35
brick wall (D26)	19.21	6.28	6	-6.41	bricks (brick wall / brick building / brick sidewalk)	4.42	bricks (brick wall / brick building / brick sidewalk)	4.42
reptile skin (D3)	-26.96	1.72	25	3.99	reptile skin (snake, crocodile, lizard)	-25.66	reptile skin (snake, crocodile, lizard)	-25.66
beach pebbles (D30)	32.32	12.90	13	21.21	rocks / pebbles / stones	22.35	rocks/pebbles/stones	22.35
pressed cork (D32)	2.76	4.43	22	-28.99	wall of a house / building	13.02	pressed cork	--
netting (D34)	-6	2.07	21	-6.64	net / netting	-6.08	net / netting	-6.08
water (D37)	26.7	24.46	9	14.85	ocean / sea	67.34	ocean / sea	67.34
lace (D41)	38.29	28.34	17	35.32	flowers	59.28	lace / embroidery	35.15
mica (D5)	6.13	13.64	14	-17.51	rocks / pebbles / stones	22.35	wall of a house / building	13.02
raffia woven with cotton (D50)	14.74	11.55	12	-10.07	tree bark	12.47	cloth / fabric	27.49
oriental straw cloth (D52)	-6.85	7.61	19	-17.04	cloth / fabric	27.49	cloth / fabric	27.49
handmade paper (D57)	-10.11	8.64	20	-29.71	wall of a house / building	13.02	paper	25.66
European marble (D58)	-12.87	-0.12	26	-28.28	ocean / sea	67.34	granite / marble	31.39
European marble (D61)	-15.84	4.76	24	-8.30	rocks / pebbles / stones	22.35	granite / marble	31.39
European marble (D62)	0.62	14.52	14	0.89	rocks / pebbles / stones	22.35	granite / marble	31.39
handwoven rattan (D64)	30.31	10.30	16	8.39	woven basket	20.31	woven basket	20.31
wood grain (D69)	19.19	13.40	11	28.18	tree bark	12.47	wood	28.52
wood grain (D71)	22.48	10.48	18	24.25	wood	28.52	wood	28.52
tree stump (D72)	19.39	13.01	6	-4.79	tree bark	12.47	tree bark	12.47
soap bubbles (D73)	-16.19	4.18	34	-32.20	bubbles	35.6	bubbles	35.6
coffee beans (D74)	13.69	13.04	12	17.61	coffee beans	40.37	coffee beans	40.37
Oriental straw cloth (D78)	10.15	14.26	18	-21.13	wood	28.52	cloth / fabric	27.49
ceiling tile (D86)	-17.99	5.65	22	-39.24	wall of a house / building	13.02	wall of a house / building	13.02
fossilized sea fan with coral (D87)	-2.99	6.94	23	-4.01	leaves	34.87	coral / coral reef	42.84
dried hop flowers (D88)	10.15	8.11	20	1.24	insects / bugs	-35.54	flowers	59.28
grass lawn (D9)	-21.42	2.72	21	-41.91	grass	31.57	grass	31.57
clouds (D90)	0.83	3.49	16	6.46	clouds	61.82	clouds	61.82
clouds (D91)	63.88	22.74	8	27.75	clouds	61.82	clouds	61.82
fur, hide of unborn calf (D93)	10.78	5.47	6	5.47	animal fur (e.g., dog, cat)	6.69	animal fur (e.g., dog, cat)	6.69
crushed rose quartz (D98)	18.11	12.65	16	8.60	rocks / pebbles / stones	22.35	crystals / diamonds / gems	52.46
chocolate spread	28.67	10.07	20	38.47	chocolate	61.7	chocolate	61.7
chocolate (Gaussian blurred)	-54.01	4.17	21	-43.37	X-ray / ultrasound	-1.04	--	--
mud	-51.91	-1.52	23	-49.88	mud	-21.09	mud	-21.09
mud (Gaussian blurred)	-66.27	1.64	21	-63.15	people	40.24	--	--
lettuce	49.81	22.30	15	34.32	flowers	59.28	lettuce	37.99
mold	-23.27	1.50	29	-41.54	dirt / gravel / soil	-9.74	mold	-66.16
rotted strawberry	-48.37	-0.27	27	-25.08	beans	8.09	rotten / spoiled food	-74.77
strawberry	-12.57	9.31	27	1.84	eggs	33.25	strawberry	49.51
infected skin	-26.48	0.78	28	-17.61	ocean / sea	67.34	infection	-71.47
tropical sea	36.73	31.83	9	30.20	ocean / sea	67.34	ocean / sea	67.34

Table A2. For the 15 visual textures whose preference was not predicted, shows the average preference rating, number of descriptions, average premium rating, and most frequently associated object description with its average valence rating. All rating scales vary from -100 to +100.

Visual Texture	Preference	nDescriptions	Premium	Most Frequent Description (MFD)	Valence of MFD
woven cane (D101)	1.88	19	15.22	cloth / fabric	27.49
woolen cloth (D11)	12.13	17	-6.75	clothing (shirt / pants / sweater / jacket)	57.07
brick wall (D25)	18.43	6	-11.12	bricks (brick wall / brick building / brick sidewalk)	4.42
beach sand (D29)	-5.24	17	-29.34	cement / concrete / pavement	7.43
beach pebbles (D31)	45.84	10	19.52	rocks / pebbles / stones	22.35
lace (D39)	11.18	28	7.95	flowers	59.28
lace (D40)	38.27	23	42.47	flowers	59.28
lace (D42)	35.24	23	44.54	curtain	20.1
straw matting (D55)	-2.11	20	-1.38	clothing (shirt / pants / sweater / jacket)	57.07
European marble (D63)	-21.67	31	-19.67	spider web	-40.72
Oriental straw cloth (D80)	-20.28	21	-28.33	carpet / rug	15.51
Oriental straw cloth (D82)	-1.41	24	-7.00	carpet / rug	15.51
woven matting (D83)	7.91	21	-0.58	cloth / fabric	27.49
dried hop flowers (D89)	-26.21	23	-8.20	insects / bugs	-35.54
brick wall (D94)	31.42	5	3.01	bricks (brick wall / brick building / brick sidewalk)	4.42

Appendix B

Equation B1gives the details for Equation 2 in the paper and is adapted from Thumfart et al. (2011):

$$MAE_{punished} = \alpha \frac{MAE}{range(MAE)} + (1 - \alpha) \times \frac{ModelComplexity}{\max(ModelComplexity)} \quad (B1)$$

where MAE stands for "mean absolute error." For a given visual texture, the absolute error is the absolute value of the difference between the measured and predicted preferences for that texture. Averaging the absolute error across all textures gives the mean absolute error, and subtracting the maximum absolute error from the minimum gives $range(MAE)$. Thus, MAE and $range(MAE)$ could easily be computed for the ecological valence model from the data in the second and third columns of table A1 in Appendix A. $ModelComplexity$ is given by the number of predictors/regressors in the model. In Thumfart et al.'s model, they used up to 50 predictors so that $\max(ModelComplexity) = 50$.

Table B1 shows the values for MAE, $range(MAE)$, $ModelComplexity$, and max($ModelComplexity$) for the ecological valence model versus Thumfart et al.'s computational model (Note: MAE for the computational model is given in Table 3 of Thumfart et al., 2011, and $range(MAE)$ was calculated using Equation B1 after E. Lughofer indicated in personal communication that they used an α value of 0.90).

Table B1. Values for the variables in Equation B1.

Model	MAE	$range(MAE)$	$ModelComplexity$	max($ModelComplexity$)
EVT model	17.29	67.50	1	50
Thumfart et al. model	9.57	143.55	6	50

Using Equation B1 and the values in Table B1, it's straightforward to show that

$MAE_{punished\ EVT} = MAE_{punished\ Thum.} = 0.10$ when $\alpha = 0.35$.