

# Dichoptic difference thresholds for uniform color changes applied to natural scenes

Ali Yoonessi

McGill Vision Research Unit, Montréal, Québec, Canada



Frederick A. A. Kingdom

McGill Vision Research Unit, Montréal, Québec, Canada



It has recently been shown that the visual system is more sensitive to uniform color and/or luminance changes applied to raw compared to phase-scrambled images of natural scenes (A. Yoonessi & F. A. A. Kingdom, 2008). Here we consider whether the mechanisms responsible for the differential sensitivity operate before or after the point at which the signals from the two eyes are combined. Knowing this should help determine the types of nonlinearities responsible. Thresholds for detecting uniform color transformations applied to raw and phase-scrambled natural scenes were measured under two conditions: monocular, in which the discriminand pairs were placed side by side, and dichoptic, in which they were dichoptically superimposed. Subjects were required to select the pair of images that were transformed from two pairs of images in which the other pair was untransformed. In the dichoptic condition, the transformed image pair was identifiable by its lustrous appearance. In line with our previous findings, thresholds in the monocular condition were higher for the phase-scrambled compared to raw scenes. However in the dichoptic condition there was no significant difference between raw and phase-scrambled thresholds, suggesting that the differential sensitivity was mediated by mechanisms lying beyond the point of binocular combination. It is suggested that cortical neurons sensitive to edges but suppressed by neighboring texture might be responsible for the higher sensitivity to transformations applied to raw compared to phase-scrambled images of natural scenes.

Keywords: natural scenes, phase scrambling, dichoptic, color vision

Citation: Yoonessi, A., & Kingdom, F. A. A. (2009). Dichoptic difference thresholds for uniform color changes applied to natural scenes. *Journal of Vision*, 9(2):3, 1–12, <http://journalofvision.org/9/2/3/>, doi:10.1167/9.2.3.

## Introduction

Color vision has been studied traditionally using simple laboratory stimuli such as patches, Gabors, and gratings. Only recently have images of natural scenes been used to study color vision (e.g., Brainard, Rutherford, & Kraft, 1997; Fine & MacLeod, 2001; Johnson & Baker, 2004; Párraga, Troscianko, & Tolhurst, 2002; Ruderman & Bialek, 1994; Webster & Mollon, 1997; Yoonessi & Kingdom, 2008). Although the complexity of natural scenes makes the data obtained from them sometimes difficult to interpret, they offer a unique opportunity to study how the structural properties of the natural visual environment influence color perception.

In a recent study we measured human sensitivity to a range of color transformations applied uniformly across images of natural scenes (Yoonessi & Kingdom, 2008). The transformations were rotations and translations applied to the color space that defined all pixel values. We found that for all types of transformation, sensitivity was higher for the raw images of natural scenes compared to their phase-scrambled counterparts. Control experiments ruled out that the differential sensitivity was due to the familiarity of the colors in the raw scenes, suggesting instead that it was due to the raw scene's unique spatial structure. Raw natural scenes typically consist of patterns

of edges separated by uniform regions, and we conjectured that this was the critical features.

One of the issues raised by Yoonessi and Kingdom's (2008) study concerns the locus of the mechanisms responsible for the higher raw-scene sensitivity. Assuming that the detection of uniform color changes is a relatively low-level visual process, one that could in principle be mediated by monocular as well as binocular neurons, a legitimate question is whether the higher sensitivity is mediated by monocular or binocular mechanisms. The principle aim of the present study is to answer this question. Knowing the locus of the differential sensitivity could be useful in determining the nonlinearities that are responsible.

The retinal images in the two eyes are very similar under normal viewing conditions, the slight differences between them arising from retinal disparity. Normally the between-eye differences due to disparity are not perceived as differences or anomalies—instead they are exploited by stereopsis to provide an impression of a unitary three-dimensional world. However, artificial differences between stimuli presented to the two eyes can result in departures from unitary vision, of which rivalry, is the most commonly studied form (Blake, 2001). Recently, Malkoc and Kingdom (2004) described a new measure of non-unitary binocular perception: the ‘Dichoptic Difference Threshold,’ or DDT. The DDT is the minimum

detectable difference between two dichoptically superimposed stimuli. The DDT is a performance-based rather than appearance-based measure. If the difference between a dichoptic pair is gradually increased from zero, a point is reached where the stimulus takes on a slightly lustrous appearance, and it is this that enables it to be distinguished from dichoptically identical stimuli. Figure 1 illustrates the effect. If one free-fuses the two pairs of stimuli, the bottom, dichoptically different stimulus should appear lustrous. Because the lustrous appearance occurs at much smaller between-eye differences than are required to elicit rivalry, DDTs are much lower than thresholds for binocular rivalry (Malkoc & Kingdom, 2004).

DDTs offer a simple means to determine whether the positive effects of natural scene structure on sensitivity to color transformations are mediated by mechanisms at an early stage within monocular channels, or after the signals from the two eyes are combined. If they are mediated within monocular channels, we would expect to find a similar pattern of results for DDTs as for discriminands in plain view i.e. lower thresholds for raw compared to phase-scrambled scenes. On the other hand if they are mediated by channels after the point of binocular



Figure 1. Sample stimuli presented in a single forced-choice trial. The top pair is identical, original images. The bottom pair has been transformed by rotation along the red–green axis in equal and opposite directions. In the monocular condition, the two pairs are shown sequentially, and the subject has to choose the pair that is different. In the dichoptic condition, the two stimuli in each pair are dichoptically superimposed; otherwise the task is the same. When free-fused the bottom stimulus should appear lustrous, which is the cue for ‘different’.

combination, then we would not expect a difference between raw and phase-scrambled DDTs. The present study will test between these two alternatives.

What are the color transformations that we have employed? They are luminance and/or chromatic changes applied uniformly across the image. The transformations are implemented by rotating or translating the three-dimensional color space defining the colors and luminances of every pixel in the image. Sample transformations applied to an image are shown in Figure 2. The color space employed here is a modified version of the MacLeod–Boynton color space (MacLeod & Boynton, 1979) designed by Ruderman, Cronin, and Chiao (1998). The axes in the color space represent the responses of three postreceptoral channels: the luminance-sensitive channel that sums the outputs of the L (long-wavelength-sensitive) and M (medium-wavelength-sensitive) cones; a chromatically sensitive channel that differences the outputs of the L and M cones and is known as the ‘L–M’ channel; a chromatically sensitive channel differences the sum of the L and M cone responses from the S (short-wavelength-sensitive) cone response and is known as the ‘S–(L+M)’ channel. Because these channels are often (though strictly speaking incorrectly) referred to as the ‘luminance,’ ‘red–green,’ and ‘blue–yellow’ channels, we will employ this terminology. In one form of MacLeod–Boynton color space the three postreceptoral channel axes are formed by appropriate combinations of cone contrast, where cone contrast is defined as  $\Delta L/L_b$ ,  $\Delta M/M_b$ , and  $\Delta S/S_b$ . The denominator in each cone contrast term is the cone response to the background, which is assumed to determine the state of cone adaptation. While this is a reasonable assumption for briefly presented stimuli such as gratings, or low contrast patches, it is arguably inappropriate for natural scenes that tend to be of high contrast and for which cone adaptation is likely determined locally rather than across the scene as a whole (Brown & Masland, 2001; Ledda, Santos, & Chalmers, 2004; Shapley & Hawken, 2002; Wallach, 1948). A particularly undesirable consequence of using conventional cone contrast to represent the three postreceptoral channel layers of natural scenes is that the red–green layer spuriously picks up pure-luminance shadows (Olmos & Kingdom, 2004b; Párraga et al., 2002). The use of logarithmic-based cone contrasts is one way to avoid this problem (Olmos & Kingdom, 2004b; Ruderman et al., 1998).

For the experiments described below we used 50 images of natural scenes—termed ‘raw’—and their phase-scrambled versions. Each image was decomposed into three layers based on the modeled responses of the luminance, red–green, and blue–yellow postreceptoral channels. Each layer was then transformed by translation and rotation. Thresholds for detecting the transformations were measured under four conditions:

1. raw scenes, with the discriminand pair placed side by side and viewed monocularly;

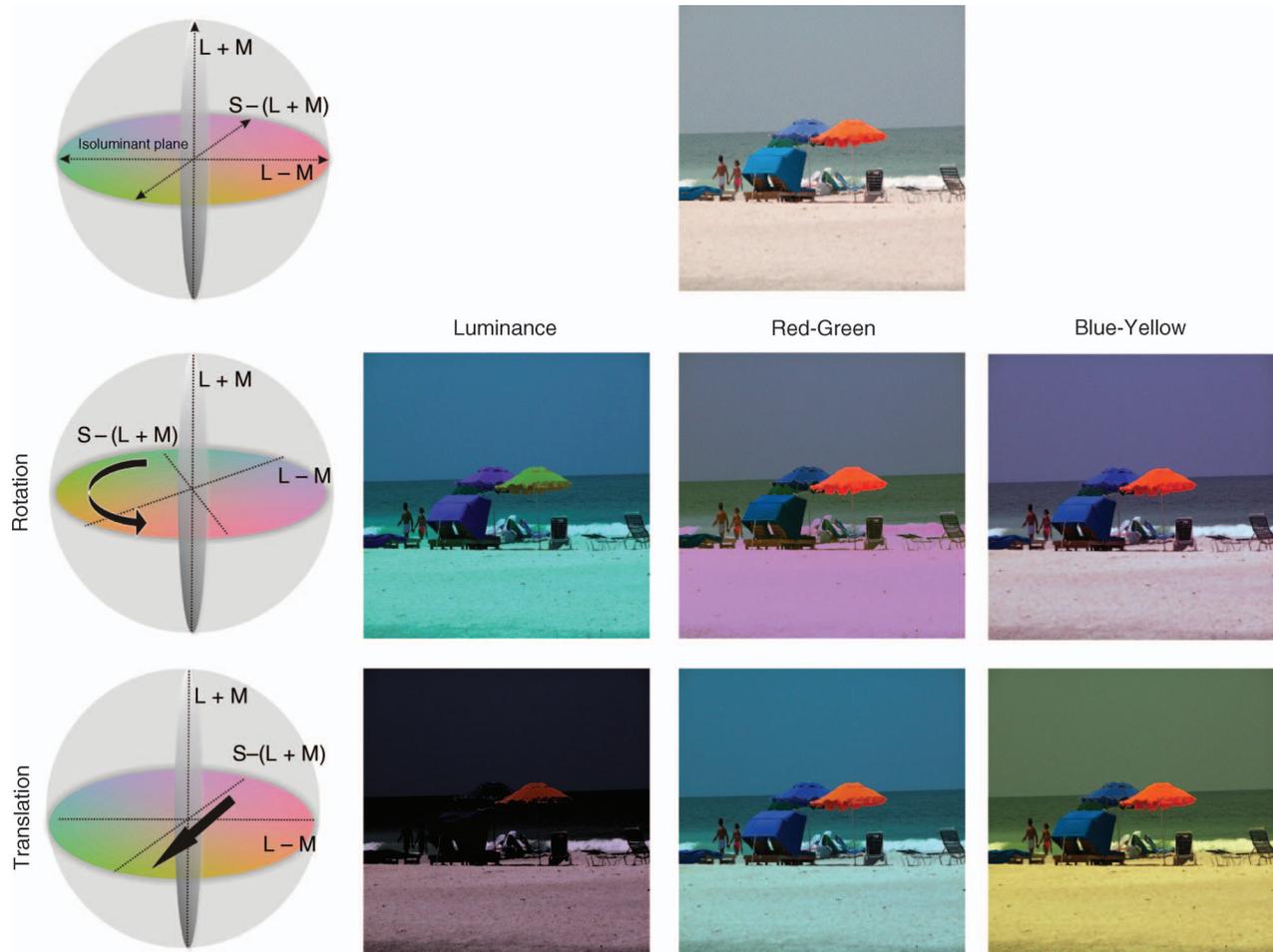


Figure 2. Transformations applied to the color space of an image of a natural scene (top). The two types of transformation are illustrated on the left, showing rotation applied to the luminance axis and translation to the blue–yellow axis. On the right the results of applying the transformations to the three axes of the color space of a natural scene are shown.

2. phase-scrambled scenes, with the discriminand pair placed side by side and viewed monocularly;
3. raw scenes, with the discriminand pair superimposed dichoptically;
4. phase-scrambled scenes, with the discriminand pair superimposed dichoptically.

In order to compare the various types of luminance/chromatic transformations, we have measured thresholds defined in terms of a simple and intuitively appealing metric of image distance: the Euclidean distance, or  $L_2$  norm.  $E$  can be calculated using the following formula:

$$E = \sqrt{\frac{\sum_{n=1}^N \sum_{i=1}^3 (p_{ni} - q_{ni})^2}{3N}}, \quad (1)$$

where  $p_{ni}$  and  $q_{ni}$  are the intensities of the corresponding pixels in the two images, with  $i$  being the image layer ( $i = 1:3$ ),  $n$  the pixel (i.e., with unique  $x, y$  coordinate), and  $N$  the number of pixels per image. Euclidean distance has

the important property that it defines a straightforward measure of the distance between two images, that is the same answer irrespective of the orthonormal basis used to represent the images, e.g., pixels, Fourier, Haar, etc. (Horn & Johnson, 1985). Euclidean distance has been previously employed to compare sensitivities to a variety of transformations applied to natural scenes (Kingdom, Field, & Olmos, 2007). It is important to state at the outset however that we are not arguing that Euclidean distance is the proper *perceptual* metric. Rather, we argue that  $E$  is a relatively neutral metric, providing a useful measure for comparing the relative sensitivities to the different types of chromatic/luminance transformations that we have used.

## Methods

### Subjects

Five subjects were employed, the two authors, and three persons who were naive as to the purpose of the experi-

ment. All had normal or corrected-to-normal vision. Color vision was tested using the Ishihara plates.

## Equipment and calibration

The scenes were photographed with a Nikon CoolPix-7500 digital camera and displayed on a Sony FD Trinitron 17", GDM F-500 using the VSG graphics board (Cambridge Research Systems) housed in a 2.8-GHz PC computer. The monitor RGB phosphors were gamma-corrected after calibration using an Optical photometer (Cambridge Research Systems). The spectral emission functions of the three phosphors were measured using an Optikon SpectroScan® PR 645 spectrophotometer, with the monitor screen filled with red, green, or blue, in the range of 400 to 700 nm at 10-nm intervals. Monitor resolution was 640 \* 480 with a refresh rate of 100 MHz. Matlab version 7 was used for all image processing tasks.

The cameras were calibrated as follows. Each one of a set of gray Munsell papers was illuminated by an incandescent light with a constant DC power, and photographed. Additionally, the luminance of the light reflected from each paper was measured with a Topcon SR-1 spectroradiometer. The average R, G, and B pixel values were plotted against the corresponding measured luminance and fitted with the following function:  $L = a(b^s + 1)$ , where  $L$  is luminance,  $s$  is the pixel level value obtained for each of the camera sensors (R, G, and B), and  $b$  is a constant that determines the slope of the curve. In addition, a white target was photographed through a series of narrowband optical interference filters from 400 to 700 nm at 10-nm intervals. Each R, G, and B value was recorded, gamma-corrected, and used to construct a spectral sensitivity function for each sensor, which was then normalized to produce equal responses to a flat-spectrum light.

For the dichoptic condition, the two side-by-side images on the monitor were brought into binocular registration using a custom-built 8-mirror Wheatstone stereoscope.

## Images

The gamma-corrected camera RGBs were mapped onto gamma-corrected monitor RGBs using a  $3 \times 3$  linear transformation matrix. The coefficients in the matrix were device specific and were chosen to produce as faithful a reproduction of the image colors as possible, using a method described elsewhere (Yoonessi & Kingdom, 2007).

Fifty ‘everyday’ scenes, representing a range of natural environments (forests, mountains, flowers, and fruits) and urban scenes (buildings, traffic signs, man-made objects), photographed under a variety of different illumination conditions (sunny and cloudy) and at a

variety of distances (0.5 m–1000 m), were taken from the McGill Calibrated Color Image database (Olmos & Kingdom, 2004a). The images were photographed by the camera (see above) and stored as uncompressed Tagged Image File Format (TIFF) files with resolution  $1920 \times 2560$  pixels and color depth of 24 bits (256 levels for each R, G, and B image). The camera’s smallest aperture setting ( $f$  7.4) was chosen to capture the images with minimum within-image differences in focus. Then, the images were resized to  $147 \times 147$  pixels wide using the nearest-neighbor interpolation algorithm and converted to bitmap file format with 24-bit depth. Each image subtended  $9 \times 9$  cm (11.4 degrees of visual angle) on the monitor at the viewing distance of 45 cm for both monocular and dichoptic conditions.

## Stimuli

### Phase-scrambled images

Phase randomization was implemented by discrete Fourier transform of the images. The phase and amplitude was calculated by following formulas:

$$\begin{aligned} \text{Amplitude} &= \sqrt{[F_r(\omega_x, \omega_y)]^2 + [F_i(\omega_x, \omega_y)]^2} \\ \text{Phase} &= \arctan \left[ \frac{F_i(\omega_x, \omega_y)}{F_r(\omega_x, \omega_y)} \right], \end{aligned} \quad (2)$$

where  $\omega_x$  and  $\omega_y$  are frequency variables and  $F_r$  and  $F_i$  are the real and imaginary parts of each Fourier frequency component of the spectrum. A random number between  $-\pi$  and  $+\pi$  was generated for each Fourier component and added to original phases of all three image planes. An inverse Fourier transform then returned the phase-scrambled image.

### Conversion of stimuli from RGB to LMS color space

Using the spectral sensitivities of the camera sensors and the sensitivities of the L (long-), M (middle-), and S- (short-wavelength-sensitive) cones from Smith and Pokorny (1975), a conventional  $3 \times 3$  linear matrix was used to convert the RGB camera values to LMS cone excitations (Kingdom et al., 2007).

### Color space and postreceptoral layers

A modified version of the Ruderman color space was used to model the three layers of human vision (Ruderman et al., 1998). Cone contrasts for each pixel were defined as

$$\begin{aligned} L_C &= \log L - \log \bar{L} \\ M_C &= \log M - \log \bar{M} \\ S_C &= \log S - \log \bar{S}, \end{aligned} \quad (3)$$

where  $\log L$ ,  $\log M$ , and  $\log S$  are log pixel cone excitations and  $\log \bar{L}$ ,  $\log \bar{M}$ , and  $\log \bar{S}$  are log pixel cone excitations averaged across the image. The three post-receptor responses to each pixel were defined as

$$\begin{aligned}\hat{l} &= (r\hat{L}_C + \hat{M}_C) \\ \hat{\alpha} &= (\hat{L}_C + \hat{M}_C - 2\hat{S}_C) \\ \hat{\beta} &= (\hat{L}_C - \hat{M}_C),\end{aligned}\quad (4)$$

where  $\hat{l}$ ,  $\hat{\alpha}$ ,  $\hat{\beta}$  are the luminance, ‘blue–yellow,’ and ‘red–green’ axes, respectively.  $r$  is a parameter that determines the relative L and M cone contrast inputs to the luminance mechanism and varies between observers. We determined  $r$  as described below.

### Image transformations

We applied two basic transformations to three axes respectively: translation and rotation. All transformations were affine, that is, all points lying on a line remained on the line after the transformation, and ratios of distances (Weisstein, 2004). Six levels of each transformation, i.e., six levels of  $E$  were employed, and these were determined through pilot experiments. The size levels of  $E$  were logarithmically spaced.

## Psychophysical procedures

### Scaling of axes

To compare the results of transformations applied to different layers of the color space, it was necessary to ‘equate’ the layers. We asked subjects to adjust the contrasts of the luminance and Red–Green layers to match that of the Blue–Yellow layer in five random images. Since the Blue–Yellow layer has the least contrast perceptually, we chose it as the base contrast for this procedure, in order to avoid exceeding the monitor range.

### Isoluminant setting

Isoluminance was determined using the method of minimum distinct border (Boynton, 1973; Yoonessi & Kingdom, 2008). Subjects altered the ratio of L to M in the ‘red–green’ layer of five natural images, in which the contrasts of the luminance and ‘blue–yellow’ layers were set to zero, until the image appeared to have least sharp borders. The measure  $r$  (Equation 4) was averaged across five images from the natural scene set, for each subject.

### Main task

In both monocular and dichoptic conditions, four images were presented on the monitor on each trial. Each

set of four consisted of two original and two transformed images. The four images were presented in two successive pairs in a conventional two-interval forced-choice procedure. In the monocular condition, each pair of images was presented side by side on the screen with a center-to-center separation of 13.5 cm (8.6 degrees). This is illustrated in Figure 1. In the dichoptic condition, each pair was superimposed dichoptically using the stereoscope (Figure 1—free-fused). In one of the forced-choice intervals the pair of images were both originals while in the other forced-choice interval the pair were transformed in equal and opposite directions. For example, in the translation-in-red–green condition, one of the transformed images was shifted toward red and the other shifted toward green, both by the same amount. The transformed images were randomly presented in either the first or second interval. In the monocular condition, the subject was required to report which interval contained the pair that looked different. In the dichoptic condition the subject was required to report which stimulus appeared lustrous or abnormal. The inter-stimulus pair interval was 200 ms and each image was displayed on the monitor for 250 ms. A tone provided feedback for an incorrect response. During each session, the type of transformation and layer that was transformed was fixed, while the level of transformation was selected randomly. There were 150 trials per session, and two sessions per condition, giving a total of 300 trials per condition. Viewing distance was 45 cm for both conditions (in the dichoptic condition this was the length along the light path in the stereoscope).

## Analysis

On every trial the Euclidean distance  $E$  of the transformed image was recorded along with the subject’s response (correct or incorrect). Although there were 6 discreet levels for each transformation, the computed values of  $E$  for each level of a given transformation varied according to image. In order to fit psychometric functions to the data, the  $E$ s were divided into 6 ‘bins’ for each transformation. The first bin was set to have a minimum of zero, while the last, sixth bin was set to have a maximum equal to the maximum  $E$  found for that transformation. The first bin ‘divider’ was determined iteratively to be that which minimized the between-bin variance in the number of trials when the remaining bin dividers were logarithmically spaced. This method ensured that the trials were distributed as evenly as possible between bins while obeying the constraint that all except the first bin were logarithmically spaced ( $E$ s in the first bin began at zero). After the  $E$ s were binned, the mean  $E$ s, proportion correct, and number of trials were calculated for each bin. Psychometric functions were fitted using psignifit version 2.5.6, which uses the maximum-likelihood method described by Wichmann and Hill (2001).

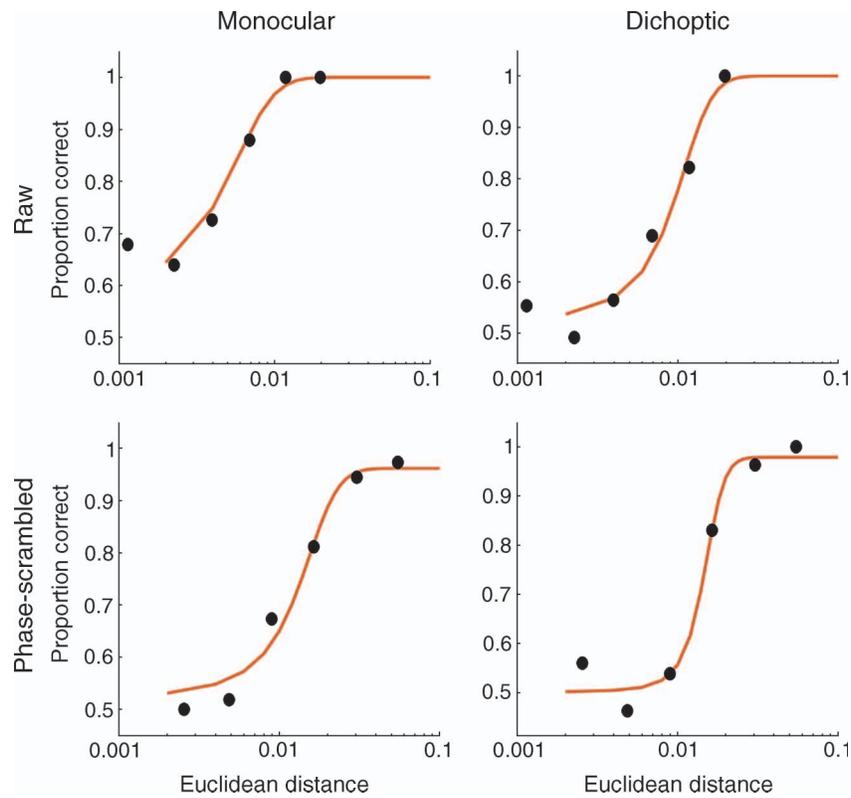


Figure 3. Psychometric functions for translation in red–green for subject AY, for both monocular and dichoptic conditions and for both raw and phase-scrambled scenes.

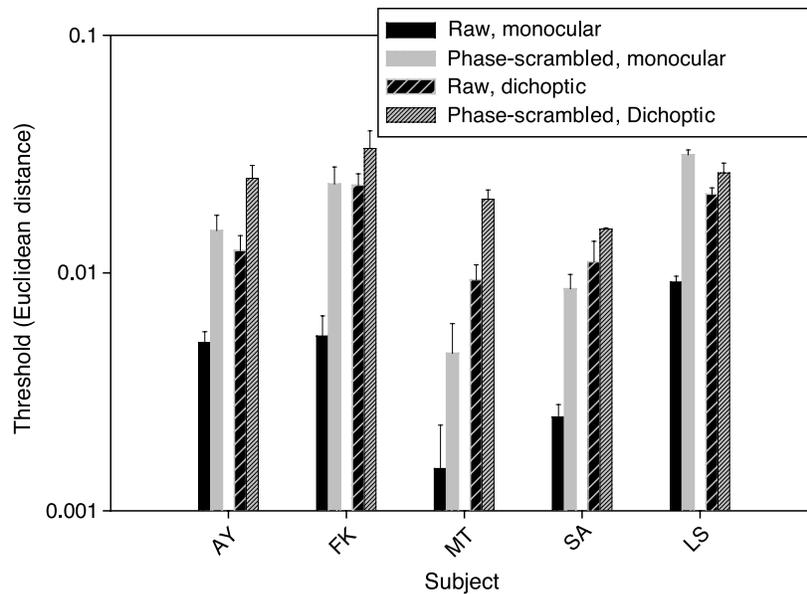


Figure 4. Euclidean distance thresholds for detecting rotation along the blue–yellow axis, for the four combinations of monocular/dichoptic and raw/phase-scrambled. Data for five subjects are shown.

## Results

Example psychometric functions for translation of the red–green layer, for both raw and phase-scrambled scenes and for both monocular and dichoptic presentations, are shown in Figure 3. Each plot gives the overall proportion of correct trials as a function of  $E$ . The threshold was calculated at the 75% correct level (see Methods section).

Thresholds for five subjects for one of the transformations, rotation along blue–yellow axis, are shown in Figure 4. In this figure, DDTs are lower for raw than phase-scrambled scenes for both dichoptic and monocular conditions, but the difference in thresholds is bigger for the dichoptic condition. If one pools the data from all the subjects and all the transformations, average thresholds in the monocular condition are  $1.43 E - 2$  for raw and  $1.99 E - 2$  for phase-scrambled scenes, whereas in the dichoptic condition average thresholds are  $3.52 E - 2$  for raw and  $3.85 E - 2$  for phase-scrambled scenes. Using a within-subject two-tailed  $t$ -test and the  $p < 0.05$  criterion, the difference between raw and phase-scrambled in the monocular condition was significant ( $t = 2.80$ ,  $df = 54$ ,  $p = 0.007$ ), but in the dichoptic condition it was not significant ( $t = 1.33$ ,  $df = 54$ ,  $p = 0.38$ ). In fact the differences between raw and phase-scrambled in the dichoptic condition shown in Figure 4 were the largest we found. This can be seen in the summary of the results

shown in Figure 5, which plots the *difference* between the raw and phase-scrambled thresholds when normalized to the raw thresholds. The gray bars, which represent the dichoptic conditions, are largest for the rotation-along-blue axis transformation. Thus although in the dichoptic condition there is variation in the size (and direction) of the difference between raw and phase-scrambled thresholds, with some possibly significant individual points, on average there is no significant difference. On the other hand the black bars, which show the monocular differences in threshold, are consistently above zero, in many cases large and, as we have shown, significant. The results in the monocular condition confirm our previous findings using binocularly presented images (Yoonessi & Kingdom, 2008).

Finally, if one pools data for both raw and phase-scrambled thresholds, the mean monocular threshold,  $1.70 E - 2$ , was significantly lower than the mean dichoptic threshold,  $3.65 E - 2$  ( $t = 5.38$ ,  $df = 114$ ,  $p < 0.001$ ).

## Discussion

In line with our previous study (Yoonessi & Kingdom, 2008), sensitivity to uniform color and/or luminance changes in images of natural scenes under normal viewing

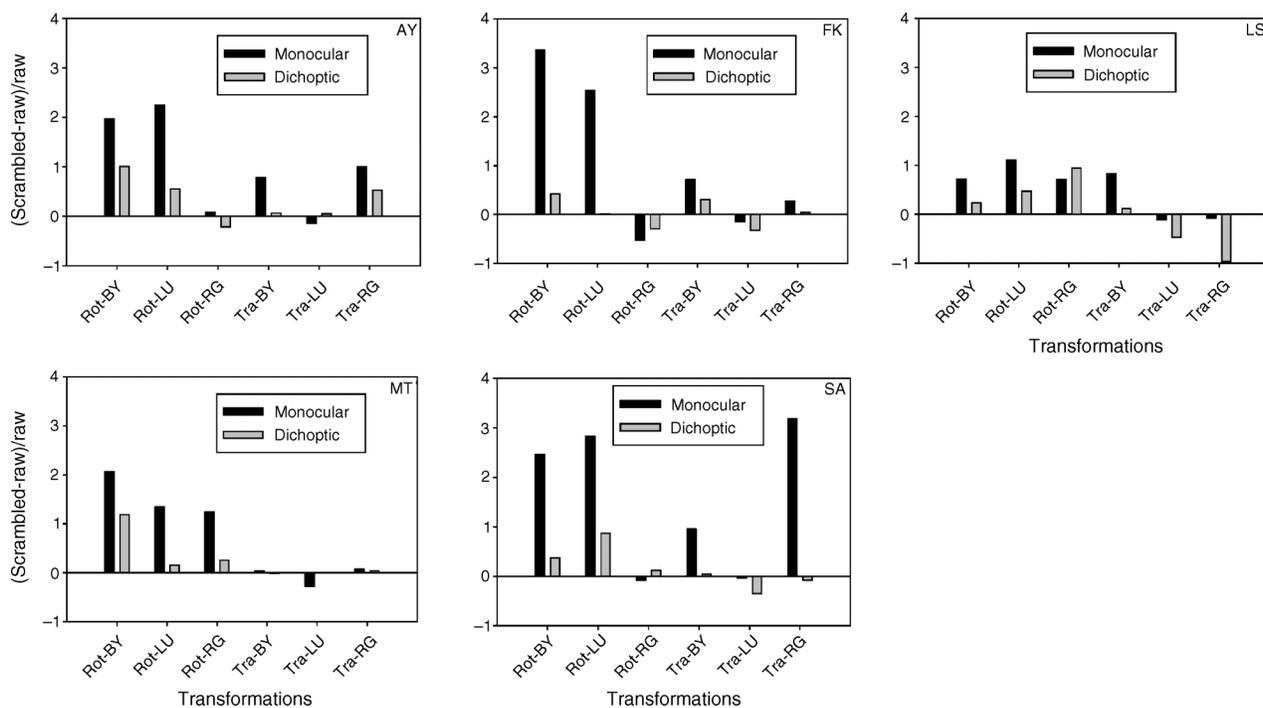


Figure 5. Each bar represents the difference in threshold Euclidean distance between the phase-scrambled and raw scene conditions, normalized to the raw scene thresholds. The first three letters of the labels on the abscissa indicate the type of transformation (Rot = Rotation, Tra = Translation), and the last two letters the axis (BY = Blue–Yellow, RG = Red–Green, Lu = Luminance).

conditions was higher for raw compared to phase-scrambled images. The principle aim of this study was to determine whether or not this result held under dichoptic viewing conditions, in which the color transformations were applied in opposite directions to a dichoptically fused image pair. It did not. This suggests that the mechanisms responsible for the higher sensitivity to raw compared to phase-scrambled scenes under normal viewing conditions operate only at or after the point of binocular combination.

In the monocular condition, subjects viewed the two stereo-halves of each image side by side. Could it be possible that they identified the manipulated image pair as the pair that looked ‘less normal’, and that this produced the relatively low thresholds? In the manipulated pair, both images were transformed in opposite directions. The difference *between* the two images of the pair would therefore be twice the difference between that of either image and the untransformed version. Thus on these grounds alone, it would be easier for the subject to compare the two images with each other than compare either of them with the untransformed, ‘normal’ images. When added to the fact that a mental representation of what is normal would be widely tuned for both average luminance and average chromaticity (as these properties vary considerably in natural scenes due to varying lighting conditions) it is extremely unlikely that the lower thresholds in the monocular condition were a result of the adoption of any sort of ‘compare-to-normal’ strategy.

What type of mechanism mediates DDTs? [Figure 6](#) shows a possible scheme for what might be happening in the dichoptic condition. The scheme illustrates both differencing and averaging the monocular signals emanating from both the dichoptically different and dichoptically identical image pairs. The idea of a binocular-differencing channel is an old one (Cohn & Lasley, 1976; DeSilva & Bartley, 1930) but not a generally accepted one. In [Figure 6](#) the ‘signal’ from each monocular image is a Gaussian distribution representing hypothetical responses ( $y$ -axis) of a set of neurons tuned to different colors ( $x$ -axis) in response to a small region of the image. The negative and positive symbols in the middle of the figure represent, respectively, the binocular-differencing and binocular-averaging channels. At the bottom of the figure are the *differences* in responses between the dichoptically different and dichoptically identical image pairs calculated for each channel (remember the task for the subjects was to decide which of the stimuli was dichoptically different). As can be seen the binocular-differencing channel provides a larger differential response to the two image pairs than does the binocular-averaging channel. Although this model is too simple to make any quantitative predictions, in particular because it does not include nonlinearities such as response compression and half-wave rectification, it is consistent with the idea that a putative binocular-differencing mechanism mediates DDTs. A stronger argument however is that it is hard to imagine how, or

why, a binocular-averaging channel would signal the ‘lustre’ that subjects consistently say mediate their judgments of dichoptic difference.

In the dichoptic condition, in addition to finding near-equal sensitivities to the raw and phase-scrambled scenes, thresholds were overall higher than in the monocular condition by a factor of about 2.2. Malkoc and Kingdom (2004) found the ratio of dichoptic to monocular thresholds to be about 3.7 in their experiments using patches differing in chromaticity.

Why are the dichoptic thresholds higher than their monocular counterparts? Suppose that dichoptic differences are signaled via the putative binocular-differencing channel in [Figure 6](#). One can reasonably assume that it will be active only if there are signals in both monocular inputs (otherwise it would be activated when one eye was closed), perhaps via some form of AND-gating. Therefore in the monocular condition this channel will be largely inactive. However when activated, for example by the dichoptically different stimuli, there might be an inhibitory input from the binocular-averaging channel; indeed the inhibition might be mutual between the two channels. The inhibition from the binocular-averaging channel would to some degree suppress the signal from the dichoptically different signal, resulting in higher DDTs in the dichoptic compared to monocular conditions. This is admittedly speculative, and only additional experiments will determine if this explanation is correct. It will also be interesting to see if current models of binocular summation (e.g., Baker, Meese, & Georgeson, 2007; Meese, Georgeson, & Baker, 2006) are able to account for the measured difference between the dichoptic and monocular conditions in the present as well as previous studies (Malkoc & Kingdom, 2004, in preparation).

What is it about the post-binocular stages of visual processing that produces lower thresholds for uniform color transformations applied to raw compared to phase-scrambled scenes? If the difference between the transformed and untransformed images were reflected in the average absolute difference in the outputs of an array of linear, or quasi-linear filters such as cortical simple cells, we would not expect a difference in thresholds between the raw and phase-scrambled scenes (Yoonessi & Kingdom, 2008). So linear cortical filtering is not the reason. What types of cortical nonlinearity might account for the difference? In the past decade, numerous studies have demonstrated that the responses of some cortical neurons to stimuli placed within their classical receptive field can be modulated by stimuli placed outside the receptive field, or within what is sometimes termed the ‘extra-classical’ receptive field, or ERF (Angelucci & Bullier, 2003; Blakemore & Tobin, 1972; DeAngelis, Freeman, & Ohzawa, 1994; Maffei & Fiorentini, 1976; Nelson & Frost, 1985; Rao & Ballard, 1999; Sengpiel, Sen, & Blakemore, 1997; Webb et al., 2002; Zetsche & Röhrebein, 2001). The ERF is nonlinear, in that it can only modulate the response of the neuron if the classical

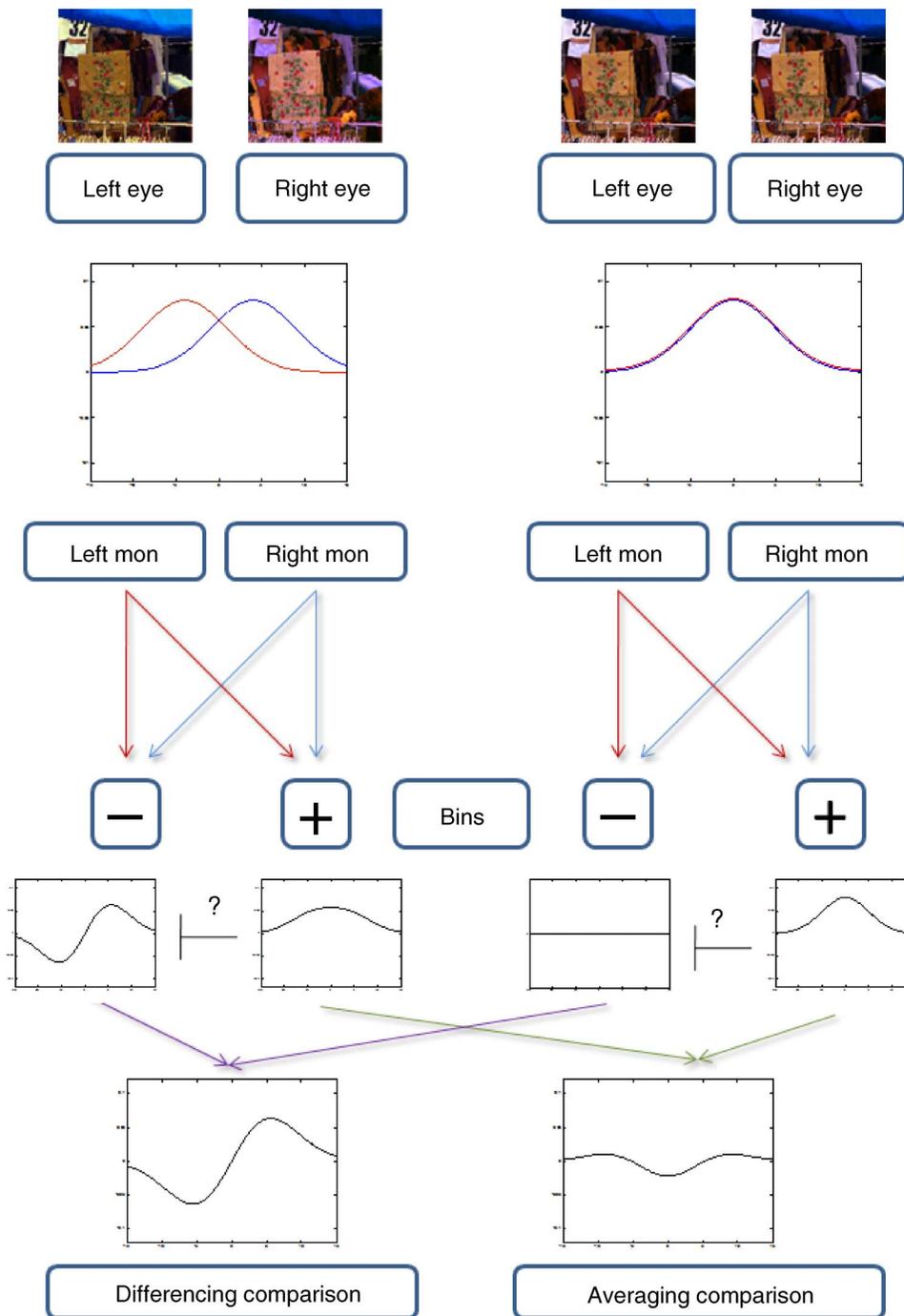


Figure 6. A possible scheme for how monocular signals might be combined binocularly to detect dichoptic differences. *Top left:* dichoptically different image pair transformed in equal and opposite directions by rotation along the red–green axis. *Top right:* dichoptically identical image pair. The Gaussian curves under each image pair show hypothetical outputs of a set of neurons tuned to different colors in response to a small region of the image. Both image pairs are detected by both a binocular-differencing (–) and binocular-averaging (+) channels. The differential activation to the two image pairs is shown for each channel at the bottom of the figure. Both channels could in principle detect the difference between the two image pairs, but the binocular-differencing channel gives the bigger differential response. The binocular-differencing channel is shown to be inhibited from the binocular-averaging channel, though the inhibition might be mutual. Mon = monocular; Bins = binocular channels.

receptive field is already activated. Could ERF neurons be responsible for the difference between the raw and phase-scrambled scenes? Grigorescu, Petkov, and Westenberg (2003) have simulated the responses of ERF neurons to images of natural scenes and revealed how they pick up isolated contours as well as orientational discontinuities in textures, but not the elements in uniformly textured regions. This finding is complemented by psychophysical data from Kingdom and Prins (2005) who showed that contour-shape-sensitive mechanisms are relatively unresponsive to contours that are flanked by parallel contours. Therefore we would expect ERF neurons to be less responsive to the discontinuities in phase-scrambled compared to raw natural scenes, because phase scrambling has the effect of spreading out energy from edges into the space between them. Therefore on the assumption that ERF neurons are at least in part responsible for detecting color changes, they are a plausible candidate for the sensitivity differences we found between raw and phase-scrambled scenes in our plain-view monocular condition. Although the chromatic properties of ERF neurons have not to our knowledge been studied, it is noteworthy that there is recent evidence for binocular color-preference cells in the macaque whose color preferences are well-matched in the two eyes (Peirce, Solomon, Forte, & Lennie, 2008). If such cells were subject to ERF, they would be strong candidates for detecting the uniform color transformations studied here.

Of course nonlinearities also exist in pre-binocular neurons. These include retinal cone adaptation (Enroth-Cugell & Shapley, 1973; Shapley & Enroth-Cugell, 1984; Smirnakis, Berry, Warland, Bialek, & Meister, 1997) and half-wave rectification in the lateral geniculate nucleus (Dan, Atick, & Reid, 1996; Duong & Freeman, 2008). One reason why these nonlinearities might not result in differential sensitivity to raw versus phase-scrambled scenes is that they are by-and-large ‘point-wise’, i.e., highly localized (Cleland & Freeman, 1988; MacLeod, Williams, & Makous, 1992; Rushton, 1965; Shapley & Enroth-Cugell, 1984; Williams & MacLeod, 1979). In other words the impact of these nonlinearities on sensitivity would be the same on average irrespective of whether the energy in the image was spread out more or less evenly or concentrated into edges.

## Conclusion

The influence of higher order statistics on the detectability of uniform color changes applied to natural scenes occurs after information from two eyes is combined. It is suggested that the reason why higher order statistics might have relatively little impact prior to the point of binocular combination is that the nonlinearities are largely point-wise. On the other hand, the influence of higher order

statistics beyond the point of binocular combination might be mediated by neurons whose responses to edges are inhibited by neighboring textures.

## Acknowledgments

We thank Dr. Adriana Olmos for calibrating the camera and providing the images of natural scenes.

This study was supported by a Canadian Institute of Health Research Grant No. 11554 to FK.

Commercial relationships: none.

Corresponding author: Ali Yoonessi.

Email: ali.yoonessi@mcgill.ca.

Address: Royal Victoria Hospital, Room H4.14, 687 Pine Avenue West Montreal, Quebec H3A 1A1, Canada.

## References

- Angelucci, A., & Bullier, J. (2003). Reaching beyond the classical receptive field of V1 neurons: Horizontal or feedback axons? *The Journal of Physiology*, *97*, 141–154. [PubMed]
- Baker, D. H., Meese, T. S., & Georgeson, M. A. (2007). Binocular interaction: contrast matching and contrast discrimination are predicted by the same model. *Spatial Vision*, *20*, 397–413. [PubMed]
- Blake, R. (2001). A primer on binocular rivalry, including current controversies. *Brain and Mind*, *2*, 5–38.
- Blakemore, C., & Tobin, E. A. (1972). Lateral inhibition between orientation detectors in the cat’s visual cortex. *Experimental Brain Research*, *15*, 439–440. [PubMed]
- Boynton, R. M. (1973). Implications of the minimally distinct border. *Journal of the Optical Society of America*, *63*, 1037–1043. [PubMed]
- Brainard, D. H., Rutherford, M. D., & Kraft, J. M. (1997). Color constancy compared: Experiments with real images and color monitors. *Investigative Ophthalmology & Visual Science*, *38*, 2206.
- Brown, S. P., & Masland, R. H. (2001). Spatial scale and cellular substrate of contrast adaptation by retinal ganglion cells. *Nature Neuroscience*, *4*, 44–51. [PubMed]
- Cleland, B. G., & Freeman, A. W. (1988). Visual adaptation is highly localized in the cat’s retina. *The Journal of Physiology*, *404*, 591–611. [PubMed] [Article]
- Cohn, T. E., & Lasley, D. J. (1976). Binocular vision: Two possible central interactions between signals from two eyes. *Science*, *192*, 561–563. [PubMed]

- Dan, Y., Atick, J. J., & Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *Journal of Neuroscience*, *16*, 3351–3362. [PubMed] [Article]
- DeAngelis, G. C., Freeman, R. D., & Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology*, *71*, 347–374. [PubMed]
- DeSilva, H. R., & Bartley, S. H. (1930). Summation and subtraction of brightness in binocular perception. *British Journal of Psychology*, *20*, 242–252.
- Duong, T., & Freeman, R. D. (2008). Contrast sensitivity is enhanced by expansive nonlinear processing in the lateral geniculate nucleus. *Journal of Neurophysiology*, *99*, 367–372. [PubMed] [Article]
- Enroth-Cugell, C., & Shapley, R. M. (1973). Adaptation and dynamics of cat retinal ganglion cells. *The Journal of Physiology*, *233*, 271–309. [PubMed] [Article]
- Fine, I., & MacLeod, D. I. A. (2001). Visual segmentation based on the luminance and chromaticity statistics of natural scenes [Abstract]. *Journal of Vision*, *1*(3):63, 63a, <http://journalofvision.org/1/3/63/>, doi:10.1167/1.3.63.
- Grigorescu, C., Petkov, N., & Westenberg, M. A. (2003). Contour detection based on nonclassical receptive field inhibition. *IEEE Transactions on Image Processing*, *12*, 729–739. [PubMed]
- Horn, R. A., & Johnson, C. R. (1985). *Matrix analysis*. Cambridge, UK: Cambridge University Press.
- Johnson, A. P., & Baker, C. L., Jr. (2004). First- and second-order information in natural images: A filter-based approach to image statistics. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, *21*, 913–925. [PubMed]
- Kingdom, F. A. A., & Prins, N. (2005). Different mechanisms encode the shapes of contours and contour-textures [Abstract]. *Journal of Vision*, *5*(8):463, 463a, <http://journalofvision.org/5/8/463/>, doi:10.1167/5.8.463.
- Kingdom, F. A., Field, D. J., & Olmos, A. (2007). Does spatial invariance result from insensitivity to change? *Journal of Vision*, *7*(14):11, 1–13, <http://journalofvision.org/7/14/11/>, doi:10.1167/7.14.11. [PubMed] [Article]
- Ledda, P., Santos, L. P., & Chalmers, A. (2004). A local model of eye adaptation for high dynamic range images. *Proceedings of the 3rd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa* (pp. 151–160). New York: ACM Press.
- MacLeod, D. I., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America*, *69*, 1183–1186. [PubMed]
- MacLeod, D. I., Williams, D. R., & Makous, W. (1992). A visual nonlinearity fed by single cones. *Vision Research*, *32*, 347–363. [PubMed]
- Maffei, L., & Fiorentini, A. (1976). The unresponsive regions of visual cortical receptive fields. *Vision Research*, *16*, 1131–1139. [PubMed]
- Malkoc, G., & Kingdom, F. A. A. (2004). Dichoptic difference thresholds for the properties of chromatic stimuli [Abstract]. *Journal of Vision*, *4*(11):64, 64a, <http://journalofvision.org/4/11/64/>, doi:10.1167/4.11.64.
- Meese, T. S., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision*, *6*(11):7, 1224–1243, <http://journalofvision.org/6/11/7/>, doi:10.1167/6.11.7. [PubMed] [Article]
- Nelson, J. I., & Frost, B. J. (1985). Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Experimental Brain Research*, *61*, 54–61. [PubMed]
- Olmos, A., & Kingdom, F. A. (2004a). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, *33*, 1463–1473. [PubMed]
- Olmos, A., & Kingdom, F. A. A. (2004b). *McGill Calibrated Colour Image Database*. From <http://tabby.vision.mcgill.ca/>.
- Párraga, C. A., Troscianko, T., & Tolhurst, D. J. (2002). Brief communication spatiochromatic properties of natural images and human vision. *Current Biology*, *12*, 483–487.
- Peirce, J. W., Solomon, S. G., Forte, J. D., & Lennie, P. (2008). Cortical representation of color is binocular. *Journal of Vision*, *8*(3):6, 1–10, <http://journalofvision.org/8/3/6/>, doi:10.1167/8.3.6. [PubMed] [Article]
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87. [PubMed]
- Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters*, *73*, 814–817. [PubMed]
- Ruderman, D. L., Cronin, T. W., & Chiao, C. C. (1998). Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America A*, *15*, 2036–2045.
- Rushton, W. A. H. (1965). The Ferrier lecture, 1962: Visual adaptation. *Proceedings of the Royal Society of London B: Biological Sciences*, *162*, 20–46.
- Sengpiel, F., Sen, A., & Blakemore, C. (1997). Characteristics of surround inhibition in cat area 17. *Exper-*

- imental Brain Research*, 116, 216–228. [PubMed] [Article]
- Shapley, R., & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3, 263–346.
- Shapley, R., & Hawken, M. (2002). Neural mechanisms for color perception in the primary visual cortex. *Current Opinion in Neurobiology*, 12, 426–432. [PubMed]
- Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W., & Meister, M. (1997). Adaptation of retinal processing to image contrast and spatial scale. *Nature*, 386, 69–73. [PubMed]
- Smith, V. C., & Pokorny, J. (1975). Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Research*, 15, 161–171. [PubMed]
- Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, 38, 310–324. [PubMed]
- Webb, B. S., Tinsley, C. J., Barraclough, N. E., Easton, A., Parker, A., & Derrington, A. M. (2002). Feedback from V1 and inhibition from beyond the classical receptive field modulates the responses of neurons in the primate lateral geniculate nucleus. *Visual Neuroscience*, 19, 583–592. [PubMed]
- Webster, M. A., & Mollon, J. D. (1997). Adaptation and the color statistics of natural images. *Vision Research*, 37, 3283–3298. [PubMed]
- Weisstein, E. W. (2004). “Affine Transformation.” From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/AffineTransformation.html>.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63, 1293–1313. [PubMed]
- Williams, D. R., & MacLeod, D. I. (1979). Interchangeable backgrounds for cone afterimages. *Vision Research*, 19, 867–877. [PubMed]
- Yoonessi, A., & Kingdom, F. A. (2008). Comparison of sensitivity to color changes in natural and phase-scrambled scenes. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 25, 676–684. [PubMed]
- Yoonessi, A., & Kingdom, F. A. A. (2007). Faithful representation of colours on a CRT monitor. *Color Research and Application*, 32, 388.
- Zetzsche, C., & Röhrbein, F. (2001). Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network*, 12, 331–350. [PubMed]