The Use of Visual Statistical Features in Empirical Aesthetics

Daniel Graham

The Oxford Handbook of Empirical Aesthetics
Edited by Marcos Nadal and Oshin Vartanian

Subject: Psychology, Cognitive Psychology Online Publication Date: Aug 2020
DOI: 10.1093/oxfordhb/9780198824350.013.19

Abstract and Keywords

Evolution generally demands that the brain take advantage of the probable statistical structure in the natural environment. Much research in recent decades has confirmed that regular statistical features in natural scenes—especially low-level spatial regularities—can help explain processing strategies in the human visual system. Basic statistical features in various classes of human-created images broadly match those found in natural scenes. Such regularities can be seen as evolved constraints on the visual structure of aesthetic images and therefore human visual aesthetics. Some researchers have also attempted to find statistical features whose variation from natural images is associated with variations in preference and other aesthetic variables. There is evidence that variations in statistical features of luminance and color could be exploited by the visual system in certain situations. However, there is much ambiguity and variability in most reported relationships between variations in image statistical features and variations in measures of human aesthetics. In contrast, basic statistical constraints that align with efficient visual system processing are almost never violated in aesthetic images. Put simply, statistical features may constrain but may not explain variability in visual aesthetics. The chapter concludes with an outlook on future directions for research.

Keywords: Statistical regularities, efficient coding, natural scenes, art perception, skewness, visual system, retina, deep learning

Statistical Regularities in Images: An Introduction

Neuroscientific investigations of statistical features in images are now a well-established field of study. A “feature” in this case means a characteristic or regularity of an image that is reflected in measures of image structure, such as pixel values. Statistical features in natural images are relevant to visual neuroscience because of evolutionary constraints on visual systems, which demand that sensory systems operate parsimoniously in their ecology. Neuroscience has come to see the brain more generally as capable of adapting over evolution and development to take advantage of the likely physical structure of the external world, whether the structure is the visual environment, the auditory scene, spoken language, or other inputs. In other words, the brain has been shaped by the need to take advantage of statistical regularities in the environment.

Regularities in Spatial Statistics

Power Spectra

In terms of vision, this line of argument can help explain the evolved structure of the early visual system. Consider spatial statistics. A simple relationship exists between any two neighboring points in a visual scene: on average, they are likely to be similar in terms of light intensity, and this similarity falls off as the two points become more distantly separated. This regularity holds for the most part regardless of how big your two "points" are, or what you are looking at. By pointing your finger in a pseudorandom direction in your environment and assessing light intensity at the pointed-to location, as well as immediately to its right (or left, up, down, etc.), you can confirm this for yourself. This property does not always hold, but it is likely to hold over many samples. As a regularity, this feature—a correlation in pairwise intensity—is something the visual system can take advantage of to make its job of encoding and transmitting information more efficient.

In the case of pairwise correlations, one can explain basic neural encoding strategies in the retina as an evolved processing strategy that takes advantage of such correlations. In other words, retinal interneurons assume that visual input from the world will have pairwise correlations like those in nature. Consequently, these cells mostly respond in areas of the scene where pairs of neighboring points are *not* correlated, such as edges (Atick & Redlich, 1992; Graham et al., 2006). This makes sense because edge contours often define objects, and object detection, segmentation, and recognition are among the most critical functions of primate vision. Efficient retinal processing strategies of this kind are shared by all primates and indeed nearly all mammals and other vertebrates. For reviews of statistical regularities in natural scenes and their relevance to models of vision coding, see for example, Field, 1994; Geisler, 2008; Simoncelli & Olshausen, 2001).

In vision science, basic spatial regularities are generally measured using the power spectrum of spatial frequencies in an image. The power spectrum measures the relative contributions of sine-wave patterns of varying spatial frequency in the composition of an image. The sine waves in this case describe the intensity of basis functions, which look like stripes of differing size, number, and orientation (see Figure 1A). Mathematically, such two-dimensional sine-wave basis functions vary in terms of frequency, amplitude, orientation, and phase. Any image can be broken down into a collection of such basis functions using Fourier analysis; it can also be reassembled from the appropriate "recipe" of basis functions using the same mathematical machinery.

The lowest spatial frequencies correspond to basis functions that alternate only once between white and black (we are ignoring color for the moment), at any orientation. A strong contribution to low frequency content in images would be a horizon line separating bright sky from darker land. And indeed, such ecological structure—being so common—contributes most to the typical power spectrum of the natural visual world. High frequencies, on the other hand, correspond to fine detail in a scene. Although humans glean important information from high spatial frequencies, they constitute only a small fraction of the image's spatial variation. Overall, when spatial frequency is plotted against the contribution of that frequency to image structure on logarithmic axes, we see a straight line with a negative slope. The relationship between power S and spatial frequency f can be approximated mathematically as: $S = f^p$, where p is about equal to –2 for natural images. As mentioned earlier, it does not matter how large the "points" are that one considers in pairwise correlations: this is the "scale invariance" implied by the relationship $S = f^{-2}$, which is equivalent to $S = 1/f^2$. That is, power spectra display "1/f (one-over-f)" scaling (see e.g., Bak et al., 1987). Mathematically, p describes the fall off of the typical natural scene power spectrum with increasing spatial frequency. In particular, p describes the slope of the power spectrum when plotted on logarithmic axes. Note that some authors alternatively investigate the spatial frequency amplitude spectrum, which is the square root of the power spectrum; plotted on logarithmic axes, its typical slope for natural scenes is –1, or simply p/2. See Figure 1B.

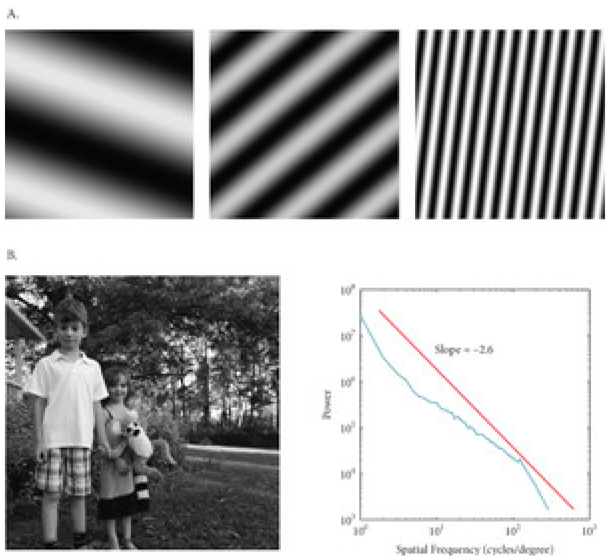

Figure 1. (A) Sine-wave gratings of various frequencies (increasing from left to right), amplitudes, and orientations. (B) Natural scene and its corresponding power spectrum. Any image can be decomposed into sine-wave components: the power spectrum (shown in blue) measures the contribution of each spatial frequency to the structure of the image, averaged over orientation. This function shows approximately linear fall off (shown in red) on logarithmic axes. Natural scenes typically have a slope of around –2 (this example has a slope of –2.6).

The power spectrum description is equivalent to measuring pairwise similarities of each point (pixel) in an image with all of its neighbors; the power spectrum description is more flexible and formalized (and therefore in most common use in vision science), but thinking of the pairwise correlation structure can be more intuitive. This chapter will continue to refer to pairwise correlations where appropriate, though the following sections are primarily elaborated in terms of the power spectrum description.

Power Spectra and Aesthetics

Art Images

Given these relationships, one may suppose that what is "natural" in terms of statistical features has a special place in human visual aesthetics. To a first approximation, this is the case. Consider a class of images often created for aesthetic purposes: artwork. Diverse collections of art images are known to have similar regularities in terms of statistics relevant to vision coding. For example, large samples of art images have a similar pairwise correlation structure as natural scenes. In particular, Graham and Field (2007) and Redies et al. (2007a) separately found essentially the same spatial regularities in different large samples of world artwork, with *p* having an average value of around –2 in both studies (see Graham & Redies, 2010, for a review).

To some extent this is unsurprising: although art styles vary widely within and across cultures, all art images are to a greater or lesser extent intended for the human eye and therefore must make use of visual patterns that our visual system is adapted to. Indeed, the variation in *p* for natural and artistic images of varying types is relatively small (see Table 1). This general consistency extends to abstract artwork including artwork produced with a degree of randomness, such as Jackson Pollock's drip paintings (Graham & Field, 2008a). Thus, even when artists are not depicting natural scenes, or when they employ randomness, they almost always follow the basic spatial regularities of natural scenes.

Table 1. Slope value (−p) for log-log plots of radially averaged Fourier power versus spatial frequency for different categories of natural and artistic images (from Graham & Redies, 2010).

	$-p$	SD[a]	n[b]
Natural scenes[c,d]	−2.0	0.3	208
Photographs of plants[c]	−2.9	0.4	206
Photographs of simple objects[c]	−2.8	0.3	179
Photographs of faces[e,f]	−3.5	0.2	3313
Graphic art of Western provenance[c]	−2.1	0.3	200
Artistic portraits (graphic art)[e]	−2.1	0.3	306
15th century	−2.0	0.2	20
16th century	−2.1	0.2	89
17th century	−2.1	0.4	34
18th century	−2.2	0.1	18
19th century	−2.2	0.4	50
20th century	−2.2	0.3	95
Etching	−2.0	0.3	50
Engraving	−2.1	0.2	17
Lithograph	−2.2	0.2	27
Woodcut	−2.4	0.4	13
Charcoal, chalk	−2.2	0.3	100
Pencil, silver point	−2.0	0.2	59
Pen drawing	−2.1	0.3	31

Scientific illustrations[c]	−1.6	0.3	209

[a] Standard deviation.

[b] Number of images in each category.

[c] Data from the study by Redies, Hasenstein et al. (2007).

[d] Images from the database of van Hateren and van der Schaaf (1998)

[e] Data from the study by Redies, Hänisch et al. (2007b).

[f] AR face database of Martinez and Benavente (1998).

White Noise

However, it is not necessary to presume that artistic images have a special place in aesthetics (cf. Nadal & Skov, 2018) in order to find evidence in support of the idea that basic statistical regularities that are relevant to vision coding also shape our visual aesthetics. For example, one can see *prima facie* that very statistically unnatural images such as white noise (Figure 2A) are not attractive. White noise is created by randomly assigning the intensity at each pixel. This class of images has an average pairwise correlation of zero (since each pixel value is chosen independently of its neighbors), and, correspondingly, a power spectrum slope $p = 0$. The latter result indicates that the contribution to image structure across spatial frequency in white noise is uniform.[1]

Because white noise images are very unlike the natural visual world to which humans are evolutionarily adapted, there is a sense in which we can’t even see them. For example, the white noise images shown in Figure 2A look indistinguishable, yet each one is utterly different from the others. Indeed, any pair of white noise images is likely to be far more different in terms of spatial structure than a given pair of natural images (see Chandler & Field, 2007). The visual system has adapted to a correlated world, and not to the uncorrelated world of white noise. In comparison, a random selection of images that has the pairwise regularities like those natural scenes but are otherwise completely random—1/*f* noise—will appear rather pleasant, perhaps reminiscent of cloud-watching (Rogowitz & Voss, 1990); such images are also readily distinguishable (Figure 2B).

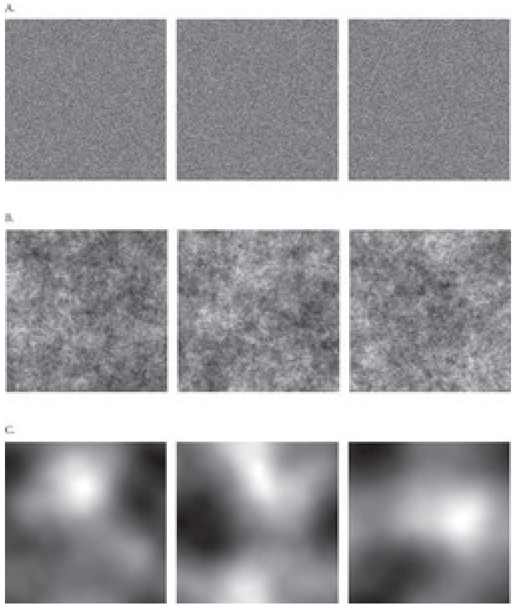

Figure 2. (A) White noise images, where each pixel value is chosen at random (in this case from a Gaussian distribution). In each image, a given pixel value is likely to be quite different from that of the corresponding pixel in the other images, and therefore these images are utterly different from one another. However, they are essentially indistinguishable. These images all have a flat power spectrum (slope = 0). (B) Noise images that possess 1/*f* scaling in their power spectra, i.e., 1/*f* noise. These images are also random, but have the same basic spatial statistical regularity as natural scenes, with a power spectrum slope of approximately −2.0. Such images are easily distinguished. (C) Blurry images that show power spectrum slope of around −6.0.

In terms of production, white noise has probably been created only twice by hand: this feat of craft was first accomplished by Attneave (1954). In this seminal paper on efficient visual system encoding, two military enlistees darkened approximately 20,000 squares by hand according to randomly generated numerical values. This feat was also accomplished by the French artist François Morellet in his painting *Random Distribution of 40,000 Squares Using the Odd and Even Numbers of a Telephone Directory*, 1960 (Mather, 2013; see Figure 3A). Viewed from the perspective of art history, humans have produced copious numbers of monochrome paintings and other handmade images that are in a way imperceptible (in the sense of having no spatial variation), but only two that resemble white noise. Interestingly, however, the red and blue pigments Morrellet used are essentially equiluminant (Figure 3B), so the image may be processed more like a monochrome. Thus, in terms of empirical aesthetics, noise images hold a special place as an almost universally disliked stimulus. However, beyond dislike, it would be interesting to know how humans treat such images in an aesthetic context—do they feel disgust or other negative valence emotions, or no emotions at all? Do they prefer a disgusting image to white noise?

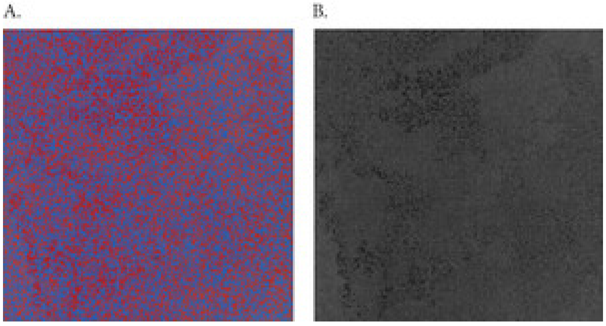

Figure 3. (A.) François Morellet, *Random Distribution of 40,000 Squares Using the Odd and Even Numbers of a Telephone Directory, 1960.* Museum of Modern Art, New York. B. The same image with color information removed: pixel values represent light intensity.

Blur

Images with power spectrum slope values of p much more negative than −2 are also difficult to perceive, and are also disliked. Such images appear very blurry. Randomly produced blurry images (e.g., generally those with $p > 3$, as in Figure 2C) are fairly easy to distinguish from one another due to differing location of the "blobs." However, they also seem to provoke frustration and aversion because of their indistinctness. When a natural scene is heavily blurred, the effect is the same. Conversely, anyone requiring strong optical correction can describe the pleasant sensation of seeing the world more clearly when they wear spectacles. That is, beyond the practical benefits of seeing the world without blur, there may be a related aesthetic dimension.

Figure 4. Philip Barlow, Refuge II, n.d. http://www.philipbarlow.com. This image has a spatial frequency power spectrum slope of ~–3.5.

We can see this in artistic production as well: professional photographs are very rarely blurry. As Ke et al. (2006) showed, in a large and diverse set of viewer-rated photographs, the statistical presence of blur in an image is an excellent predictor of low image quality ratings. In fact, blur was a far better predictor of preference in this study than statistical models that considered contrast, color, edge structure, or brightness. Anecdotally, artwork that employs substantial blur exists (e.g., the work of South African painter Philip Barlow; see Figure 4), but is also very rare. However, even Barlow's work has a power spectrum that is well fit by a straight line on logarithmic axes. And though the image in Figure 4 has a slope p that is somewhat outside the normal range for natural scenes (around −3.5 by the author's calculation), it is not nearly as far as that of the very blurry images in Figure 2C.

Barlow's work is of particular interest because he captures in paint something like the perceptual experience of humans with substantial optical errors. Since optical errors are highly prevalent in industrialized societies (e.g., 96% of 19-year-old males in Seoul were found to have myopia; Jung et al., 2012), perhaps researchers should consider how this myopia epidemic can affect aesthetics. If most adults see only a blur beyond a certain distance in depth (without optical correction), do they generally discount aesthetic considerations in this depth regime, for example in urban architecture? Such questions require research but it can be concluded that blur is rare and disliked aesthetically, though it is perhaps not as rare or as disliked as white noise.

There is also evidence that spatial statistics that differ in other ways from $1/f$ can be aversive. Fernandez and Wilkins (2008), for example, have shown that excess power in the mid-range of spatial frequency (especially high-contrast stripes at around 3 cycles/° in spatial frequency)—relative to natural spectra—generates discomfort. They argue that this discomfort may be related to conditions of optically induced headache and seizure. Thus, images that deviate strongly from natural spatial statistics may be disliked not only because they are imperceptible, but perhaps also because they negatively interfere with other brain processes.

The Perceptibility Hypothesis

In terms of image statistics, then, we like what is typical. As it turns out, what is preferred is also what we see best, as has been shown by Spehar et al. (2015). These authors tested human viewers on preference and acuity—the latter measured as detection at increasing contrast and as discrimination in terms of just-noticeable differences—for random noise stimuli that varied in their power spectrum slope p from –0.2 to –5.0. They found that there was a strong relationship between images that were best perceived and those that were preferred, which in both cases corresponds to a value of p of roughly –2. In other words, there is a very similar inverted U-shaped distribution for both preference and visibility as a function of power spectrum slope p, which is centered around p = –2. At the extremes of the distribution of p values, blurry images were somewhat preferred to white noise-like images. The same relationship between acuity and preference held for sine-wave gratings of varying spatial frequency. Thus, humans cannot detect or discriminate images with statistical features characteristic of white noise and blur, and we dislike those images, too. Noise images with natural spatial statistics, on the other hand, can be readily detected and discriminated, and are therefore preferred (see Figure 2). The idea that we generally prefer what is typical because we see it best has been termed the *perceptibility hypothesis* (Graham & Field, 2008a).

Fractal Dimension Statistics

Before considering statistical features beyond power spectrum slope, it is worth a brief diversion to discuss the concept of "fractal dimension," since it is related to the power spectrum. Fractal dimension is defined as a measure of the degree to which a single line fills up a plane, or the degree to which a single plane fills up a volume, due to fractal structure (see e.g., Mandelbrot, 1982). Fractals in this case are mathematical objects that are both self-similar and scale invariant: they show the same patterns of structure at all spatial scales, and are generated according to deterministic recursive functions.

The idea of fractal dimension has been applied to real-world (nonmathematical) objects as well, but it is important to note that it can only be approximated. Often, fractal dimension is approximated by calculating *box dimension*, which works as follows: an object is defined in terms of a binary boundary. A grid of boxes is overlain on the boundary, and the fraction of filled boxes is tallied. This is repeated with larger and smaller grids over several orders of magnitude. The function relating box size to the fraction of boxes required to cover the boundary at that size determines the box dimension. This procedure can be used to understand the space-filling quality of real-world binary boundaries such as coastlines, where the boundary is meaningful. But its application to images is ambiguous. Typically, images are thresholded to produce such boundaries (e.g., Viengkham & Spehar, 2018). But thresholding produces quite unnatural silhouette-like images, whose power spectra (and other image statistics) are greatly altered relative to the original image. Indeed, few if any parts of the natural visual world consist solely of binary boundaries; most boundaries are in fact low contrast (Ruderman & Bialek, 1994), which is likely the case with artwork as well given the limitations in its dynamic range of luminances (see Graham, 2011).

Moreover, even if such boundaries were meaningful, Normant and Tricot (1991) have shown that the box dimension metric is only an accurate measure of fractal dimension for images that are both scale invariant and self-similar (see also Kube & Pentland, 1988). Since natural images and art images are not self-similar, box dimension measurement is an unreliable measure of their fractal dimension (see Soille & Rivest, 1996; Theiler, 1990). It should also be noted that the application of box dimension alone as a characteristic of authorship of abstract artwork (e.g., for stylometrics or attribution, e.g., Taylor et al., 1999) has been refuted (Jones-Smith & Mathur, 2006).

If a correspondence between visual perception and fractal dimension did exist, it would likely involve the "2D" box dimension (the degree to which a 2D plane fills 3D space), rather than the "1D" box dimension (the degree to which a 1D boundary fills a 2D plane). This could be done for example by considering an image as an intensity surface filling a 3D volume. In this case, fractal dimension D_f is linearly related to the slope p of the spatial frequency amplitude spectrum: $p = 8 - 2D_f$ (Knill et al., 1990). This relation holds for all images whose spatial frequency power spectra are well-described by the function $1/f^p$.

However, nearly all research concerning “fractal dimension” of images measures instead the “1D” box dimension. Ultimately, then, box dimension is a flawed statistical feature in natural and artistic images, whereas the spatial frequency power spectrum slope yields a robust and accurate measure of the same spatial regularities. Researchers are thus encouraged to work within spatial frequency statistics, a formalism particularly suited to patterns in the physical world, as opposed to fractal analysis, which is most germane to mathematical objects.

Higher-Order Spatial Statistics

Are Artistic Images Special?

So far we have only considered spatial regularities of the lowest statistical order, which concern relationships between pairs of points (pixels). In an image, there may exist relationships among more than two pixels, which are called higher-order spatial statistical regularities. Such regularities are generally difficult to measure in images because of the combinatoric explosion of possible triplets, quadruplets, and so on; the difficulty in measurement is further exacerbated by the difficulty of describing any such regularities graphically or numerically. However, higher-order structure that is relevant to vision coding can be explored using spatial filters that resemble those employed in early visual system spatial processing, such as the Gabor-like spatial filters of V1 simple cells. Because simple cell-like filters appear to efficiently encode “sparse” higher-order statistical structure in natural scenes (Bell & Sejnowski, 1997; Olshausen & Field, 1996), their responses give an indication of the higher-order spatial regularities most important to primate vision coding.

One line of argument holds that artistic images are a special class of images due to their higher-order statistical properties. Christoph Redies and colleagues have performed numerous experiments that provide evidence for this proposition.

First, consider faces: as a class, faces are processed in the visual system using specialized mechanisms. For example, face processing such as symmetry detection is most effective for upright faces (Rhodes et al., 2005). If artists aim to exploit these mechanisms, they should reproduce the typical statistics of real faces. However, artistic portraits deviate in their pairwise spatial statistics compared with photographs of human faces (Redies et al., 2007b). Similar findings have been found in terms of higher-order statistics, as described below. And Graham and colleagues have found evidence for differences in basic structure between faces and handmade face representations, such as frontal portrait paintings (Graham et al., 2014), and masks from many world cultures (Prescott & Graham, 2020).

Special statistical features for artwork as an image class may not be limited to low-level structure. Redies and colleagues have also shown that, in artworks of different styles, local spatial structure is distinct from what is typical in other image classes, such as photographs of objects and facades. The higher-order structure they measured, termed edge orientation entropy, captures regularities in how edges continue in an image. In this analysis, edge elements are detected using a standard image-processing algorithm. Then each edge element's orientation is compared with that of its neighbors, or to all other edges in the image. One can then create a distribution of edge orientations as well as a distribution of edge relationships (as circular histograms). Uniformity in these distributions can be described in terms of their entropy. In the case of edge relationships, high entropy (i.e., uniformity) means that knowing a given edge's orientation tells one little about the orientation of other edges in the image, such as neighboring edges. Redies and colleagues found that artworks from numerous Western "high" art styles show quite similar distributions of edge orientation entropy, which are similar to those of certain natural scene categories such as large vistas. Similar results were found when the art was grouped by subject matter. Artworks are distinct in this sense from faces, as well as from several classes of human-created objects and architecture (Redies et al., 2017). Related results have been shown for artificial neural network systems that learn higher-order statistical regularities: diverse artwork from across Eurasia tends to group together in terms of its optimal neural network representation, in a way distinct from other human-created objects and buildings, as well as certain varieties of natural scenes (Brachmann et al., 2017; see also section Deep Learning, below). Also, when artists depict natural scenes, they tend to de-emphasize spatial frequency energy of horizontal and vertical orientations compared with natural scenes, which have disproportionate energy in cardinal orientations (Schweinhart & Essock, 2013).

However, it is not clear that all or even most artistic images occupy a special place in terms of human aesthetics because of their statistical properties (to say nothing of their purported specialness vis-à-vis cognitive appraisal or cultural context). To the extent that artwork is different from other image classes in terms of statistical features, this may be due to materials and compositional factors. For example, artists have freedom to create edges on a blank 2D canvas, typically using styli whose width falls in a small range. The 2D retinal image of the 3D physical world, on the other hand, can be modeled in a generative fashion from objects with power-law scaling in size, which can occlude each other, along with biases for cardinal orientations due to horizons, buildings, and trees (Ruderman, 1994, Switkes et al., 1978). In other words, the causes of statistical structure in natural images are distinct from those in art images because they are composed in fundamentally different ways. It is also possible that occlusions of objects in handmade art occur less or in different ways than in other classes of images. That is, humans may generally try to avoid depicting objects in handmade flat media using arrangements that produce a "bad" Gestalt due to occlusion. There is currently no evidence that "good" arrangements of objects are reflected in basic or higher-order spatial statistics. The distinctness of art from other artificial objects and buildings may also be explained by photographic biases in the former (McManus et al., 2011) and by rectilinearity in the latter. In any case, when considering the diversity of art styles beyond Western "high" art (Redies & Brachmann, 2017), as well as variations across time (Mather, 2018) and in art by neuroatypical individuals (Graham & Meng, 2011a), there is considerable variation in statistical regularities at low and high orders, which overlaps with what is typical in other image categories. However, to a first approximation, nearly all artwork shows the same $1/f$ power spectra as that typical of natural scenes. Moreover, despite differences in image statistics, artistic representations show similar basic perceptual responses: for example, portrait and landscape paintings can be discriminated by humans with similar accuracy compared with photographs of faces and landscapes when stimuli are presented rapidly (Graham & Meng, 2011b).

Bearing in mind these complexities, we may now ask whether variations in aesthetic response may be predicted by variations in higher-order statistical features. That is, given that low-level statistical regularities exist and are associated with preference, are variations in higher-order statistics correlated with or predictive of particular dimensions of aesthetic perception? Redies and colleagues have found that aesthetic ratings of images of several kinds of human-created objects and artificial patterns are correlated with higher entropy in edge orientations for those images, which is in turn partially predicted by the perceived curvilinearity of those images (Grebenkina et al., 2018). However, for the most nominally "artistic" of the image categories tested in this study—music album cover art—higher-order statistics (and post-hoc combinations thereof) explained only a quarter or less of data variance in aesthetic ratings. In this experiment, human-judged curvilinearity—a feature that to date has yet to be fully characterized in terms of low- or higher-order statistical regularities—explained much greater proportions of aesthetic rating data variance.

Though summarizing other research in this vein is beyond the scope of the present chapter (see Brachmann & Redies, 2017), we can summarize the affirmative statistical evidence related to the argument of Redies and colleagues as follows:

- Artwork of faces is different from real faces, and more like natural scenes.
- Artwork as a class is distinct from other human-created objects as well as faces and some natural scene categories.
- There is an association between aesthetic preference and a statistical feature that resembles curvilinearity for artificial objects and patterns.

Nevertheless, there is considerable ambiguity in the relationships between variations in higher-order statistics and aesthetics, and indeed in the statistical specialness of "artwork" (whether or not one believes artwork holds a special place in aesthetics).

On the other hand, it has been established that humans do have a preference for low-level spatial statistical regularities associated with natural and easily perceptible visual stimuli. One could speculate that, beyond general preference for pairwise spatial regularities due to efficient and widely shared coding strategies in the early visual system, visual aesthetics may not be an important enough neural function to warrant rigorous optimization for particular variations in low-order (or, for that matter, higher-order) spatial statistical features.

As we shall see once we look beyond spatial statistics, the visual system does not have a default aesthetic response to specific statistical properties in images, although general “rules” can be described for certain applied perceptual situations. Nor is what is natural necessarily what is aesthetically preferred in these cases.

Regularities in Luminance Statistics

One can appreciate the complexity of the situation by considering luminance distribution statistics, or the proportion of high-, middle-, and low-intensity light in an image. This area of research traces back to some of the first empirical work on efficient coding of natural scene regularities, which involved studies of natural luminances in insect vision (Laughlin, 1981). To a first approximation, scene luminances (light intensities) in natural daylight follow a lognormal distribution (Attewell & Baddeley, 2007; Brady & Field, 2000; Dror, Leung, Adelson, & Willsky, 2001). A lognormal distribution is a distribution whose logarithm is Gaussian. This means that most pixels in a natural scene send relatively low intensities of light to our eyes (the peak in the low intensities), but natural scenes also have intensities that extend far into the high intensities (the “heavy tail” of the distribution). In terms of descriptive statistics, symmetrical distributions like a Gaussian distribution have *skewness* (the third statistical moment) of 0, whereas natural scenes typically possess intensity distributions with skewness greater than 0, due to their approximately lognormal shape. See Figure 5.

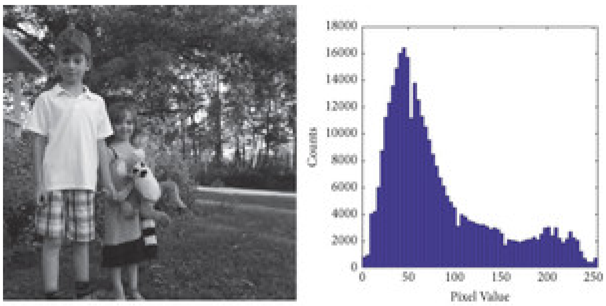

Figure 5. Natural scene and its intensity (pixel value) histogram. This histogram is roughly lognormal in shape and has a skewness of +1.1. Note that the scene image is not calibrated for luminance: because cameras compress high luminances in scenes more than low luminances, the true distribution of luminances in this scene extends much farther into the high luminances, and would likely have a correspondingly higher skewness.

Thus, we might expect to find that humans prefer this regularity. Indeed, this is what we would predict if we posit that low-level preferences follow from their statistical efficiency relative to sensory encoding. In particular, because cone photoreceptors respond in roughly logarithmic fashion at increasing light intensity, lognormally distributed scene intensities would be encoded as Gaussian responses, thereby making maximal information-theoretic use of the encoding mechanism of cone excitation.

In certain circumstances, there does appear to be a preference for high skew. For example, Yang et al. (2011) found that 7–8-month-old infants preferred to look at computer-simulated 3D objects that had high skewness in their intensity distributions (Figure 6). However, this was only the case when the high skewness corresponded with high glossiness (i.e., when bright highlights in the image accurately indicated glossy material reflectance). But there is also evidence that humans can accurately judge the freshness of photographs of real fruit and vegetables from small image patches, and that luminance distribution skewness by itself partly contributes to the accuracy of such judgments (Arce-Lopera et al., 2012; Wada et al., 2010). There is also evidence that higher values of luminance distribution kurtosis (the fourth statistical moment of the distribution, which is generally higher for distributions with high skew) are associated with preference for abstract art, explaining 25% of data variance (Graham et al., 2010). Taken together, this line of work suggests that some image classes (including those depicting single glossy objects) that show high luminance distribution skewness may be generally preferred.

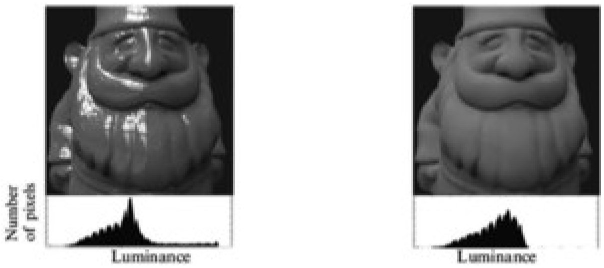

Figure 6. Glossiness and histogram (Yang et al., 2011).

Conversely, however, Graham et al. (2016) have shown that versions of a natural image manipulated to have high skew are systematically disliked compared with those manipulated to have low skew. For outdoor scenes, Graham et al. (2016) found that humans generally prefer skew lower than what was present in the original scene. That is, we prefer natural scene images to be substantially altered in terms of their luminance distributions relative to what is natural. In a suite of studies, artistic photographs of Western U.S. landscapes as well as images from a standard natural scene database were manipulated to change their skew. This procedure leaves the mean and variance of the intensity distribution (and, of course, spatial relationships of pixels in the image) unchanged. Regardless of the inclusion of glossy objects in the scenes—and even when pixels were randomly scrambled—human viewers consistently preferred low-skew versions of the images. This finding may not be so surprising when we consider that human painters overwhelmingly produce low-skew images (Graham & Field, 2008a). In part, the low-skew bias in artwork is due to the limited dynamic range of luminances available for 2D reflective objects (see Graham et al., 2008b), but aesthetic demands may also bias human artists toward low-skew images. Indeed, it is possible to make high-skew images by hand, but because of the limited dynamic range, such images will necessarily be mostly quite dark, with a few highlights.

Otaka, Shimakura, and Motoyoshi (2019) have also investigated luminance regularities in images of skin patches. Since the material properties of skin are closely associated with biologically relevant aesthetic judgments such as attractiveness, this is a potentially useful arena for understanding statistical features related to aesthetics. Otaka and colleagues (2019) found that a range of perceptual judgments such as glossiness, smoothness, healthiness, and evenness can be summarized into two subjective factors: "goodness" and "glossiness." These factors in turn showed high correlations to simple luminance and color statistics, particularly the variance, skew, and kurtosis of the images' luminance and color distributions. Interestingly, both the glossiness attribute and the "goodness" factor were anticorrelated with luminance distribution skewness. In other words, low skew is preferred.

Thus, it appears that humans may have a general taste for low-skew scenes (at least for complex outdoor scenes, and to a lesser extent, skin), which artists have, on average, attempted to sate. Again, the artwork result (Graham & Field, 2008a) may be related to perceptibility: although one can make a high-skew image by hand, it will be mostly dark; lower-skew images may have more detail to explore, making them in a way more perceptible. However, in certain functional situations, on the other hand, such as judging fruit and vegetable ripeness we may be particularly attuned and attracted to high skewness associated with glossiness. But it should be said that researchers have criticized this approach on theoretical and empirical grounds (see e.g., Anderson & Kim; 2009; Fleming, 2014). It is certainly not the case that a particular skew of the luminance distribution will guarantee preference (or, for that matter, glossiness). The possible neurobiological underpinnings of such a process also remain undefined.

Regularities in Color Statistics

Uniform Color Patches

Color is another realm where basic statistical features have been investigated in terms of empirical relationships with aesthetics.

Karen Schloss, Stephen Palmer, and colleagues have performed extensive research across diverse populations to determine mean preferences for small patches of color presented in isolation. In a series of experiments, Palmer and Schloss (2010) established a logical chain of evidence connecting numerical liking ratings for color patches and liking ratings of natural objects of the same color. In particular, color patches whose hue is commonly associated with objects whose average rated valence is positive are generally rated more highly than color patches whose hue is commonly associated with objects whose average rated valence is negative. For example, mean preference across individuals generally favors blue and disfavors brown: in turn, Palmer and Schloss (2010) show systematic evidence that humans judge canonically blue objects like clear sky and clean water as positive, while we judge canonically brown objects like feces and rotten food as negative.

What can these results tell us regarding color statistics? Generalizing from color patches to objects is difficult—we may make like a particular color only in a particular context, or in a particular functional situation (Schloss et al., 2013). Indeed, the spectral composition of the color could be the same in differing objects where a color is alternatively liked and disliked. We expect our food, for example, to have a predictable color appearance. Wheatley (1973), in a classic study, asked diners to enjoy a steak dinner in a room with dim lighting. When lighting was increased to a normal level partway through the meal, several diners reportedly became ill upon realizing that their otherwise delectable steak had been artificially colored blue (see Spence, 2015). So, whereas we may like a wavelength spectrum with increased energy in the short (blue) wavelengths when it is produced by, say, the sky, we will not necessarily like a similar collection of wavelengths when produced by something that does not normally appear blue. In any case, an understanding of the human species' most and least favorite colors is not necessarily a question that is usefully described using the language of statistics—this is in part because, even in a simple experiment, human responses can only be sampled for a limited number of the millions of discernible colors (Linhares et al., 2008); the combinatorics problem gets much worse when even simple color combinations are considered.

Artwork

Again bearing in mind the nonidentity of aesthetics and artwork, we can glean a meaningful understanding of relationships between color statistics and aesthetics by examining art. Empirical understanding of color statistics in art may be useful because Western artists in particular have historically paid close attention to natural color appearance and its vicissitudes, and they have invented diverse pigments for this purpose (though art traditions in the West and elsewhere have also greatly emphasized abstract color symbolism).

Montagner et al. (2016) used multispectral imaging to capture narrow-band measurements of wavelength content across the visible spectrum in Western painted artwork. They found that artists use more saturated red colors and fewer greens than are typical in natural scenes, but otherwise artists mostly matched natural colors. This result may be related to aesthetics, though other explanations are possible, such as the danger posed by standard green pigments, which were historically likely to contain arsenic (Zhao, Berns, Taplin, & Coddington, 2008), leading to their disuse. There are also potential biases in natural scene image databases, which may include disproportionate amounts of greenery.

However, further experiments using multispectral images of paintings suggest additional links between preference and color statistics. In an innovative study, Nascimento et al. (2017) investigated preference for rotations of the color gamut in paintings. Such modifications have the effect of changing the color of each pixel according to its location in a perceptually uniform color space, but maintaining the geometric relationships among all pixel colors; light intensity (luminance) and spatial relationships of pixels remain constant as well. Thus, during gamut rotation, pixels with similar hues in the original painting will stay similar to one another as their hues change together, whereas pixels with very different original hues will maintain large differences as their hues change. Multispectral imaging—which captures fine-grained wavelength information invisible to the eye—is required in this case to accurately simulate images with rotated color gamuts; see Figure 7. What is interesting with regard to empirical aesthetics is that when viewers are given a painting whose color gamut has been randomly rotated from its original state and are asked to rotate it according to their preference, they generally choose the original color axes. In particular, average viewer preference for the absolute colors in a given work of art is nearly indistinguishable perceptually from the preference for the original work created by the artist (Nascimento et al., 2017). This result holds for representational and highly abstract works, though, interestingly, preferred gamuts for naturalistic paintings were further away from the original colors compared with what was found for abstract works. This pattern holds for both art experts and laypersons. Thus, while artists do not necessarily all use the same approach to color harmony, each may find a maximally preferred color representation for her particular subject matter. However, preferred color gamuts do not fully align with what is typical in nature (nor, it was found, for represented materials of some aesthetic importance such as skin). Thus, from the point of view of color statistics, artists may in some respects prioritize aesthetic composition over natural rendering.

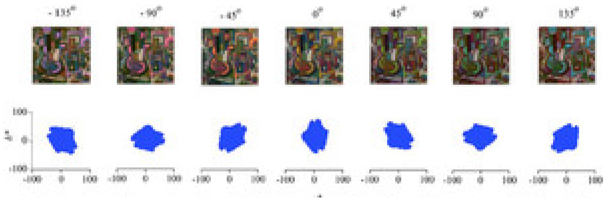

Figure 7. Experimental stimuli from Nascimento et al. (2017). The original painting is shown in the center image (0•). The color values for pixels in the painting in the CIELAB colorspace (with coordinates a* and b*) are depicted as the blue blob below the painting. As this blob is rotated through the colorspace to different angles of orientation, the corresponding pixels change in their a* and b* color values, though the overall spatial structure of the image remains the same. When humans viewers are able to rotate the blob themselves according to preference (from a random starting angle, and without prior exposure to the work), they overwhelmingly chose rotation values very close to 0•. This result holds for highly abstract works like the one shown as well as naturalistic paintings though, interestingly, preferred colors for naturalistic paintings were further away from the original colors compared to the result for abstract works.

These results represent evidence both of the utility of empirical aesthetics investigations using digitized artwork, andthe insights granted by statistical understanding of image features. As Nascimento et al. (2017) note, their conclusions are not predicted by the Palmer–Schloss model of preference for color patches in isolation.

In related experiments involving artworks and natural scenes, Nascimento and Masuda (2014) and Masuda and Nascimento (2012) used digital simulations of lighting changes as well as real-world lighting installations to investigate preferred illuminants. They found that humans preferred illumination to be rather different from what is typical in daylight: the preferred lighting spectra had more peaks and valleys across wavelength compared with natural daylight, which varies more smoothly. Also, humans preferred lighting that generated more saturated colors in the art and scenes compared with what is produced by daylight illumination.

Thus, while blue may be liked and brown disliked in general (due perhaps to these colors' associations with positive and negative valence natural substances, respectively), and while artists and viewers seem to agree on the most pleasing lighting and color relationships in artwork, there is variation in preference compared with what is natural—as with the preference for the distribution of luminances in scenes, described in the next section.

Regularities in Spatiotemporal Statistics

As we have seen, regularities of spatial, luminance, and color features can grant insight into empirical aesthetics, especially when considered in relation to natural regularities and the visual system mechanisms for encoding the retinal image. However, human vision is inherently dynamic: if one were to fully stabilize the retinal image, it would fade to nothingness (this so-called Troxler fading can be accomplished by physically affixing the image frame to the eyeball so that image and retina move together, or by mimicking the side-to-side jitter of the eyeball in a display; see, for example, Martinez-Conde et al., 2013).

Yet little research has been done to investigate spatiotemporal regularities from the point of view of empirical aesthetics. This may be because video stimuli take longer for viewers to evaluate compared with still images, and thus fewer can be shown during an experiment (though some researchers have developed novel methods for continuous evaluative responses, e.g., Muth et al., 2016). Many existing neuroscientific analyses of "natural" spatiotemporal patterns study instead Hollywood film (see Hasson et al., 2008) This is because it is very challenging to create natural movies that fully account for body, head, and eye movements, as well as accommodation and binocular disparity. Summarizing how statistical regularities (especially higher-order spatial regularities) in static image statistics vary over time is also challenging.

As soon as humans developed the capability of manipulating images over time, they very quickly deviated from what is natural. In particular, they invented the cut, whereby two different scenes or two different views of the same scene (i.e., shots) are ordered in sequence. But while the approaches filmmakers employ within cuts—as well as the sequence of information displayed across cuts—differ from those we experience in daily life, their frequency appears to have grounding in natural vision. Cutting et al. (2010) have found that cut length in Hollywood film generally follows a 1/*f* temporal pattern. That is, the length of a given cut is well predicted by the length of cuts nearby in time, and the overall the distribution of cut lengths follows a 1/*f* distribution. Cutting et al. (2011) found that, over film history, absolute cut length has gone down, while camera and object motion within cuts has gone up. These results may be partly due to technological and cultural changes, but the 1/*f* nature of cuts may also reflect variations in human attentional capacity over time, which are also found to exhibit 1/*f* scaling in time (Gilden, 2001).

Of perhaps greater relevance to this chapter is the finding that mean intensity (pixel value) of frames in Hollywood film has gone down over time (Cutting et al., 2011). A further study by Cutting (2014) showed that pixel values show roughly lognormal (positively skewed) distributions throughout film history, though more recent films have a greater proportion of low- and high-intensity pixels compared with earlier film (suggesting higher skewness for recent films, though skewness was not reported). Thus, films considered as a whole share a basic regularity of natural luminance distributions. However, because of variability in luminance nonlinearities in film and projection, as well as considerable scattering of projected light in a large theater (which would make “black” pixels brighter), the luminances received by the retina may not be fully described by pixel value distributions.

Despite producing spatiotemporal sequences using cuts that would never be experienced in nature, filmmakers have produced temporal statistical regularities that may take advantage of statistical regularities in visual attention over time. This is perhaps not so surprising since human eye movements are so rapid and frequent that our instantaneous view of the world is itself fragmentary. The retinal image continually experiences large shifts, though not in the same way as film cuts, which are typically structured around story-related functions rather than egocentric exploration. In any case, given the dominance of film in the aesthetic marketplace, it is clear that capturing attention in this way is effective. It also appears that film generally reproduces the intensity distribution of natural scenes, which feature mostly low intensities.

The Reliability of Aesthetic Responses

In this chapter, as in nearly all published research in empirical aesthetics, it has been assumed that human aesthetic responses can be accurately and consistently measured in humans using empirical methods. That is, when someone rates an image as a “5” in terms of liking on a 0–9 scale, we grant this measurement an implicit reliability. It is commonly acknowledged that individuals’ aesthetics vary considerably, and for reasons that are poorly understood (but see recent work by Vessel et al., 2018 and Leder et al., 2016, addressing individual differences in aesthetics for certain image classes). And very few studies have investigated the ways our aesthetics vary at different stages of development. Nevertheless, we tend to assume that any individual’s preferences are largely the same from one day or week to the next. Therefore, we assume that our measurements of their preferences are an accurate empirical gauge of an underlying psychological reality termed "aesthetics."

Yet humans at all stages of the lifespan have been found to be quite unreliable in their preferences for identical stimuli from week to week, with young children being most unstable. In the first study of its kind, Pugach et al. (2017) measured stability (i.e., test–retest reliability) in individuals age 3–99 using a suite of ranking tasks. Participants separately ranked images in four classes, containing face photographs, landscape photographs, and artistic paintings of the same faces and landscape features. They then repeated the same task 2 weeks later. In all age groups, participants made on average at least one rank change per image in each image class (young children made upwards of two rank changes per image).

Other studies, though not specifically concerned with stability/reliability, report correlations within observers of below .75 for aesthetic ratings of faces, natural images, and architecture over shorter intervals (Hönekopp, 2006). We can infer that, at most, around 50% of the variance of a given individual's preference for visual stimuli can be explained by their own previous ratings of the same stimuli. Indeed, even over very short (~15-min) intervals, individuals' aesthetic responses to the same stimuli can be rather different, with mean consistency measured at less than .9 for ratings of identical images (Vessel et al., 2018). This may in part be because of decreased liking with repeated exposures (Biederman & Vessel, 2006).

On the other hand, the same approach of studying stability/reliability has produced evidence that supports the notion of a "core" visual aesthetics, one that is surprisingly robust to brain damage. For example, in people with Alzheimer's related dementia and frontotemporal dementia, aesthetic stability for images is not significantly different from age-matched controls, though explicit memory is substantially worse in the diseased cohorts (Graham et al., 2013; Halpern et al. 2008; Halpern & O'Connor, 2013).

Thus, the study of aesthetic stability suggests that, on average, a hierarchy of basic preferences in the visual domain inheres in all individuals, even in the face of brain damage. However, the same research approach suggests that, outside of a fairly consistent aesthetic heuristic, there is considerable variability within all individuals in terms of their aesthetic responses to individual stimuli over the course of days or weeks. Fortuitously for researchers in empirical aesthetics, undergraduates are the most stable age cohort (Pugach et al., 2017). However, the variable nature of our aesthetic response for identical stimuli is a special concern when we try to discover empirical relationships between statistical features of images and measures of aesthetic response since statistical features in images are unchanging and devoid of context. Further, this problem is an impediment to a mechanistic neurobiological understanding of the relationships that may be observed between image statistics and aesthetics.

Summary

Given the evidence considered here, we can summarize how statistical features of images can illuminate the study of empirical aesthetics:

1) Humans prefer images with basic spatial statistics in a range around what is typical for natural scenes, and everybody dislikes images that have very different spatial statistics compared with what is typical in natural scenes, such as white noise and very blurry images. Irrespective of contextual factors, a $1/f$ power spectrum with slope around –2 is our default expectation for aesthetic images.
2) Smaller variation in slope around $p = -2$ is not known to be predictive of specific aesthetic responses. However, there is some evidence that artistic images from across cultures typically share distinctive higher-order statistical structure, though the influence of materials and composition cannot be ruled out.
3) Adults and children prefer high skew (i.e., less Gaussian) distributions of light intensity for isolated objects, fruits, and vegetables when there is associated glossiness. But more generally, we prefer lower-skew (i.e., more Gaussian) distributions of light intensity compared with what is natural. Handmade artwork is also low in skewness, though Hollywood films may have more highly skewed distributions of light intensity.
4) Observers prefer the way Western painters represent color compared with other color representations in a uniform space of color transformations. However, there are significant differences between natural color statistics and the preferred statistics of color use in artwork.
5) Temporal statistics of cuts in Hollywood films are also described by a $1/f$ distribution and therefore may fit human attentional capacity. However, cuts are themselves rather unnatural representations of spatiotemporal human vision.

Outlook: Where Do We Go From Here?

In conclusion, let us consider a variety of grand questions for future research.

Statistical Features: Comparative Approaches

Given the high similarity between human vision and other ape visual systems (as well as monkey visual systems, and indeed those of most other mammals), we can ask whether nonhuman animals experience the same generalized preferences for image statistical features such as natural scene-like power spectra. Our understanding of human visual aesthetics in the context of natural statistical regularities may tell us something about the “aesthetic primitives” that apply across many species. If other ape, primate, and mammal species show maximal preference and acuity for low-level natural scene-like regularities such as $1/f$ scaling, this would be strong evidence for the generality of the perceptibility hypothesis. As such, this finding would open new territory in the study of empirical aesthetics, since other species would be seen to operate on the same foundation of low-level visual biases or preferences. Studies of luminance statistics, color statistics, and other features in nonhuman animals may be profitable for similar reasons.

Spatiotemporal Features in Real Environments

Given the unnaturalness of film cuts, researchers have made efforts toward understanding more naturalistic aesthetic experiences that play out over time, such as museum visits (e.g., Pelowski et al., 2017). However, there is much yet to learn about how the statistics of spatiotemporal patterns of real-world visual stimulation relate to aesthetic experience. Indeed, as we go from considering single images to temporal sequences of images, the space of possible patterns of stimulation expands greatly. In relation to efficient coding of image statistical features, there is evidence that eye movements provide an important transformation of natural scenes (Kuang et al., 2012). However, accurately correlating eye movements in real environments to points of fixation in 3D space remains an unsolved problem.

Deep Learning

A new frontier in image statistics has opened in the recent acceleration in research on computer vision and machine learning systems that employ "deep" artificial neural networks. Such "deep learning" networks (e.g., convolutional neural networks) are powerful because they can learn higher-order spatial regularities of an input class of images. In particular, such systems find regularities by learning a set of basic representational units (i.e., spatial filters) that efficiently characterize commonalities of image structure within that image class. Basic spatial features are learned in lower levels of the network, while higher (deeper) levels use supervised learning to make more abstract associations between those features and semantic or conceptual categories. For example, such systems can learn to distinguish the stylistic category or authorship of patches extracted from artworks, and can match human performance on this task (see e.g., Gatys et al., 2016). It would not be inconceivable for such a system to learn to distinguish between aesthetic preferences of individual human observers, or to predict individual preference ratings in a given context.

However, the deep learning revolution has been countered by skeptics (e.g., Lake et al., 2017) who argue that a given deep learning network's performance is often poor when trained with a different set of images, applied to a different set of test images, or deployed to a related but distinct task. It is also impossible to interrogate a deep learning model to understand how it achieves a particular result since its "knowledge" consists entirely of thousands or millions of network weights, which together comprise a statistical model. That is, such models do not specify or describe mechanistic or functional relationships among real-world variables.

Thus, deep learning models that learn an individual's aesthetic taste might capture subtle higher-order idiosyncrasies in a particular set of images, and learn to associate these idiosyncrasies with what the subject reports to be their preference for a given tested image.

Deep learning might then be able to accurately predict the individual's aesthetic judgments of previously untested images from the same set. However, even highly accurate models of this kind would likely fail in a different but related task or context. In addition, such models do not contribute to explaining human aesthetic preferences or their neural underpinnings.

In contrast, what is powerful about considering natural regularities in image statistical features is that we can generate explanatory models—at least about constraints on aggregate taste—which align with efficient neurobiological mechanisms in human vision.

Acknowledgments

Thanks to Chris Redies, Andy Gartus, Patrick Markey, Helmut Leder, and Ed Vessel for helpful discussions.

References

Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of Vision*, *9*(11), 10.

Arce-Lopera, C., Masuda, T., Kimura, A., Wada, Y., & Okajima, K. (2012). Luminance distribution modifies the perceived freshness of strawberries. *i-Perception*, *3*(5), 338.

Atick, J. J. and Redlich A. N. 1992. What does the retina know about natural scenes? *Neural Computing*, *4*, 196–210.

Attewell, D., & Baddeley, R. J. (2007). The distribution of reflectances within the visual environment. *Vision Research*, *47*(4), 548–554.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychology Reviews*, *61*, 183–193.

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the 1/f noise. *Physical Review Letters*, *59*(4), 381.

Bell, A. J. and Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, *37*, 3327–3338.

Biederman, I., & Vessel, E. A. (2006). Perceptual pleasure and the brain: A novel theory explains why the brain craves information and seeks it through the senses. *American Scientist*, *94*(3), 247–253.

Brachmann, A., Barth, E., & Redies, C. (2017). Using CNN features to better understand what makes visual artworks special. *Frontiers in Psychology*, *8*, 830.

Brachmann, A., & Redies, C. (2017). Computational and experimental approaches to visual aesthetics. *Frontiers in Computational Neuroscience*, *11*, 102.

Brady, N., & Field, D. J. (2000). Local contrast in natural images: Normalisation and coding efficiency. *Perception*, *29*, 1–15.

Chandler, D. M., & Field, D. J. (2007). Estimates of the information content and dimensionality of natural scenes from proximity distributions. *The Journal of the Optical Society of America A*, *24*, 922–941.

Cutting, J. E. (2014). How light and motion bathe the silver screen. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(3), 340.

Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C., & Candan, A. (2011). Quicker, faster, darker: Changes in Hollywood film over 75 years. *i-Perception*, *2*(6), 569–576.

Cutting, J. E., DeLong, J. E., & Nothelfer, C. E. (2010). Attention and the evolution of Hollywood film. *Psychological Science*, *21*(3), 432–9.

Dror, R. O., Leung, T. K., Adelson, E. H., & Willsky, A. S. (2001). Statistics of real-world illumination. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. II–II.

Fernandez, D., & Wilkins, A. J. (2008). Uncomfortable images in art and nature. *Perception*, 37, 1098–1113.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, *6*(4), 559–601.

Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research*, *94*, 62–75.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2414–2423.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167–192.

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review*, *108*(1), 33.

Graham, D. J. (2011). Visual perception: Lightness in a high dynamic range world. *Current Biology*, *21*(22), R914–R916.

Graham, D. J., Chandler, D. M., & Field, D. J. (2006). Can the theory of “whitening” explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research*, *46*, 2901–2913.

Graham, D. J., & Field, D. J. (2007). Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spatial Vision*, *21*, 149–164.

Graham, D. J., & Field, D. J. (2008a). Variations in intensity statistics for representational and abstract art, and for art from the eastern and western hemispheres. *Perception*, *37*, 1341–1352.

Graham, D. J., & Field, D. J. (2008b). Global nonlinear compression of natural luminances in painted art. *Computer Image Analysis in the Study of Aart*, *6810*, 68100K.

Graham, D. J., Friedenberg, J. D., McCandless, C. H., & Rockmore, D. N. (2010). Preference for artwork: Similarity, statistics, and selling price. *Proceedings of SPIE: Human Vision and Electronic Imaging*, *7527*, 75271A.

Graham, D., & Meng, M. (2011a). Altered spatial frequency content in paintings by artists with schizophrenia. *i-Perception*, *2*(1), 1–9.

Graham, D. J., & Meng, M. (2011b). Artistic representations: clues to efficient coding in human vision. *Visual Neuroscience*, *28*(4), 371–379.

Graham, D.J., Pallett, P. M., Meng, M., & Leder, H. (2014). Representation and aesthetics of the human face in portraiture. *Art & Perception*, *2*(1–2), 75–98.

Graham, D. J., & Redies, C. (2010). Statistical regularities in art: Relations with visual coding and perception. *Vision Research*, *50*(16), 1503–1509.

Graham, D., Schwarz, B., Chatterjee, A., & Leder, H. (2016). Preference for luminance histogram regularities in natural scenes. *Vision Research*, *120*, 11–21.

Graham, D., Stockinger, S., & Leder, H. (2013). An island of stability: art images and natural scenes–but not natural faces–show consistent esthetic response in Alzheimer's-related dementia. *Frontiers in Psychology*, *4*, 107.

Grebenkina, M., Brachmann, A., Bertamini, M., Kaduhm, A., & Redies, C. (2018). Edge-orientation entropy predicts preference for diverse types of man-made images. *Frontiers in Neuroscience*, *12*, 678.

Halpern, A. R., Ly, J., Elkin-Frankston, S., & O'Connor, M. G. (2008). "I know what I like": stability of aesthetic preference in Alzheimer's patients. *Brain and Cognition*, *66*(1), 65–72.

Halpern, A. R., & O'Connor, M. G. (2013). Stability of art preference in frontotemporal dementia. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 95.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections*, *2*(1), 1–26.

Hönekopp, J. (2006). Once more: is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 199–209.

Jones-Smith, K., & Mathur, H. (2006). Fractal analysis: Revisiting Pollock's drip paintings. *Nature*, *444*(7119), E9.

Jung, S. K., Lee, J. H., Kakizaki, H., & Jee, D. (2012). Prevalence of myopia and its association with body stature and educational level in 19-year-old male conscripts in Seoul, South Korea. *Investigative Ophthalmology & Visual Science*, *53*(9), 5579–5583.

Ke, Y., Tang, X., & Jing, F. (2006). The design of high-level features for photo quality assessment. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, *1*, 419–426.

Knill, D. C., Field, D., & Kersten, D. (1990). Human discrimination of fractal images. *The Journal of the Optical Society of America A*, *7*(6), 1113–1123.

Kuang, X., Poletti, M., Victor, J. D., & Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Current Biology*, *22*(6), 510–514.

Kube, P., & Pentland, A. (1988). On the imaging of fractal surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *10*(5), 704–707.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253. doi:10.1017/S0140525X16001837

Laughlin, S. B. (1981) A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch*, 36c, 910–912.

Leder, H., Goller, J., Rigotti, T., & Forster, M. (2016). Private and shared taste in art and face appreciation. *Frontiers in Human Neuroscience*, *10*, 155. doi:10.3389/fnhum.2016.00155

Linhares, J. M. M., Pinto, P. D., & Nascimento, S. M. C. (2008). The number of discernible colors in natural scenes. *The Journal of the Optical Society of America A*, *25*(12), 2918–2924.

Mandelbrot, B. B. (1982). *The fractal geometry of nature* (Vol. 1). New York: WH Freeman.

Martinez, A. M., & Benavente, R. (1998). The AR face database. CVC Technical Report, 24.

Masuda, O., & Nascimento, S. M. (2012). Lighting spectrum to maximize colorfulness. *Optics Letters*, *37*(3), 407–409.

Mather, G. (2013). *The psychology of visual art: Eye, brain and art*. Cambridge, England: Cambridge University Press.

Mather, G. (2018). Visual Image Statistics in the History of Western Art. *Art & Perception, 6*(2–3), 97–115.

Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: Towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, *14*, 83–96.

McManus, I. C., Stöver, K., & Kim, D. (2011). Arnheim's Gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images. *i-Perception*, *2*(6), 615–647.

Montagner, C., Linhares, J. M., Vilarigues, M., & Nascimento, S. M. (2016). Statistics of colors in paintings and natural scenes. *The Journal of the Optical Society of America A*, *33*(3), A170–A177.

Muth, C., Raab, M. H., & Carbon, C. C. (2016). Semantic stability is more pleasurable in unstable episodic contexts. On the relevance of perceptual challenge in art appreciation. *Frontiers in Human Neuroscience*, *10*, 43.

Nadal, M., & Skov, M. (2018). The pleasure of art as a matter of fact. *Proceedings of the Royal Society London B: Biological Sciences*, *285*(1875), 20172252.

Nascimento, S. M., Linhares, J. M., Montagner, C., João, C. A., Amano, K., Alfaro, C., & Bailão, A. (2017). The colors of paintings and viewers' preferences. *Vision Research*, *130*, 76–84.

Nascimento, S. M. C., & Masuda, O. (2014). Best lighting for visual appreciation of artistic paintings—experiments with real paintings and real illumination. *The Journal of the Optical Society of America A*, *31*(4), A214–A219.

Normant, F., & Tricot, C. (1991). Method for evaluating the fractal dimension of curves using convex hulls. *Physical Review A*, *43*(12), 6518.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

Otaka, H., Shimakura, H., & Motoyoshi, I. (2019). Perception of human skin conditions and image statistics. *Journal of the Optical Society of America A, 36*, 1609–1616.

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.

Pelowski, M., Forster, M., Tinio, P. P., Scholl, M., & Leder, H. (2017). Beyond the lab: An examination of key factors influencing interaction with "real" and museum-based art. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(3), 245.

Prescott, N., & Graham, D. J. (2020). Face Representation and Perceptual Processing in Mask-making. PsyArXiv, doi:10.31234/osf.io/mdpaz<**http://osf.io/mdpaz**> Retrieved from psyarxiv.com/mdpaz<**http://psyarxiv.com/mdpaz**>

Pugach, C., Leder, H., & Graham, D. J. (2017). How stable are human aesthetic preferences across the lifespan? *Frontiers in Human Neuroscience*, *11*, 289.

Redies, C., & Brachmann, A. (2017). Statistical image properties in large subsets of traditional art, bad art, and abstract art. *Frontiers in Neuroscience*, *11*, 593.

Redies, C., Brachmann, A., & Wagemans, J. (2017). High entropy of edge orientations characterizes visual artworks from diverse cultural backgrounds. *Vision Research*, *133*, 130–144.

Redies, C., Hasenstein, J., & Denzler, J. (2007a). Fractal-like image statistics in visual art: Similarity to natural scenes. *Spatial Vision*, *21*(1), 137–148.

Redies, C., Hänisch, J., Blickhan, M., & Denzler, J. (2007b). Artists portray human faces with the Fourier statistics of complex natural scenes. *Network: Computation in Neural Systems*, *18*(3), 235–248.

Rhodes, G., Peters, M., Lee, K., Morrone, M. C., & Burr, D. (2005). Higher-level mechanisms detect facial symmetry. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*(1570), 1379–1384.

Rogowitz, B. E. and Voss, R. F. (1990). Shape perception and low-dimensional fractal boundaries. *Proceedings of SPIE*, *1249*, 387–394.

Ruderman, D. L. (1994). The statistics of natural images. *Network: Computation in Neural Systems*, *5*(4), 517–548.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters, 73*(6), 814–817.

Schloss, K. B., Strauss, E. D., & Palmer, S. E. (2013). Object color preferences. *Color Research & Application*, *38*(6), 393–411.

Schweinhart, A. M., & Essock, E. A. (2013). Structural content in paintings: Artists overregularize oriented content of paintings relative to the typical natural scene bias. *Perception*, *42*(12), 1311–1332.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*(1), 1193–1216.

Soille, P., & Rivest, J. F. (1996). On the validity of fractal dimension measurements in image analysis. *Journal of Visual Communication and Image Representation*, *7*(3), 217–229.

Spehar, B., Wong, S., van de Klundert, S., Lui, J., Clifford, C. W. G., & Taylor, R. (2015). Beauty and the beholder: The role of visual sensitivity in visual preference. *Frontiers in Human Neuroscience*, *9*, 514.

Spence, C. (2015). On the psychological impact of food colour. *Flavour*, *4*(1), 21.

Switkes, E., Mayer, M. J., & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: Anisotropyand the carpentered environment hypothesis. *Vision Research*, *18*, 1393–1399.

Taylor, R. P., Micolich, A. P., & Jonas, D. (1999). Fractal analysis of Pollock's drip paintings. *Nature*, *399*(6735), 422.

Theiler, J. (1990). Estimating fractal dimension. *The Journal of the Optical Society of America A*, *7*(6), 1055–1073.

Van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society B: Biological Sciences*, *265*, 359–366.

Vessel, E. A., Maurer, N., Denker, A. H., & Starr, G. G. (2018). Stronger shared taste for natural aesthetic domains than for artifacts of human culture. *Cognition*, *179*, 121–131.

Viengkham, C., & Spehar, B. (2018). Preference for fractal-scaling properties across synthetic noise images and artworks. *Frontiers in Psychology*, *9*, 1439. doi:10.3389/fpsyg.2018.01439

Wada, Y., Arce-Lopera, C., Masuda, T., Kimura, A., Dan, I., Goto, S. I., Tsuzuki, D. & Okajima, K. (2010). Influence of luminance distribution on the appetizingly fresh appearance of cabbage. *Appetite*, *54*(2), 363–368.

Wheatley, J (1973) Putting colour into marketing. *Marketing*, *October*, 24–29, 67.

Yang, J., Otsuka, Y., Kanazawa, S., Yamaguchi, M. K., & Motoyoshi, I. (2011). Perception of surface glossiness by infants aged 5 to 8 months. *Perception*, *40*(12), 1491–1502.

Zhao, Y., Berns, R. S., Taplin, L. A., & Coddington, J. (2008, February). An investigation of multispectral imaging for the mapping of pigments in paintings. *Computer Image Analysis in the Study of Art*, *6810*, 681007).

Notes:

(1.) Curiously, we perceive white noise images as having only high frequency structure, despite having equivalent amounts of low, medium, and high frequency structure.

Daniel Graham

Department of Psychology, Hobart & William Smith Colleges