Detection, Color Matching, and Discrimination: An Exploration of the Nature and Number of Chromatic Mechanisms

by Timothy G. Shepard

B.A. in Psychology, University of Massachusetts-Boston
M.A. in Psychology, Northeastern University

A dissertation submitted to

The Faculty of
the College of Science of
Northeastern University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

August 10, 2016

Dissertation directed by

Rhea T. Eskew, Jr.
Professor of Psychology
Abstract of Dissertation

Narrowly-tuned, selective noise masking of chromatic detection has been taken as evidence for the existence of “higher-order” color mechanisms. Experiment 1 replicates earlier observations of selective masking of tests in the (L, M) plane of cone space when the noise is placed near the corners of the detection contour. We used unipolar Gaussian blob tests, with three different noise color directions, and show that there are substantial asymmetries in the detection contours, asymmetries that would have been missed with bipolar tests such as Gabor patches. A new chromatic detection model, which is based upon probability summation of linear cone combinations, incorporates a linear contrast energy vs. noise power relationship that predicts how the sensitivity of these mechanisms changes with noise contrast and chromaticity is presented. With only six unipolar color mechanisms – the same number as the cardinal model – the new model accounts for the threshold contours across the different noise conditions, including the asymmetries and the selective effects of the noises. Our model is different from the cardinal model, in at least one important way. Four of the six detection mechanisms here have opposed L and M cone inputs and that is the feature that allows it to predict selective masking across noise conditions without a large number of mechanisms.

In Experiment 2, results of asymmetric color matching of the threshold-level tests from Experiment 1 were used to test the detection model. The assumption is that the mechanisms are univariant labelled lines (Rushton, 1972; Watson & Robson, 1981), implying here that the colors of physically-different stimuli that are detected by a single mechanism should all be the same, and suggesting that the colors of two stimuli detected
by different mechanisms should differ. Physically different stimuli that lie along one mechanism threshold line should therefore produce metameric color matches – post-receptoral metamers – while tests that lie on two different mechanism lines should be matched with different colors. The results show that color matches fall into six clusters in CIE (u’, v’) space (across all the noise conditions), and these clusters correspond closely to the six mechanisms in the model. Most importantly, where the detection model determines that a given test angle is detected by different mechanisms in different noise conditions (due to differential masking), the hue of that test angle changes in a consistent way. These color matches essentially apply a color label to each of the mechanism threshold lines, confirm the six-mechanism model, and quantify the hue signaled by each mechanism.

In a discrimination experiment (Experiment 3), on a given trial, two threshold-level stimuli (one "standard" and one “test”) were presented in random order, and observers were asked to select the standard. Across trials, many different standards and tests in the (L,M) plane of cone contrast space were used, with and without chromatic masking noise (Shepard et. al., 2016). Qualitatively, (i) if both stimuli are detected by the same chromatic mechanism, performance should be at chance level (50%); (ii) if the stimuli are detected by two different chromatic mechanisms, discrimination should be as good as detection (ca. 82%); (iii) if one or both stimuli are detected by multiple mechanisms, performance should be intermediate. Quantitatively, a Bayesian Color Classifier (Eskew et al., 2001) was used to combine the outputs of the six mechanisms in the detection model from Shepard et. al., (2016); the classifier’s predictions are computed
without any free parameters. In general, the classifier’s predictions for each observer and each condition are excellent, confirming the detection model.

In conclusion, the three experiments presented in this dissertation utilized unipolar blobs along with chromatic detection, color matching, and discrimination to search for higher order mechanisms and test a variety of models. In Experiment 1, a new detection model is presented that can produce selective masking with only six linear mechanisms—adding mechanisms does not significantly improve the fit. In Experiment 2, we use a color-matching paradigm to apply a color label to each of the mechanism threshold lines in the model fit, confirm the six-mechanism model, and quantify the hue signaled by each mechanism. In Experiment 3, discrimination was analyzed by a Bayesian classifier model to test the six-mechanism detection model. We show that the classifier’s predictions for each observer in every condition are excellent, confirming the detection model.
# Table of Contents

Abstract of Dissertation .......................................................... ii

Chapter 1: Introduction .......................................................... 1

Chapter 2: General Methods ..................................................... 10

Chapter 3: A model of selective masking in chromatic detection .... 18

Chapter 4: Labelling the lines: A test of the six-mechanism model ... 71

Chapter 5: Discrimination-A Bayesian Classifier Model ............ 106

Chapter 6: General Discussion ............................................... 145

References ............................................................................ 154

General Appendices .............................................................. 155
Chapter 1 - Introduction

Color vision is vital for many of the tasks that we carry out in our daily lives: identifying objects (e.g. an orange or a grapefruit?), detecting harmful foods (e.g. mushrooms, berries, and moldy fruits), responding to aesthetics and art, perceiving emotional expression (e.g. emoticons), and diagnosing health problems (e.g. jaundice, dermal rashes, dehydration). From the early stages of processing to the much later cognitive stages, it is important that we understand how the color visual system operates.

In 1860, Hermann von Helmholtz reformulated Thomas Young’s theory (Young, 1802) on color vision and published the influential Young-Helmholtz trichromatic theory (Helmholtz, 1925), which proposed that the human retina contains three types of photoreceptors that are necessary for normal color vision. Over the last 150 years, research in color vision has progressed immensely. Initially, much of color vision research focused on the first stage of processing which further unveiled the actual functions of three receptors, each containing different photopsins that respond differently to specific wavelengths of light—(L)ong, (M)edium, and (S)hort wavelengths (Wald et al., 1955; Svaetichin, 1956; for review see Gegenfurtner, 2003). It also became apparent that cones absorb light at all visible lengths and are not chromatically selective to one specific “color”. Furthermore, it was widely accepted that there is not a direct pathway from the cone receptors to the visual processing centers of the brain.

Over the last three decades, researchers have paid less attention to the first stage of color vision processing, instead focusing on the interaction/comparison of cone signals—the ‘mechanisms’ required for the detection and discrimination of color. Here we define a color mechanism as a fixed (relative) combination of cone signals that is
correlated with the observer’s behavior in psychophysical experiments (Eskew, 2008). Mechanisms are univariant, which implies that when two stimuli are detected by one and only one mechanism, there is some relative intensity at which the two stimuli cannot be discriminated from one another. They are also “labelled lines” which implies that when two stimuli are detected by different mechanisms, they must be discriminable from one another at all relative intensities. These two properties of a mechanism will be used throughout this dissertation.

Post-receptiorally, cones pass on information to bipolar cells and then relay information to the retinal ganglion cells (Fig. 1a). According to a standard model, three channels then transmit information from the ganglion cells to the visual processing center of the brain (LGN to V1-V8). In the L-M color opponent channel, signals from both the L and M cones are subtracted from each other (reddish-greenish channel). In the S-(M+L) color opponent channel, the sum of both the L and M cones is subtracted from the short wavelength cones (yellowish-bluish channel). In the L+M channel, L and M cone signals are summed to produce an achromatic or luminance channel (Boynton, 1979) (Figure 1b). These ‘cardinal directions’ were first suggested by LeGrand (1949, 1968), after a reanalysis of MacAdam’s chromatic discrimination data (1942), (i.e. MacAdam’s ellipses) showed that discrimination thresholds appear to be mediated by two variables (an L-M and an S-cone pathway).
This so-called “cardinal axes” model assumes the presence of a red/green detection mechanism along with an opponent yellow/blue mechanism, and a non-opponent luminance mechanism (Krauskopf et al., 1982). An elaborated version of the cardinal axes model, with the original bipolar color channels split into unipolar pairs (Eskew, 2009; Sankeralli & Mullen, 1997, 2001) is shown in Figure 2. This model contains three symmetric pairs of rectified mechanisms—a different mechanism may be responsible for the detection of reddish and another for greenish stimuli (blue and a yellow, increment and a decrement mechanism).
Higher-order Mechanisms

However, Krauskopf (1999) and Hansen & Gegenfurtner (2013) showed that the simple cardinal axis-like model is incorrect. Krauskopf and colleagues (1986) argued that there were instead many “higher order” color mechanisms, initially conceived of as recombinations of the cardinal mechanisms, tuned to a large variety of hues.

Krauskopf (1982) utilized a habituation experiment where observers first habituated to a light modulated along one particular direction of color space. The stimuli were large chromatic fields; detection thresholds were measured before and after exposure to a modulation of chromaticity along the cardinal axes and intermediate directions. They found that thresholds for detecting reddish and greenish changes from achromatic white were higher after viewing a reddish/greenish field but not after viewing a field in the yellowish/bluish direction—and thresholds for detecting blue and yellowish...
changes from white were higher after viewing a bluish/yellowish field but not after viewing a field in the reddish/greenish direction. Habituation along one of three directions, which they called “cardinal”, thresholds were elevated along the same “cardinal” direction but not along the other directions in color space. However, Krauskopf et al. (1986) later reanalyzed their data from the habituation experiment and reported that, in the equiluminant plane, thresholds were elevated most for tests that had the same chromaticity as the habituation stimuli, consistent with the presence of detection mechanisms that are tuned to specific hues in that plane. Krauskopf et al. referred to these mechanisms as ‘higher-order’ color mechanisms because they could be thought of as the result of a recombination of the signals generated by the classical mechanisms; the recombination(s) of signals from the cardinal axis mechanisms are narrowly tuned to specific colors (e.g. orange, pink, purple). These so-called ‘higher order’ mechanisms are in addition to the six aforementioned cardinal directions and thought to be processed beyond the LGN neurons at a cortical level (De Valois, 2000; Gegenfurtner, Kiper, & Levitt, 1997; Kiper, Fenstemaker, & Gegenfurtner, 1997).

Many experimental tasks have provided evidence for non-classical/non-cardinal mechanisms, tuned to specific colors including: detection (D’Zmura, Lennie, & Krauskopf, 1987; Gegenfurtner & Kiper, 1992; D’Zmura & Knoblauch, 1998) color appearance (Webster, 1994; Webster, 1991; Mizokami, 2004; Krauskopf, 1986), visual search (D’Zmura, 1991; Nagy, 2004; Monnier, 2001), tilt after-effect (Flanagan, 1990; Clifford, 2003), classification images (Hansen, 2005; Bouet, 2004), first-order discrimination (Zaidi, 1993; Li, 1997; Krauskopf, 1999; Krauskopf, 1992; Hansen, 2006), second-order texture discrimination (Goda, 2001; Li, 1997), spatial alignment
(McKeefry, 2004; McGraw, 2004) detection of Glass patterns (Wilson, 2005; Cardinal, 2003), and motion coherence (Krauskopf, 1996). An even more recent study suggests a large number of mechanisms (4,096 mechanisms) are necessary to account for color discrimination of textured patterns with noise masking (Hansen & Gegenfurtner, 2006).

In the present study, psychophysical procedures, specifically masking of chromatic detection, were used to investigate the number of color mechanisms required for chromatic detection and discrimination. The logic of masking and chromatic detection is straightforward. Observers’ chromatic detection thresholds are measured for multiple color tests. Then, masking noise is added at a color angle and detection thresholds for the same tests are re-measured. Figure 3 shows one example where baseline detection thresholds were measured (white open symbols), and then reddish-greenish masking noise was added at the color direction of 135°-315° and detection thresholds were measured again (black filled symbols). When noise was added, thresholds along the flanking red and green mechanisms shift outwards (i.e. threshold elevation), with no change in slope of the flanks.

Figure 3: Detection thresholds (open symbols) were measured in the LM plane. Detection thresholds with masking noise at the color direction of 135°/315° (filled symbols). Figure from Eskew (2009).
There was little effect on the detection thresholds at the color angles of 45° and 225° (labeled on figure 3), which suggests the presence of, at least, one pair of mechanisms other than the red-green pair. Most likely, desensitization occurred in the red and green mechanisms, which exposed, in this case, a pair of luminance mechanisms. These increment and decrement luminance mechanisms sum the L and M cone signals, so their detection contour has a negative slope (i.e., if \( L+M=1 \) at threshold, then \( M=1-L \)).

Even though others have found evidence for higher order mechanisms, these results are consistent with a cardinal axis-like model, with a limited number of mechanisms. For example, Hansen & Gegenfurtner, 2006; Gegenfurtner et al., 1992; and Gegenfurtner et al., 2010 also used a noise-masking paradigm with multiple noise masks in the corner of the detection contour (e.g. 38° and 47°). What they found was that placing noise in the corners of the detection contour led to selective masking at specific angles and little masking effects at adjacent angles, which suggests the presence of multiple mechanisms tuned to those specific directions. The difference in results between this work and past research, where no “higher order” mechanisms were found, is most likely due to the implementation of noise masking, at multiple angles, near the corners of the detection contour. One reason for this discrepancy may have to do with our selection of tests and noise angles—this will be addressed in Chapter 3.

Acknowledging Hansen and Gegenfurtner’s (2013) criticism about the selection of our tests and noise directions, we measured detection contours in the L,M plane with and without masking noise; when noise was actually placed near the corners of the contour where the underlying mechanisms have equal sensitivity. Shepard and Eskew (2013) replicated the main effect of Hansen and Gegenfurtner (2013), with a Gabor (i.e.,
bipolar) stimulus, finding that the effect of the noise was selective: the maximum threshold elevation was at or near the noise direction. We demonstrated that this result could be accounted for with a model that contained a very limited number of symmetrically paired mechanisms. Notably, both studies used bipolar test stimuli that contain equal and opposite chromatic contrasts. The symmetry of such stimuli enforces symmetry on the threshold contour and the model fit that can obscure the presence of asymmetric mechanisms—ones that do not form pairs with equal and opposite cone weights. As discussed by (Eskew, 2009), unipolar mechanisms like this can act as if there were more mechanisms than actually exist, since different mechanisms can detect the two poles of bipolar stimuli independently.

Three experiments presented here utilized unipolar blobs along with chromatic detection, color matching, and discrimination to search for higher order mechanisms and test a variety of models. With only six unipolar color mechanisms—the same number as the cardinal model—a new model is presented that accounts for the threshold contours across the different noise conditions, including the asymmetries and the selective effects of the noises. This model is different from the cardinal model in at least one important way—four of the six detection mechanisms here have opposed L and M cone inputs and that is the feature that allows it to predict selective masking across noise conditions without a large number of mechanisms. Most importantly, it will be shown that adding mechanisms (i.e. higher order model) does not significantly improve the fits.

Experiment 1 (Chapter 3) aimed to determine the number of mechanisms required to account for the pattern of threshold elevations produced by chromatic noise masking in the L,M plane of cone space. The purposes of Experiment 2 (Chapter 4) were to
determine the hue signaled by each mechanism and to test the detection model in Experiment 1. In Experiment 3 (Chapter 5), the same experimental strategy that has been used in other contexts (Eskew et al., 2001) was employed, combining the detection data with discrimination data to further test the model. Taken together, this dissertation provides considerable evidence that only a limited number of mechanisms are required for chromatic detection and discrimination and that each mechanism signals a mechanism-specific hue.
Chapter 2 - General Methods

Observers

Four well-practiced observers (TGS, CLM, SAF, and NO) participated in the experiments. All had normal scores on the Farnsworth-Munsell 100 hue test (Farnsworth, 1943) and the Ishihara Plates. Northeastern University’s Institutional Review Board approved the research protocol; the procedures comply with the Declaration of Helsinki.

Apparatus

Stimuli were created on a Power Macintosh computer and displayed on a SONY GDM-F520 CRT monitor running at 75hz, using an ATI Radeon 7500 video board with a driver verified to support the 10-bit digital-to-analog converters (DACs). Spectroradiometric calibration was performed across the spectrum using a Photo Research PR-650 spectroradiometer, and the monitors were linearized with the Gamma correction lookup tables (further discussion of the gamma correction method can be found in General Appendix A). Observers were corrected to normal visual acuity using a trial lens that was placed in front of their dominant eye; the other eye was patched. Head position was stabilized with a chin and forehead rest. All experiments were conducted in a dark room.

In Experiment 2 (Chapter 3), a 15-inch MacBook Pro using an Intel HD 4000 graphics card was placed off to the side; a piece of black cardboard was used to mask the screen, showing only the stimulus of interest.
Test and Noise Stimuli

The test stimuli were circular Gaussian blobs ($\sigma=1^\circ$) presented against a grey background field with a rapid-start ‘sawtooth’ profile of 333ms total duration (Figure 4). The sawtooth profile was chosen to maximize the likelihood of separating ON- and OFF-responses. Fixation was guided by four black diagonal lines, pointing at the center of the screen, ending 1.5° from the center, which were present throughout the experiment.

The masking noise consisted of horizontal lines that were superimposed on the test (Figure 5). The lines randomly and independently changed from one chromaticity to a symmetrically opposite chromaticity (on the opposite side of the white point), so that the mean chromaticity was unchanged. Each line switched chromaticity with probability 1/2 at 18.75 Hz. The power spectrum of the noise is plotted in Figure 2 of Wang et al. (2014); it is dominated by low spatial and temporal frequencies. The noise contrast was always 90% of the maximum available at the noise direction; in cone contrast units, the noise cone contrast vector length $|\mathbf{n}|$ (see Appendix A in Shepard et al., 2016) was 0.498, 0.414, and 0.267 for the 42°/222°, 48°/228°, and 64°/244° noises, respectively.

All of the stimuli were “half-toned”; two pixel horizontal lines of noise alternated with two pixel lines of test (see Giulianini & Eskew, 1998). The high spatial frequency components in the test that were created by this half-toning procedure were not generally visible, except occasionally for tests with color directions near the ends of the contour in the no-noise condition. Note that the test was always half-toned, whether the noise was present or not, and thus the test profile was the same in all conditions. The reported contrasts in this dissertation have been halved (from nominal, peak value) to facilitate comparison with studies not using half-toned stimuli.
Detection thresholds were measured at many different chromatic angles in the (L,M) plane (3-5 runs of 100 trials at each angle).

Figure 4: (above) Stimuli are presented with a rapid-start profile ‘sawtooth’ of 333ms total duration (below) the test stimuli are circular Gaussian blobs (σ=1°)

Figure 5: Chromatic noise. The lines randomly and independently change from one chromaticity to a symmetrically opposite chromaticity (on the opposite side of the white point), so that the mean chromaticity is unchanged. Each line switches chromaticity with probability 1/2 at 18.75 Hz.

Color representation

Stimuli are represented in two color spaces in this paper. Because its units are non-arbitrary, cone contrast space is the primary representation used, and when angles and distances are given they refer to cone contrast space. In this space, the axes represent...
the modulation of the cones, relative to the steady excitation produced by the mean adapting field in meaningful units. The contrast of the stimulus, $|t|$, is defined as the Euclidean distance from the origin of the $(ΔL/L, ΔM/M)$ point representing the peak of the stimulus: $|t| = \sqrt{\left(\frac{ΔL}{L}\right)^2 + \left(\frac{ΔM}{M}\right)^2}$

The second color representation used here is a variant of a cone-excitation space, here called MacLeod-Boynton-Derrington-Krauskopf-Lennie (MBDKL) space. In the (L,M) plane of this space, the horizontal axis represents the difference of the L and M cone excitations, and the vertical axis their sum; L cone excitation increases to the right and upward. This space represents changes in the cone quantal catch from the mean adaptive field, in arbitrary units; here we normalize the distance along both the horizontal and vertical axes by threshold, using the averages of the two L-M (135°/315°) and two L+M (45°/225°) no-noise thresholds for a given observer.

Detection contours in the $(ΔL/L, ΔM/M)$ plane are typically quasi-elliptical, with their long axis along the 45° and 225° direction. As pointed out by Hansen and Gegenfurtner (2013), relative to cone contrast space, the MBDKL representation expands the chromaticities near the ends of this detection contour and can make it easier to see potential selective masking effects in this region. For this reason, we sometimes show our data in both color spaces (Chapter 3 and 4), as did Hansen and Gegenfurtner (2013).

**Procedure**

**Measuring Detection Thresholds**

A 2AFC adaptive staircase procedure was used to measure detection thresholds. Observers adapted to the grey background field for 90 secs before each run of 100 trials.
In the noise conditions, the observer adapted to the grey background plus the modulating noise. In each run, a single test color direction (and noise color direction in the masking conditions) was used. Test angle runs were intermixed and most observers measured the thresholds for ~12 different test angles, per session.

Each trial consisted of two 333 ms intervals signaled by tones and separated by 200 ms and the observer was asked to determine which interval the test stimulus appeared in. Observers made their responses using a keypad that was positioned off to the right side of the observer. Each trial began when the observer pressed the “Enter” key and the responses were made by pressing the “1” key for interval one and the “2” key for interval two. Observers were given auditory feedback (i.e. a higher tone for correct response and a lower tone for an incorrect response) after each detection trial in all of the experiments.

The Wetherill method was used to select the test contrast on each trial (Figure 5). The x-axis denotes the trial number in the block and the y-axis is the stimulus intensity. The black squares indicate a correct response on a given trial, at a specified contrast, and the red squares represent an incorrect response. If the subject picks the correct interval for three trials in a row, the contrast is lowered by 0.1 log units. If the observer selects the wrong interval the contrast of the test stimulus is raised by 0.1 log units. This method, which is commonly known as the three-down, one-up staircase method, will approach the point at which the observer is 79% correct (i.e. this joint event is the cube root of 0.50 which is equal to 0.794). The underlying assumptions are that there is an absolute threshold for detecting the stimulus and that the observer will always respond incorrectly when the stimulus is below this absolute threshold and correctly when the contrast is
above this threshold. These assumptions are incorrect, but this method is appropriate for selecting the test contrast in the current experiment.

Weibull functions were fit to the accumulated frequency-of-seeing data for each run using a maximum likelihood method to estimate two parameters of the psychometric function—a threshold estimate, corresponding to a detection rate of 82%, and an estimate of the psychometric slope. After fitting the Weibull functions, thresholds from multiple runs were averaged (3-5 runs at each color angle); standard errors were calculated using between-run (mostly between-session) variances. Additional runs were added in cases where the coefficient of variation was unusually high.

Figure 5: Wetherill method. The x-axis denotes the trial number in the block and the y-axis is the stimulus intensity. The black squares indicate a correct response on a given trial, at a specified contrast, and the red squares represent an incorrect response. In a 2AFC task, if the subject picks the correct interval for three trials in a row, the contrast is lowered. If the observer selects the wrong interval the contrast of the test stimulus is raised.

**Color Matching**

In Experiment 2 (Chapter 4), the fixed contrast, threshold-level tests were used as stimuli. These were presented in the same noise conditions used in the detection
experiment: one no-noise, and up to three noise conditions (i.e. 42°/222°, 48°/228°, and 64°/244° in cone contrast space).

For color matching, test angles were blocked by noise condition, and run in random order within that block; the observer did not know which test was being presented. For each test angle, the first step was to confirm that the test contrast was correctly set at threshold. The observer was shown the test, fixed at its measured threshold from the detection experiment (Shepard et al., 2016), in one of two temporal intervals, just as in the threshold measurement procedure. If the observer correctly selected the test interval, then the color match was performed. If the observer did not correctly select the test interval, the color match was not performed on that trial and the next trial in that block (i.e. at the same test angle) was presented. In each block, the observers selected the correct trial ~80% of the time, confirming that thresholds did not change between the collection of detection thresholds and the color matching experiment. After looking at the test and making a correct response, the observer turned to the laptop positioned off to the side and selected a matching chromaticity on a H.S.V. color wheel Figure 4, Chapter 4. When satisfied, the observer confirmed their match by viewing a disc on the same peripheral display that was roughly the same size as the threshold level test. This disc was presented on a grey background with the same chromaticity coordinates as the main experimental monitor. The H.S.V. values of the confirmed match were then recorded for the test and a new test angle was selected. Five matches were made at each test angle, chosen in random order, in each noise condition; matches in a given noise condition were collected in at least two sessions on different days. Although
observers were free to vary the brightness (i.e. value) of the matching stimulus, none of them did, so the analysis here is based only upon hue and saturation.

After the experiment was completed, the (x, y, L) chromaticity coordinates of each of the five matching discs were measured on the laptop with an X-Rite i1Display Pro. These five chromaticities were averaged; the average of the five matches was then transformed to CIE (u’,v’) space. The averaged match coordinates of each test were used as the basic data in all further analysis of the color matches.

**Discrimination**

In the discrimination experiment (Chapter 5), two stimuli were fixed at their detection thresholds from Experiment 1 (Chapter 3), and presented in two intervals in random order. One color direction was selected as the target (the ‘standard’), and the observer learned, from the feedback tones, which was the standard stimulus in the first 10 practice trials of the 50 trial block. The observer was asked to select the interval that contained the ‘standard’. The practice trials were discarded and performance (i.e. percent correct) was averaged across multiple runs for the same test, in the same noise condition, with the same standard. The discrimination task was also performed in the same noise conditions from the detection experiment: one no-noise, and up to three noise conditions (i.e. 42°/222°, 48°/228°, and 64°/244° in cone contrast space).
Chapter 3: A model of selective masking in chromatic detection
Abstract

Narrowly-tuned, selective noise masking of chromatic detection has been taken as evidence for the existence of a large number of color mechanisms (i.e., higher order color mechanisms). Here we replicate earlier observations of selective masking of tests in the (L, M) plane of cone space when the noise is placed near the corners of the detection contour. We used unipolar Gaussian blob tests, with three different noise color directions, and show that there are substantial asymmetries in the detection contours, asymmetries that would have been missed with bipolar tests such as Gabor patches. We develop a new chromatic detection model, which is based upon probability summation of linear cone combinations, and incorporates a linear contrast energy vs. noise power relationship that predicts how the sensitivity of these mechanisms changes with noise contrast and chromaticity. With only six unipolar color mechanisms – the same number as the cardinal model – the new model accounts for the threshold contours across the different noise conditions, including the asymmetries and the selective effects of the noises. The key for producing selective noise masking in the (L,M) plane is having more than two mechanisms with opposed L and M cone inputs, in which case selective masking can be produced without large numbers of color mechanisms.
Introduction

By the mid-1980s, a consensus had developed on the outlines of a model of color detection and discrimination, one that seemed to explain both psychophysics and LGN physiology (Boynton, 1979; Lennie & D'Zmura, 1988). This model was based primarily on data collected by Krauskopf, Williams, and Heeley (1982), who used habituation to find the “cardinal axes” (mechanism isolating directions) of color space. An elaborated version of a model based upon this work, with the original bipolar color channels split into unipolar pairs (Eskew, 2009; Sankeralli & Mullen, 1997, 2001) is shown in Figure 1. However, the cardinal consensus was shattered by evidence, summarized in Eskew (2009), Krauskopf (1999) and Hansen and Gegenfurtner (2013), showing that this simple model is incorrect. Instead, Krauskopf and colleagues (1986) argued that there were instead many “higher order” color mechanisms, initially conceived of as recombinations of the cardinal mechanisms, tuned to a large variety of hues.

Some of the best evidence against the cardinal axis model consists of selective masking or habituation, in which the threshold-elevating effect is tuned to very specific color directions, rather than being broadly-tuned as the cardinal axis model would predict. For example, Hansen and Gegenfurtner (2013) put masking noise near the corner of the detection contour in the (L,M) plane – the plane of cone space in which the S cones are unmodulated – and found highly selective masking. Noises near the corners produced maximum masking at those specific angles, and much less even at very close angles. Their interpretation was that there are a great many higher-order mechanisms.
Following Hansen and Gegenfurtner (2013), we measured detection contours in the (L,M) plane in the present study, with and without masking noise; the noise was placed near the corners of the contour (where the underlying mechanisms have equal sensitivity). We replicate the main effect of Hansen and Gegenfurtner (2013) by finding that the effect of the noise was in some cases selective: the maximum threshold elevation was at or near the noise direction. However, we show that a model can produce selective masking effects with only six mechanisms, the same number as the cardinal mechanisms model. The key feature of the model that allows for selective masking in the (L,M) plane is the presence of more than two mechanisms with opposed L and M cone inputs.

Methods

Observers

Four well-practiced observers (TGS, CLM, SAF, and NO) participated in the experiment. TGS and CLM participated in all four noise conditions. Observer SAF participated in 3 conditions and observer NO in two. All had normal scores on the Farnsworth-Munsell 100 hue test (Farnsworth, 1943) and the Ishihara Plates. Northeastern University’s Institutional Review Board approved the research protocol; the procedures comply with the Declaration of Helsinki.

Apparatus

Stimuli were created on a Power Macintosh computer and displayed on a SONY GDM-F520 CRT monitor running at 75hz, using an ATI Radeon 7500 video board with a
driver verified to support the 10-bit digital-to-analog converters (DACs).

Spectroradiometric calibration was performed at 4 nm intervals across the spectrum using a Photo Research PR-650 spectroradiometer, and the monitors were linearized with the Gamma correction lookup tables. Observers were corrected to normal visual acuity using a trial lens that was placed in front of their dominant eye; the other eye was patched. Head position was stabilized with a chin and forehead rest. All experiments were conducted in a dark room.

Test and Noise Stimuli

The test stimuli were circular Gaussian blobs ($\sigma=1^\circ$) presented against a grey background field with a rapid-start ‘sawtooth’ profile of 333ms total duration (Fig. 2). The sawtooth profile was chosen to maximize the likelihood of separating ON- and OFF-responses. Fixation was guided by four black diagonal lines, pointing at the center of the screen, ending 1.5° from the center, which were present throughout the experiment.

The masking noise consisted of horizontal lines that were superimposed on the test (Fig. 3). The lines randomly and independently changed from one chromaticity to a symmetrically opposite chromaticity (on the opposite side of the white point), so that the mean chromaticity was unchanged. Each line switched chromaticity with probability 1/2 at 18.75 Hz. The power spectrum of the noise is plotted in Figure 2 of Wang et al. (2014); it is dominated by low spatial and temporal frequencies. The noise contrast was always 90% of the maximum available at the noise direction; in cone contrast units, the noise cone contrast vector length $|\mathbf{n}|$ (Appendix A) was 0.498, 0.414, and 0.267 for the $42^\circ/222^\circ$, $48^\circ/228^\circ$, and $64^\circ/244^\circ$ noises, respectively.
All of the stimuli were “half-toned”; two pixel horizontal lines of noise alternated with two pixel lines of test (see Giulianini & Eskew, 1998). The high spatial frequency components in the test that were created by this half-toning procedure were not generally visible, except occasionally for tests with color directions near the ends of the contour in the no-noise condition. Note that the test was always half-toned, whether the noise was present or not, and thus the test profile was the same in all conditions. Detection thresholds were measured at many different chromatic angles in the (L,M) plane (3-5 runs of 100 trials at each angle).

Procedure

Detection

A 2AFC adaptive staircase procedure was used to measure detection thresholds. Observers adapted to the grey background field for 90 secs before each run of 100 trials. In the noise conditions, the observer adapted to the grey background plus the modulating noise (when present). In each run, a single test color direction (and noise color direction in the masking conditions) was used.

Each trial consisted of two 333 ms intervals signaled by tones and separated by 400 ms. The observer was asked to determine which interval the test stimulus appeared in and received feedback after each response. The stimulus contrast was decreased by 0.1 log units after three consecutive correct responses and increased by the same amount after one incorrect response. Weibull functions were fit to the accumulated frequency-of-seeing data for each run using a maximum likelihood method to estimate two parameters of the psychometric function —a threshold estimate, corresponding to a detection rate of
82%, and an estimate of the psychometric slope. After fitting the Weibull functions, thresholds from multiple runs were averaged (3-5 runs at each color angle); standard errors were calculated using between-run (mostly between-session) variances. Additional runs were added in cases where the coefficient of variation was unusually high. The data reported here represent results from more than 110,000 forced-choice trials.

Color representation

Stimuli are represented in two color spaces in this paper. Because its units are non-arbitrary, cone contrast space is the primary representation used, and when angles and distances are given they refer to cone contrast space. In this space, the axes represent the modulation of the cones, relative to the steady excitation produced by the mean adapting field in meaningful units. The contrast of the stimulus, |t|, is defined as the Euclidean distance from the origin of the (ΔL/L, ΔM/M) point representing the peak of the stimulus: |t| = \sqrt{\left(\frac{ΔL}{L}\right)^2 + \left(\frac{ΔM}{M}\right)^2}

Reported contrasts have been halved (from the nominal, peak value) to facilitate comparison with studies not using half-toned stimuli.

The second color representation used here is a variant of a cone-excitation space, here called MacLeod-Boynton-Derrington-Krauskopf-Lennie (MBDKL) space. In the (L,M) plane of this space, the horizontal axis represents the difference of the L and M cone excitations, and the vertical axis their sum; L cone excitation increases to the right and upward. This space represents changes in the cone quantal catch from the mean adaptive field, in arbitrary units; here we normalize the distance along both the horizontal and vertical axes by threshold, using the averages of the two L-M (135°/315°) and two L+M (45°/225°) no-noise thresholds for a given observer.
Detection contours in the ($\Delta L/L$, $\Delta M/M$) plane are typically quasi-elliptical, with their long axis along the $45^\circ$ and $225^\circ$ direction. As pointed out by Hansen and Gegenfurtner (2013), relative to cone contrast space, the MBDKL representation expands the chromaticities near the ends of this detection contour and can make it easier to see potential selective masking effects in this region. For this reason, we show our data in both color spaces, as did Hansen and Gegenfurtner (2013).

**Model**

The models used in this paper have, as elements, rectified linear chromatic mechanisms: weighted sums, with weights $W_L$, $W_M$ for the L and M cones respectively, with that sum defined to be 1.0 at threshold. The mechanisms are half-wave rectified, so that each mechanism only responds in one half of the chromatic space (present Fig. 1; Eskew, 2009). The mechanisms are combined by an approximation to probability summation, with a Minkowski exponent of 4.0. These features of the model are identical or nearly identical to those used previously by several authors (Cole, Hine, & McIlhagga, 1993; Eskew et al., 2001; Eskew, Jr., McLellan, & Giulianini, 1999; Giulianini & Eskew, 1998; Newton & Eskew, 2003; Sankeralli & Mullen, 1996)

The novel aspect of the present model is how the mechanisms vary across noise conditions. We assume here that thresholds mediated by a single mechanism follow an Energy vs. Noise (EvN) relationship based upon the theory of noise masking (Giulianini & Eskew, 2007; Wang et al., 2014), in which the contrast energy of the test is linearly related to the noise contrast power. The threshold contrast—proportional to the square root of the test contrast energy—for a test at angle $\tau$, in the presence of the noise of contrast power $|n|^2$ at angle $\nu$, is given by
\[ |\mathbf{t}| = \sqrt{\frac{1}{Q_t |\mathbf{f}|^2 \left[ \cos(\alpha - \tau) \right]^2} + b \frac{Q_n |\mathbf{n}|^2}{Q_t} \left[ \cos(\alpha - \nu) \right]^2} \]  

Eq. (1)

(derived in Appendix A). The half-brackets \((\text{−},)\) represent half-wave rectification (i.e., values less than zero are set to zero). The vector of cone contrast weights, the “mechanism vector” (Eskew, Jr. et al., 1999), is \(\mathbf{f}\), which takes polar angle \(\alpha\) in the \((L,M)\) plane \((\alpha=0\) is the \(L\) cone increment direction). The subscripted \(Qs\) are calculated constants that represent the energy and power in (t)est and (n)oise, respectively, and the fitted parameter \(b\) represents the sensitivity of the mechanism to the spatio-temporal characteristics of the noise. The first term in the radical represents the baseline, no-noise added condition, and the second term raises the threshold due to the noise.

The thresholds for all the tests for a given mechanism and noise condition are then combined by probability summation to account for the set of thresholds in that condition. The three free parameters for each mechanism are the two cone contrast weight components of \(\mathbf{f}\) (or, equivalently, the magnitude \(|\mathbf{f}|\) and its angle \(\alpha\)), and the noise sensitivity parameter \(b\). Below, we report the polar angle of the mechanism vector and vector length in the \((\Delta L/L, \Delta M/M)\) plane (see Table 1), with asymptotic standard error estimates from the fits (and after applying the appropriate propagation of error formulae in converting from the cone weights to angles and vector lengths). The degrees of freedom for fitting the model is the number of thresholds across all noise conditions (e.g., 92 for TGS) minus 3 times the number of mechanisms (e.g., 74 df for TGS’s six-mechanism model).

Eq. (1) tightly constrains the model. Each mechanism must have the same polar angle in the no-noise and all the noise conditions, and thus its threshold loci are lines of
the same slope in every condition; those lines are orthogonal to the mechanism vector \( \mathbf{f} \).

Even more importantly, for every test angle, the masking effect of the noise on the mechanism must be proportional to \( \cos^2(\alpha - \nu) \), with the constant of proportionality \( bQ_n/Q_t \) the same for all test and noise angles detected by that mechanism. These model features are derived from theory and are consistent with prior results in both luminance and chromatic detection (Gegenfurtner & Kiper, 1992; Giulianini & Eskew, 1998; Legge, Kersten, & Burgess, 1987; Pelli, 1981; Wang et al., 2014) and collectively they make model comparisons much more powerful: for example, comparing models with different numbers of these mechanisms is not based upon curve-fitting, but rather is a comparison of theoretical accounts of the data.

**Results and Discussion**

**No noise condition**

Figure 4 shows the thresholds for Gaussian blobs without noise for three observers (TGS CLM, and SAF). Subject NO, who only participated in two conditions, is not shown; her limited results are consistent with the other observers. The points denote measured thresholds and the small black lines through those points indicate plus and minus one standard error (based upon between-session variability only); in many cases the error bars are smaller than the symbols. Colored lines on the plots represent mechanism thresholds (discussed below) and the smooth closed contour is the probability sum of those mechanisms. The line color is a rough indication of the hue of stimuli that lie on that line, according to informal observations of the observers. For example, stimuli
on or near the orange line appear ‘orangey-red,’ and stimuli on or near the blue line appear ‘bluish’.

The long flanks that comprise the majority of the measured thresholds lie near two parallel lines with the approximate slope of +1 for all four observers. These data are qualitatively similar to many previous studies (Cole et al., 1993; Cole, Stromeyer, & Kronauer, 1990) and (Giulianini & Eskew, 1998) showing approximate symmetry about the main diagonal (45°/225°), and very high sensitivity of the thresholds along those long flanks (Chaparro, Stromeyer, Huang, Kronauer, & Eskew, 1993). However, there are individual differences between observers. CLM’s contour is narrower than TGS’s and SAF’s. The aspect ratio of CLM contour (the average of thresholds at 45° and 225° divided by the thresholds at 135°/315°) is 9.32, whereas TGS’s only 3.02, SAF is 3.88, and NO is 3.93; compared to some previous studies (Cole et al., 1993; Sankeralli & Mullen, 1996), TGS, SAF, and NO are relatively insensitive along the flanks (see Eskew, Jr. et al., 1999 for review).

In addition, the sets of threshold along the flanking regions converge in the first quadrant (QI), especially for CLM – an effect never observed previously – resulting in a trapezoidal detection contour when plotted in MBDKL.

There are differences between the increment and decrement thresholds at the ends of the contour. For all four observers, increment sensitivity is higher than decrement sensitivity (red points)—the decrement thresholds at 225° are between 1.25-1.9 fold higher than the increment thresholds at 45°. A similar asymmetry has been observed previously (e.g., Giulianini & Eskew, 1998), Fig 3b).
Selective Masking in the presence of chromatic noise

Thresholds in the presence of 42°/222° noise are shown in Figure 5, for 48°/228° in Figure 6, and for 64°/244° noise in Figure 7 (note change of scale compared to Fig. 4; thresholds are anywhere from 5-20 times higher when these noises are added). The detection contours are highly elongated in quadrants I and III (QI and QIII), as observed previously with 45° noise (Hansen & Gegenfurtner, 2013).

More importantly, as shown by Hansen and Gegenfurtner (2013), the masking effect of the noise is selective. Figure 8 shows threshold elevations (relative to the no noise fitted contour), as a function of angular deviation between the test and noise angles. For the 42°/222° and 48°/228° noise conditions, threshold elevations are highest when the test angles are located very near the direction of the noise, and fall steeply for tests that are only a few degrees away.

Figure 7 shows the 64°/244° noise condition for TGS, CLM, and SAF. The contours substantially broader than those in Figures 5 and 6, suggesting a transition, as the noise vector rotates toward QII/QIV, towards the broad, nonselective noise effects observed in many studies previously with noises in QII/QIV (Eskew et al., 2001; Giulianini & Eskew, 1998). Figure 8 (red symbols) shows the less-selective effect of this noise: there is relatively more elevation at test angles that are further removed from the noise.

In all of the noise conditions, there are also asymmetries between increment and decrement thresholds. These asymmetries are of the opposite direction to those found in the no noise condition (cf. Vingrys & Mahon, 1998): thresholds near the upper end of the noise contour in QI (increments) are 1.2-1.5 larger than the QIII thresholds (decrements)
in all of the noise conditions. Figure 8 shows the asymmetry in terms of elevation: the peak on the left is higher than the peak on the right. This result – more masking of increments than decrements – is in the same direction as the S cone masking studied by Wang et al. (2014). The asymmetries may be partially due to the sawtooth temporal profile of the tests, which may help separate responses from ON and OFF pathways (Wang et al., 2014).

Six-mechanism model

Only four linear mechanisms suffice to provide an excellent fit to the thresholds in all noise conditions when each observer and noise condition is considered separately (fits not shown), consistent with previous findings (Eskew et al., 2001; Giulianini & Eskew, 1998). In particular, in the no-noise condition for TGS and CLM, the set of mechanisms (R, G, Y, and P) are almost identical to mechanisms found previously (e.g. Giulianini & Eskew, 1998) – SAF has a small intrusion of two additional mechanisms even in the no-noise case. However, across the set of four noise conditions, a total of six mechanisms are required, for all observers, as discussed next.

Our model combines the outputs of linear chromatic mechanisms by probability summation, to fit detection thresholds for multiple noise conditions. The model is fit to all of the noise conditions simultaneously, with each mechanism following the Test Energy vs. Noise Power relationship (Eq. A.1; Appendix A). As will be shown, a model with six mechanisms is able to account for data across multiple noise conditions, and produce selective masking where it exists in the data (and less-selective masking where that is what the data show, in the 64°/244° condition). Table 1 shows the fitted
parameters, with single letter names chosen as mnemonics based upon the approximate hue of the tests at threshold. The table also provides asymptotic standard errors for the parameters. The statistical uncertainty implied by these standard errors may be visualized for one observer (TGS) in Figure S1: the general result is that the data tightly constrain the angle and sensitivity of mechanisms responsible for detection along the flanks of the contours, but the higher thresholds at the ends of the contour make for greater uncertainty in the other mechanism angles.

Generally, the model fits the data extremely well across all of the noise conditions, accounting for a very large portion of the variance ($R^2 \geq 0.98$), but there are a few areas along some of the contours where the fit is poor (see Figs. 5-7). For observer TGS, the model slightly overestimates the thresholds along the flanks and slightly underestimates thresholds in the corners of the $48^\circ$/228° condition. For observer CLM, the model overestimates thresholds in QI in the $64^\circ$/244° condition. There is also a slight over-estimation along the greenish flank (QII) in the $42^\circ$/222° noise condition. These small discrepancies result from the constraint that the slopes of each mechanism line must the same in across all the conditions, for each observer, since the model was fit to all of the noise conditions simultaneously (Eq. A.1; Appendix A).

In some conditions for some observers, a given mechanism does not contribute to any of the thresholds. An example is shown by the blue line in Figure 4 for observers TGS and CLM. This mechanism threshold lies well outside the data, especially for TGS; its position is determined, not by the no-noise thresholds shown in Figure 4, but by the thresholds in the three other noise conditions, as it obeys the relationship given in Eq. 1: compare the slope of the blue line in Figure 4 with the blue lines in Figures 5, 6, and 7.
This example illustrates the fact that the model fit is applied to all of the noise conditions simultaneously, and how the model constrains the fit in any one noise condition.

For observers TGS and SAF, the mechanism vectors of R and G, and of O and B, are separated by approximately 180° and are approximately equal in magnitude (i.e. they have approximate odd-symmetry) – and can be thought of as ‘quasi-paired’. CLM is the exception. For all observers, mechanisms Y and P are not symmetric in angle or sensitivity. However, the uncertainty on the weights of the Y and P mechanisms is large, since (with the partial exception of the 64°/244° noise condition) only a small segment of the mechanism threshold line is revealed in the data (Figure S1). Importantly, fitting the model with pairs of mechanisms being required to be exactly symmetric in angle and sensitivity produced significantly worse fits (discussed below).

Table 2 gives the peaks of the detection contours in all the noise conditions. These are not necessarily separated by 180°, because the probability summation contour of asymmetric mechanisms is not symmetric. The peaks do not exactly match the noise direction but there is clearly a selective effect.

Figure 9 summarizes the main features of the model fit with nested plots that contains all of the probability summation contours for TGS, CLM, and SAF plotted on the same scale. The colored points on the contours represent the noise directions and the black dotted line indicates the main diagonal, 45°/225° axis for reference. The change in the shape of the contours demonstrates that there is less selectivity as the noise angle moves away from the end of the contour (i.e., the contour is broader), a result consistent with previous studies (see Giulianiini & Eskew, 1998); compare with Figure 8.
Alternative models

As noted, at least two of the six mechanisms (Y & P) are clearly asymmetrical (i.e., unpaired), with asymmetrical cone weights and different sensitivities. The other four mechanisms (R, G, B, and O) are quasi-paired— the estimated weights are not exactly equal and opposite, but the stimuli they detect appear to be approximately symmetrical along the flanks. Perhaps these quasi-paired mechanisms are actually two parallel pairs. Therefore, we re-fit the model, trying various symmetry constraints on the mechanisms. We constrained R and G, and B and O, to have cone weights that were equal in magnitude and opposite in sign (i.e. we made them two opponent pairs). All three model fits had $R^2 > 0.97$. However, the model fits with these symmetry and sensitivity constraints were significantly worse than the original six-mechanism model where both the mechanism’s sensitivity and cone weights were free to vary, even after taking into account the reduced number of free parameters produced by the symmetry constraint. These conclusions hold whether or not the noise sensitivity parameter $b$ was constrained to be the same for the two members of a pair. Details of these analyses are given in the Supplemental material. Allowing slight asymmetries in the model is necessary to satisfactorily fit the data across multiple noise conditions.

A major aim of this study was to determine whether a “higher order” model, one with many mechanisms, was required to produce selective masking and account for the data along the detection contours. For this reason, we tested several variations on models with eight and sixteen linear mechanisms. These did not provide a significantly better fit; for one model (the base model of Hansen & Gegenfurtner (2013) with 16 symmetric mechanisms), the fit was significantly worse than our six-mechanism model. There is no
Thus, for this extensive set of data, the six-mechanism model could not be significantly improved by adding additional mechanisms. Other possible models might include mechanism nonlinearities, which are particularly plausible given the high noise contrast power used in the present study. One type of nonlinearity we tried was raising the cosine terms in Eq. 1 to an exponent before being squared in the power calculation (e.g., $\cos^{2\gamma}(\alpha - \upsilon)$, with $\gamma$ a free parameter), for all or some of the mechanisms. This type of nonlinearity, which has been used in several previous studies (D’Zmura & Knoblauch, 1998; Goda & Fujii, 2001; Hansen & Gegenfurtner, 2006; McKeefry, McGraw, Vakrou, & Whitaker, 2004), forces a symmetric tuning curve for the masking effect of the noise, and it failed here.

Another important and plausible nonlinear model involves adaptive changes in cone weights, resulting from the high contrast of the noises (Atick, Li, & Redlich, 1993; Zaidi & Shapiro, 1993). There is no question that some sort of adaptive model could fit our data. We have fit four linear mechanisms to each noise condition considered independently, and the fit is outstanding in all 13 cases (including NO whose data are not shown). Thus we could easily fit the entire set of data simply by slightly altering the cone contrast weights of the R and G mechanisms in each noise condition to align the long flanks of the detection contour approximately with the noise vector, and only two of the other mechanisms would be required. However, without some theoretical constraint on the adjustment of the weights across noise conditions, this four-mechanism adaptive model would be nearly impossible to disprove (Eskew, 2009).
In addition, informal observations by our observers suggest that the hues of the thresholds fall into six, not four, categories. Therefore, as attractive an idea as the adaptive mechanism model is, it is unlikely that such a model could account for the subjective experience resulting from these mechanisms (i.e., the hues produced by the mechanisms). A study of the color appearance of tests detected by these mechanisms is currently in progress.

**General Discussion and Conclusions**

The asymmetries in thresholds we observed in this study, taken together with related asymmetries observed in other studies (e.g., Vingrys & Mahon, 1998; Wang et al., 2014) (Krauskopf et al., 1982; Krauskopf & Zaidi, 1986) imply that experiments using bipolar stimuli, such as Gabor patches or gratings, are likely to miss important features of the data, since detection contours measured with such stimuli are required to be symmetric. It seems likely that bipolar stimuli will generate detection contours that are the inner envelope produced by the set of mechanisms (i.e. the most sensitive mechanisms will dominate), and perhaps therefore miss theoretically important features of the data.

We replicated the main result of Hansen and Gegenfurtner (2013): the effect of the noise could be selective. The maximum threshold elevation was at or near the noise direction in the 42°/222° and 48°/228° conditions (Fig. 8). Masking at 64°/244° was less selective, suggesting a transition to nonselective masking as the noise is moved away from the ends of the detection contour; this is consistent with previous results showing nonselective masking when noise was placed in QII/QIV of this plane in color space (i.e., noises at 90°/270° (M cone noise), 120°/300°, 135°/315°, and 0°/180° (L cone noise) (R.)
Eskew et al., 2001; Giulianini & Eskew, 1998), although those noises were of lower power than those used in the present study and in Hansen & Gegenfurtner (2013).

Hansen and Gegenfurtner (2013) noted this difference between their study and our earlier experiments (Giulianini & Eskew, 1998): their noise had a peak cone contrast vector length of 0.4, whereas Giulianini and Eskew used peaks averaging about 0.04 in most conditions. In the present study, our peak cone contrast vector lengths (after accounting for half-toning) are about 0.50, 0.41, and 0.27 (for the 42°/222°, 48°/228°, and 64°/244° noise conditions, respectively). However, the noise used by Hansen and Gegenfurtner (2013) was drawn from a uniform probability distribution, whereas our noise (like that of (Giulianini & Eskew, 1998) was binary, which produces greater power (Gegenfurtner & Kiper, 1992): for the same peak values, the variance (and hence the contrast power) of a uniform distribution is 3-fold lower than that of the Bernoulli (binary) distribution with the same maximum values, because many of the samples from the uniform distribution lie close to zero. Thus the Hansen and Gegenfurtner (2013) peak noise contrast of 0.4 has power that is equivalent to a binary noise with peak contrast of $0.23 \ (0.4 / \sqrt{3})$, comparable to our weakest noise. Therefore the less-selective effect of the 64°/244° noise in our study, which our model predicts, is not merely because the noise does not have enough power.

We show that a model can produce this selective masking with only six linear mechanisms, the same number as the cardinal mechanism model. Adding more than six mechanisms never significantly improved the fit. The main point of this study is that selective masking is not evidence for large numbers of mechanisms, which differs from the conclusion of Hansen and Gegenfurtner (2013). In fact, our model predicts selective
masking for noises across a range of angles near 45°/225° in cone contrast space (calculations not shown), and the much wider corresponding range of angles in threshold-scaled MBDKL space. The model also predicts much less selectivity for noises away from the ends of the detection contour (e.g., Fig. 7), with the same six mechanisms. This conclusion does not depend upon the use of the unipolar test stimulus. We have shown the same selective masking with Gabor patch tests, and a six-mechanism (necessarily symmetric) model could account for those selective masking results as well (J. R. T. Eskew & Shepard, 2013; Shepard, Swanson, & Eskew, 2013).

The model of Table 1 is depicted in Figure 10. S cone input is shown in two of the mechanisms, to be as similar as possible to the cardinal model of Figure 1. The assignment of this S cone input to these particular mechanisms is speculative here, since we did not modulate S cones in the present experiment. However, it is important to keep in mind that any linear postreceptoral mechanism(s) receiving S cone input will be active in the (L,M) plane (the sole exception being a mechanism that gets only S cone input, for which there is no evidence), and since some of the mechanisms of Figure 10 must get S cone input we have depicted it in the figure for completeness. Although the cardinal model (Fig. 1) has the S cone input opposed by a sum of L and M, substantial prior evidence indicates that S cone detection mechanisms have long-wavelength inputs of opposite sign, as shown here (De Valois & De Valois, 1993; McLellan & Eskew, 2000; Wang et al., 2014; Wisowaty, 1983).

Although our model has six mechanisms like the elaborated cardinal model of Figure 1, our model differs in one important respect: four of our six detection mechanisms have opposed L and M cone inputs, rather than only two. This is the
essential feature of our model – not its asymmetry (see Supplemental Material in Chapter 3) – that allows it to account for selective masking when noise is placed near the ends of the detection contour in the (L,M) plane; noises that are nearly parallel to the long flanking detection contours can cause different mechanisms to become most sensitive and thus determine threshold, tilting the overall detection contour. Further study of the properties of these mechanisms is ongoing.

It is difficult to relate any psychophysical threshold model to the activity of visual cortical neurons, in part because behavioral thresholds are likely dominated by the most sensitive subset of cells. Nonetheless, there are several possible points of comparison between our model and cortical neurons, at least those with foveal or near-foveal receptive fields. First, the mechanisms in our model are, at least to a good approximation, linear combinations of cone signals. Some cortical neurons respond to linear cone combinations, especially in V1, but others are nonlinear and thus more narrowly or more broadly tuned (De Valois, Cottaris, Elfar, Mahon, & Wilson, 2000; Horwitz & Hass, 2012; Lennie, Krauskopf, & Sclar, 1990); it is not generally clear how sensitive the linear cells compared to the nonlinear ones. Second, there are only six mechanisms in our model, which might suggest that the especially sensitive cells would fall into six clusters in terms of their cone weights. Many physiological studies report cells with a large variety of chromatic tunings (De Valois et al., 2000; Horwitz & Hass, 2012; Komatsu, 1998), but again it is not clear that these are highly-sensitive, nor even that they all actually have to do with color vision (e.g., Horwitz & Hass, 2012); the fact that they respond to chromatic stimuli might only be a result of irrelevant variation in cone connectivity (Conway, 2009). Third, based upon our model, most of the sensitive cells
should have opposed L and M cone inputs; L/M opponency is common in many cortical
cells but it is not clear that it predominates (e.g., Horwitz & Hass, 2012). Fourth – and
most optimistically – the cells might have cone contrast weights similar to those in Table
1. Of course, it is important to keep in mind that, even among cortical cells that actually
serve color vision, there are likely to be neurons that do not satisfy the definition of
psychophysical color mechanisms: univariant labelled lines with fixed relative spectral
tuning (e.g. tuning may change with contrast; Horwitz & Hass, 2012; Namima, Yasuda,
Banno, Okazawa, & Komatsu, 2014; Solomon & Lennie, 2005; see Eskew, 2009 for
discussion).
Appendix A

The threshold contrast energy of an achromatic (Burgess, Wagner, Jennings, & Barlow, 1981; Pelli, 1981) or chromatic (Gegenfurtner & Kiper, 1992; Giulianini & Eskew, 1998) test stimulus may be described by a linear function of the contrast power of the noise. We refer to this relationship as an empirical energy versus noise (EvN) function, expressed as:

\[ E = N_0 + N \]  \hspace{1cm} (A.1)

where \( E \) represents the test contrast energy (proportional to squared contrast), \( N \) the applied noise contrast power (also proportional to squared contrast), and \( N_0 \) is a constant representing the level of intrinsic noise in the detection mechanism. For a rectified but otherwise linear chromatic detection mechanism, Eq. (A.1) may be written as

\[ Q_n \left[ f \cdot t \right]^2 = N_0 + bQ_n \left[ f \cdot n \right]^2 \]  \hspace{1cm} (A.2)

(compare Eq. 9 of Wang et al., 2014). In Eq. (A.2), \( f \) is the vector of cone contrast weights of the mechanism, \((W_L, W_M)\). The vector \( t \) represents the cone contrasts \((\Delta L/L, \Delta M/M)\) of the test, and \( n \) is the corresponding vector representing the noise. The half-brackets \( \left[ \right] \) represent half-wave rectification (i.e., values less than zero are set to zero).

\( Q_n \) is the constant of proportionality between the noise contrast squared and the noise power. Its value was taken to be the unit contrast noise power at DC, which is 1.07 x 10^3 deg sec, after accounting for half-toning. This value was calculated in Wang et al (2014) for radially-symmetric noise; the value here is the same, considering the vertical
dimension of the pattern. The analogous constant for the contrast energy of the blob test, again considering only the vertical dimension, is

\[ Q_t = \int_0^\infty \int_0^\infty \left( h(t)w(y)e^{-\frac{y^2}{2}} \right)^2 \, dy \, dt = 0.111 \frac{\sqrt{\pi}}{2} \text{deg} \cdot \text{sec} \] (A.3)

in which \( h(t) \) gives the time course of the test presentation (Fig. 2b) and \( w(y) \) is the half-toning function (which sets alternate 2 pixel lines to zero contrast). Explicitly including \( Q_t \) and \( Q_n \) in the model factors these stimulus-specific constants out of the mechanism vector and, in principle, makes the cone contrast weights independent of the spatio-temporal characteristics of the test and noise. Comparison with previously published cone contrast mechanism vector lengths from studies that did not explicitly take the \( Q \)'s into account (e.g., the summary in Eskew et al., 1999) requires multiplying the vector lengths in the present Table 1 by \( \sqrt{Q_t} \).

Without loss of generality we can let \( N_0=1 \), effectively scaling in terms of the intrinsic noise. It is convenient to express the relationship of Eq. (A.2) in polar coordinates:

\[ Q_t |\mathbf{f}|^2 |\mathbf{f}|^2 \left[ \cos(\alpha - \tau) \right]^2 = 1 + bQ_n |\mathbf{n}|^2 |\mathbf{n}|^2 \left[ \cos(\alpha - \nu) \right]^2 \]

where \( \alpha \), \( \tau \), and \( \nu \) are the angles of the mechanism, test, and noise vectors, respectively, in the \((L, M)\) plane. The vector length of the test at threshold is then
\[ |t| = \sqrt{\frac{1}{Q_t |f|^2 \left\{ \cos(\alpha - \tau) \right\}^2 + b Q_n |n|^2 \left\{ \cos(\alpha - \nu) \right\}^2}} \]

(A.4)

The first term in the radical represents the mechanism thresholds in the absence of noise, and, at a given test angle, is inversely proportional to the mechanism vector length, $|f|$. The second term in the radical represents the effect of the noise (of contrast $|n|$). Because the mechanism sensitivity $|f|$ affects the response to the noise and the test equally, the mechanism vector length does not appear in the second term, which is to say that the relative degree of elevation by the noise does not depend upon the sensitivity of the mechanism. Equation (A.4) was applied to each mechanism, simultaneously across all the noise conditions, to estimate values of the cone weights ($W_L, W_M$) (determining $|f|$ and $\alpha$), and $b$, with the responses of the mechanisms combined by probability summation (Minkowski exponent of 4.0; see Eskew et al., 1999).

Because thresholds in the no noise condition are lower, and thus more similar to one another, their variance is less than in the noise conditions (especially the $42^\circ/222^\circ$ and $48^\circ/228^\circ$ noise conditions). Without compensating for this difference in some way, the no-noise condition data would contribute much less to the fitting of the model. Therefore, in fitting the model for each observer, the data from each noise condition were weighted inversely to the variance of the thresholds in that condition.

It is worth comparing our modeling approach to the approach of Hansen and Gegenfurtner (2006). These authors did not fit their model, which contained a total of 4,096 mechanisms after the 16 pairs of base cone weights were randomly perturbed, to
their data. Instead, a Monte Carlo procedure was used to find tests $t$ that would produce a threshold response in the presence of chromatic noise, with the vector of cone weights $f$ fixed for each hypothetical mechanism. Because we were fitting measured thresholds, we varied the mechanism parameters instead of varying simulated tests. Stated simply, for a single mechanism and noise condition, we found the vector of cone weights $f$ that satisfied $f \cdot t = 1$, with $t$ determined by experiment, whereas Hansen and Gegenfurtner (2006) found values of $t$ satisfying the same relationship, with $f$ determined by assumption.


Namima, T., Yasuda, M., Banno, T., Okazawa, G., & Komatsu, H. (2014). Effects of luminance contrast on the color selectivity of neurons in the macaque area v4 and


Figure 1: An extension of the cardinal model that contains three rectified symmetric pairs of mechanisms: a separate mechanism for the detection of L-M and another for the detection of M-L, separate mechanisms for S-(L+M) and (L+M)-S, and separate mechanisms for a luminance increment (I) and decrement (D). The dashed lines represent sign inversion. Modified from Eskew (2009).
Figure 2: a) The test stimuli were circular Gaussian blobs ($\sigma=1^\circ$) b) The stimuli were presented with a rapid-start profile ‘sawtooth’ of 333ms total duration.
Figure 3: Chromatic noise. The lines randomly and independently changed from one chromaticity to a symmetrically opposite chromaticity (on the opposite side of the white point), so that the mean chromaticity was unchanged. Each line switched chromaticity with probability 1/2 at 18.75 Hz.
Figure 4: No-Noise condition. Detection thresholds (black discs) and model fits for three observers. Colored lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms. The same data and model are represented in Cone Contrast space (left column) and MacLeod, Boynton, Derrington, Krauskopf, and Lennie (MBDKL) space (right column). The red discs denote the 45° and 225° stimulus thresholds, for reference.
| TGS Mechanisms | $R^2=0.99$ | $\log_{10} b$ | Mech. Angle $\alpha$ (deg) | Mech. Vector Length, $|f|$ | Line Color |
|----------------|-----------|----------------|--------------------------|----------------|--------------|
| $R$            | 1.35 (0.13) | 313 (2.9)      | 510.0 (26.2)             | Red            |
| $G$            | 1.39 (-0.95) | 134 (2.5)      | 510.0 (22.6)             | Green          |
| $O$            | 1.10 (-1.04) | 333 (5.6)      | 90.3 (8.5)               | Orange         |
| $B$            | 0.50 (-0.76) | 152 (5.2)      | 90.2 (7.9)               | Blue           |
| $Y$            | 1.57 (-0.44) | 18 (8.6)       | 200.9 (14.2)             | Yellow         |
| $P$            | 1.26 (-0.98) | 238 (9.6)      | 141.21 (23.9)            | Purple         |

| CLM Mechanisms | $R^2=0.99$ | $\log_{10} b$ | Mech. Angle $\alpha$ (deg) | Mech. Vector Length, $|f|$ | Line Color |
|----------------|-----------|----------------|--------------------------|----------------|--------------|
| $R$            | 1.87 (0.71) | 317 (2.9)      | 1373.8 (70.1)            | Red            |
| $G$            | 1.71 (0.25) | 133 (2.9)      | 1123.1 (57.8)            | Green          |
| $O$            | 1.04 (-1.30) | 312 (5.4)      | 458.7 (43.2)             | Orange         |
| $B$            | 1.38 (-0.83) | 139 (3.9)      | 325.6 (22.3)             | Blue           |
| $Y$            | 1.56 (-1.41) | 11 (0.6)       | 180.3 (10.5)             | Yellow         |
| $P$            | 1.36 (-1.27) | 193 (0.5)      | 158.1 (6.2)              | Purple         |

| SAF Mechanisms | $R^2=0.99$ | $\log_{10} b$ | Mech. Angle $\alpha$ (deg) | Mech. Vector Length, $|f|$ | Line Color |
|----------------|-----------|----------------|--------------------------|----------------|--------------|
| $R$            | 1.33 (-0.61) | 310 (0.7)      | 694.8 (8.1)              | Red            |
| $G$            | 2.11 (-0.95) | 133 (1.5)      | 799.8 (20.7)             | Green          |
| $O$            | 6.94 (11.1)  | 334 (3.3)      | 567.3 (31.4)             | Orange         |
| $B$            | 2.19 (-0.76) | 149 (0.6)      | 459.7 (5.1)              | Blue           |
| $Y$            | 1.53 (-0.44) | 67 (6.6)       | 277.8 (27.7)             | Yellow         |
| $P$            | 1.28 (-0.98) | 234 (5.9)      | 157.1 (15.2)             | Purple         |

Table 1: Best fitting parameter values for the six-mechanism model (with standard errors). The ‘line color’ is the color of the mechanism threshold line in the detection contour plots. Observers: TGS, CLM, and SAF.
Table 2: Peaks of the fitted contour in the no-noise, 42°/222°, 48°/228°, and 64°/244° noise conditions for observers TGS, CLM, and SAF.

<table>
<thead>
<tr>
<th></th>
<th>No-Noise</th>
<th>42°/222° Noise</th>
<th>48°/228° Noise</th>
<th>64°/244° Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGS</td>
<td>54° &amp; 217°</td>
<td>44° &amp; 225°</td>
<td>47° &amp; 228°</td>
<td>57° &amp; 234°</td>
</tr>
<tr>
<td>CLM</td>
<td>46° &amp; 229°</td>
<td>43° &amp; 223°</td>
<td>46° &amp; 226°</td>
<td>57° &amp; 234°</td>
</tr>
<tr>
<td>SAF</td>
<td>48° &amp; 227°</td>
<td>40° &amp; 222°</td>
<td>N/A</td>
<td>58° &amp; 233°</td>
</tr>
</tbody>
</table>
Figure 5: 42°/222° Noise condition. Detection thresholds (black discs) and model fits for three observers. Colored lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms. The same data and model are represented in Cone Contrast space (left column) and MacLeod, Boynton, Derrington, Krauskopf, and Lennie (MBDKL) space (right column). The red discs denote the 45° and 225° stimulus thresholds and the cyan discs indicate the 42° and 222° stimulus thresholds and also the noise direction. Notice the scaling change compared to Figure 4: cone contrast thresholds are larger (5-20 fold) when noise is added.
Figure 6: 48°/228° Noise condition. Detection thresholds (black discs) and model fits for two observers. Colored lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms. The same data and model are represented in Cone Contrast space (left column) and MacLeod, Boynton, Derrington, Krauskopf, and Lennie (MBDKL) space (right column). The red discs denote the 45° and 225° stimulus thresholds and the cyan discs indicate the 48° and 228° stimulus thresholds and the noise direction.
Figure 7: 64°/244° Noise condition. Detection thresholds (black discs) and model fits for three observers. Colored lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms. The same data and model are represented in Cone Contrast space (left column) and MacLeod, Boynton, Derrington, Krauskopf, and Lennie (MBDKL) space (right column). The red discs denote the 45° and 225° stimulus thresholds and the cyan discs indicate the 64° and 244° stimulus thresholds and the noise direction.
Figure 8: Threshold elevations (relative to the No Noise contour), as a function of angular deviation between the test and the noise angle in QI. Observers TGS, CLM, and SAF in the three added noise conditions (green-42°, blue-48°, and red-64°). The colors of the lines and discs are arbitrary.
Figure 9: Predicted six-mechanism model contours in multiple noise conditions for three observers: No Noise (black), 42°/222° (green), 48°/228° (blue), 64°/244° (red) noise conditions. The colored discs indicate the noise direction and the black dotted line indicates the main diagonal, 45°/225° axis, for reference. The contour tilts toward the noise direction, and, consistent with previous results (R. Eskew, Newton, & Giulianini, 2001; Giulianini & Eskew, 1998), there is less selectivity as the noise is moved away from the ends of the contour.
Figure 10: Six-mechanism model. The dashed lines indicate sign inversion (subtraction). Each mechanism is half-wave rectified. Four of these mechanisms (R and G, and B and O) are ‘quasi-paired’, having nearly equal and opposite L and M cone weights. Two mechanisms (Y and P) have additive L and M cone inputs, and are asymmetric (unpaired). The S cone input to B and O is by analogy to the cardinal model (Figure 1); it was not studied in the present experiments.
Supplementary Information

Uncertainty in Model Parameter Estimates Across Noise Conditions

Table 1 gives the best-fitting values of the parameters of the model, with asymptotic standard errors. Figure S1 provides a visualization of the effects of those standard errors for observer TGS, in which the best-fit value is perturbed by 0, +1, or -1 standard errors, and the EvN model (Eq. 1) is used to calculate the effect of the noise on each of the perturbed mechanism thresholds.

Since there are three parameters for each mechanism, there are 27 combinations of perturbed parameter values. In each panel of Figure S1, there are thus 27 threshold lines for each mechanism (colored as in Figures 4-7). Most of those 27 lines fall into three clusters, because the error in estimating the mechanism angle dominates the other errors. Naturally the effects of the uncertainty are greater farther away from the actual measurement (represented here by the contour).
Figure S1: The colored lines represent mechanism thresholds and the smooth closed contour (black) is the probability sum based on the best-fitting mechanism values (i.e., it is the same as in Figures 4-7). The solid lines are generated from the best fitting parameters and the dashed lines are a perturbation of those values (Table 1)—plus/minus one asymptotic standard error for observer TGS.

**Six-Mechanism Model With Symmetry Constraints**

For all observers, six mechanisms were required to generate the best fit to all of the data, combining all of the noise conditions. At least two of those six mechanisms (Y & P) are clearly asymmetrical (i.e. unpaired), with mechanism angles that do not differ by 180° and different sensitivities. The other four mechanisms (R, G, B, and O) are quasi-paired—they do not have exactly equal and opposite cone weights but appear to be roughly symmetrical along the flanks. Perhaps these are actually two parallel pairs.

For this reason, we re-fit the model with various symmetry constraints on the mechanisms. We constrained R and G, and B and O, to have cone weights that were equal in magnitude and opposite in sign (i.e., we made them two opponent pairs). The b values were constrained to be the same for each pair of mechanisms. Thus there were 9 free parameters. Although this model provided a good overall fit to the data of all three main observers ($R^2 > 0.97$), the fits were significantly worse than the original six-mechanism model, even after taking into account the greater number (18) of free parameters in the original model—observers CLM: $F(85,76) = 1.63, p=0.015$, TGS: $F(83,74) = 2.03, p=0.001$, and observer SAF: $F(53,44) = 2.29, p=0.002$.

Thresholds were also fit utilizing the same approach but with 12 free parameters, where the symmetry constraint was enforced on 3 pairs of mechanisms, but the sensitivity for each mechanism was allowed to vary: 1.) L- and M- cone weights for each of the three ‘paired’ mechanism (six free parameters) 2.) The b values were free to vary for each pair of mechanisms (six additional free parameters). All model fits again had $R^2 > 0.97$
but, again the fit with symmetry constraints was significantly worse than the original six-mechanism model—observer CLM: $F(82,76) = 1.55, p=0.027$, observer TGS: $F(80,74) = 1.68, p=0.012$ and observer SAF: $F(50,44) = 1.91, p=0.015$.

In summary, for all observers, the symmetry and sensitivity constraints placed on the six-mechanism model produced significantly worse fits, even accounting for the reduced number of free parameters. Although these models provided a worse fit, all of them did produce selective masking.

*Eight-Mechanism Model*

Here we add two mechanisms to the original six (six additional free parameters) and refit the data (24 free parameters total). For each observer, the first four mechanisms (R, G, B, and O) in Figure S2 (only TGS shown) are similar to those Results shown in Table 1. Adding two mechanisms did not significantly improve the original model: TGS: $F(74,68) = 0.89, p=0.68$ and CLM: $F(76,70) = 1.10, p=0.34$ and SAF: $F(44,38) = 1.46, p=0.12$. The simpler six-mechanism model provides just as good a fit as the eight-mechanism model. It is unnecessary to add more mechanisms.
| Mechanisms | Log$_{10} \ b$ | Mech. Angle $\alpha$ (deg) | Mech. Vector Length, $|\mathbf{f}|$ |
|------------|--------------|-----------------|-----------------|
| $R$        | 1.22         | 310             | 292.1           |
| $G$        | 1.32         | 133             | 487.1           |
| $O$        | 1.11         | 334             | 85.6            |
| $B$        | 6.03         | 156             | 243.5           |
| $V$        | 0.71         | 167             | 66.2            |
| $VI$       | 1.62         | 20              | 198.7           |
| $VII$      | 1.89         | 316             | 379.5           |
| $VIII$     | 6.36         | 278             | 207.7           |

| Mechanisms | Log$_{10} \ b$ | Mech. Angle $\alpha$ (deg) | Mech. Vector Length, $|\mathbf{f}|$ |
|------------|--------------|-----------------|-----------------|
| $R$        | 1.49         | 318             | 251.4           |
| $G$        | 1.29         | 133             | 459.5           |
| $O$        | 0.99         | 313             | 425.3           |
| $B$        | 0.69         | 151             | 108.9           |
| $V$        | 10.15        | 334             | 454.0           |
| $VI$       | 0.57         | 133             | 1154.0          |
| $VII$      | 12.26        | 196             | 148.4           |
| $VIII$     | 11.32        | 313             | 1178.8          |
| SAF Mechanisms | $\log_{10} b$ | Mech. Angle $\alpha$ (deg) | Mech. Vector Length, $|f|$ |
|----------------|-------------|--------------------------|-------------------|
| $R$            | 1.19        | 310                      | 575.1             |
| $G$            | 1.59        | 133                      | 831.3             |
| $O$            | 1.28        | 336                      | 77.0              |
| $B$            | 2.50        | 158                      | 370.4             |
| $V$            | 15.13       | 15                       | 308.8             |
| $VI$           | 7.55        | 101                      | 298.7             |
| $VII$          | 7.92        | 294                      | 351.6             |
| $VIII$         | 9.08        | 320                      | 715.3             |

Table S1 Best fitting parameter values for the six-mechanism model. Observers a) TGS b) CLM c) SAF
Figure S2: Eight mechanism EvN model. Measured thresholds (red discs) for observer TGS in multiple noise conditions: (a) No Noise (b) $42^\circ/222^\circ$, (c) $48^\circ/228^\circ$, and (d) $64^\circ/244^\circ$. The black lines represent mechanism thresholds and the smooth closed contour (blue) is the probability sum of these mechanisms. The observer’s thresholds are represented in MBDKL space.
Sixteen Equally Spaced Mechanisms

A major aim of this study was to determine whether a “higher order” model—with many mechanisms—was required to produce selective masking and account for the data along the detection contours. We began with the model of (Hansen & Gegenfurtner, 2006). Their model was based initially on having 16 equally spaced mechanisms in the (L,M) plane of MBDKL space. This set of angles included the four cardinal mechanism directions along with intermediates (i.e. 0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5°, 180°, 202.5°, 225°, 247.5°, 270°, 292.5°, 315°, 337.5°). Hansen and Gegenfurtner (2006) found that this model did not fit their data, so the mechanism angles were randomly perturbed across the spatial extent of their test stimulus, such that the actual model applied to the data had as many different mechanisms as there were test stimulus elements (i.e. 16 x 16 (pattern of squares) x 16 (equally spaced mechanisms) =4,096 total mechanisms across the extent of the test). Here we did not apply this random perturbation, but instead only used the starting 16 mechanisms. Each of our mechanisms obeyed the same EvN relationship as used in our six-mechanism model (Appendix A), unlike those of Hansen and Gegenfurtner (Hansen & Gegenfurtner, 2013). While we did not randomly perturb the mechanism angles, we attempted to vary the starting angle of the set of mechanisms to produce the best fit possible; for example, we shifted all sixteen mechanism directions by 10° counterclockwise. However, the best fits were obtained with the basic set, containing the four cardinal mechanisms ((L-M), (M-L), S-(L+M), and (L+M)-S) plus the intermediates.

For half of the mechanisms – those in the right half plane of MBDKL space – the vector length |f| (Eq. 1) was allowed to vary, as was the constant of proportionality b. The
other half of the mechanisms formed symmetric pairs with the first half. The net result is that 16 parameters were free to vary.

Figure S3 shows thresholds from observers TGS, CLM, and SAF with the fitted mechanism contours for this 16 mechanism model (only MBDKL space is shown). The black lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms.

This 16 mechanism model does produce selective masking. However, for all three observers, the fits are significantly and substantially worse than the six-mechanism model fits ($F_{(76,74)} = 1.93$, $p=0.002$ for TGS; $F_{(78,76)} = 2.30$, $p=0.0001$ for CLM; $F_{(46,44)} = 3.25$, $p=0.00007$ for SAF). For each observer, the predicted detection contour obviously fails to fit the thresholds in one or more regions of the data (e.g., the $42^\circ/228^\circ$ noise condition for TGS). In particular, the increment/decrement asymmetry seen in our thresholds cannot be captured by this symmetric model; this is particularly noticeable for CLM.

If the angles of the sixteen-mechanisms were allowed to vary freely (i.e. there was no equal-spacing constraint), the fit would clearly be as good as the 6-mechanism model, since the six-mechanism model would in that case be merely a subset of the sixteen mechanisms.
Figure S3: Sixteen-mechanism model—each mechanism is equally spaced & equally sensitive. Measured thresholds (red discs) for three observers TGS, CLM, and SAF in multiple noise conditions: (a) No Noise (b) 42°/222°, (c) 48°/228°, and (d) 64°/244°. Black lines represent the mechanism thresholds and the smooth closed contour (blue) is the probability sum of these mechanisms. The observer’s thresholds are represented in MBDKL space. Subject SAF did not participate in the 48°/228° condition (panel (c) is missing).
Chapter 4: Labelling the lines: A test of a six-mechanism model of detection
Abstract

Six linear chromatic mechanisms are sufficient to account for the pattern of threshold elevations produced by chromatic noise masking in the L,M plane of cone space (Shepard et al, 2016). The key feature that allows this model to predict selective masking across multiple noise conditions is having more than two detection mechanisms with opposed L and M cone inputs. Here, we report results of asymmetric color matching of the threshold-level tests from that detection study, and use those matches to test the detection model. We assume the mechanisms are univariant labelled lines (Rushton, 1972; Watson & Robson, 1981), implying here that the colors of physically-different stimuli that are detected by a single mechanism should all be the same, and suggesting that the colors of two stimuli detected by different mechanisms should differ. Physically-different stimuli that lie along one mechanism threshold line should therefore produce metameric color matches – post-receptoral metamers – while tests that lie on two different mechanism lines should be matched with different colors. The results show that color matches fall into no more than 6 clusters in CIE (u’, v’) space (across all the noise conditions), and these clusters correspond closely to the six mechanisms in the model. Most importantly, where the detection model determines that a given test angle is detected by different mechanisms in different noise conditions (due to differential masking), the hue of that test angle changes in a consistent way. These color matches essentially apply a color label to each of the mechanism threshold lines, confirm the six-mechanism model, and quantify the hue signaled by each mechanism.
Introduction

Recently we described a chromatic detection model that contains only six mechanisms (Shepard et al, 2016), but, unlike the cardinal model (Krauskopf, Williams, & Heeley, 1982), accounts for selective masking in the L,M plane of color space (Hansen & Gegenfurtner, 2006). This model differs in one important way from the cardinal model: four of our six detection mechanisms have opposed L and M cone inputs, rather than just two as in the cardinal model. This is the critical feature that allows the model to account for selective masking when noise is placed near the ends of the detection contour in the (L,M) plane; noises that are nearly parallel to the long flanks of detection contours can cause different mechanisms to become most sensitive and thus determine threshold, tilting the overall detection contour (see figure 9 in Shepard et al, 2016). This model has not yet been tested in full LMS color space, and it is possible that additional color mechanisms may have to be added to account for other data, but in the LM plane the six-mechanism model provides an excellent and parsimonious account of detection across a variety of masking noise conditions.

An outline of the model is shown in Figure 1. In Shepard et al., (2016), the mechanisms were labelled with single letter mnemonics for the hue they signal ((R)ed, (G)reen, (Y)ellow, (B)lue, (O)range, and (P)urple); those hues were determined by the observers informally, by naming. In the current paper we formally measure those hues. Throughout this paper, we refer to colors and color matches by color names (e.g., “orange”), and detection mechanisms by the same single letters (e.g., “O”) used in Shepard (2016).
The top and bottom pairs of mechanisms in Figure 1 take weighted differences of the L and M cone signals, with different weights for all four mechanisms (within each row, the weights are similar in magnitude but of opposite sign). The middle pair of mechanisms takes the sum of the L and M cone signals, again with weights that were similar in magnitude but of opposite sign for the two mechanisms in the row; an S cone input of opposite sign is shown; this input is speculative because the model has not yet been tested with S cone modulations.

In our detection model, thresholds mediated by a single mechanism follow an Energy vs. Noise (EvN) relationship based upon the theory of noise masking (Giulianini & Eskew, 2007; Wang, Richters, & Eskew, 2014), in which the contrast energy of the test is linearly related to the noise contrast power. Mechanisms are required to have the same relative cone weights (the same mechanism angles) across all of the noise conditions (see Table 1 in Shepard et al, 2016 and Appendix for details); this means that mechanism threshold lines have the same slope in all noise conditions (i.e., the threshold lines are perpendicular to the mechanism vector Eskew, McLellan, & Giulianini, 1999). Moreover, the threshold elevation produced by chromatic noise at a given angle is proportional to the squared cosine of the difference in the mechanism and noise angle, meaning that the relative distance to the mechanism threshold line is determined by theory rather than being freely fit to data. These model features are derived from theory and highly constrain the detection model.

We define a mechanism as a “fixed (relative) combination of cone signals that is correlated with the observer’s behavior in psychophysical experiments (Eskew, 2008).” Mechanisms are assumed to be univariant, which implies that when two stimuli are
detected by one and only one mechanism, there is some relative intensity at which the
two stimuli cannot be discriminated from one another. In addition, they are “labelled
lines” which implies that when two stimuli are detected by different mechanisms, they
must be discriminable from one another at all relative intensities. Taken together, these
assumptions imply that each mechanism is characterized by one and only one hue, and so
color matches should cluster into no more than six groups. By comparing the color
matches with the model mechanism threshold lines, we can ‘label the lines’ and further
test the model. It is important to keep in mind that the detection model was determined
solely by the two-alternative threshold data; it was fit to those data prior to measuring the
color matches.

The main purpose of the present color matching study was to use asymmetric
color matching to determine the hue of the tests that are detected by each of the
mechanisms in Figure 1. The aforementioned properties of a mechanism were used to
interpret the color matches. Examples are shown in Figure 2. A M-cone increment test
may produce the same color as a L-cone decrement test because both test thresholds lie
on the same univariant mechanism threshold line (Figure 2a). All of the other color
matches along this mechanism line should also be of the same chromaticity. The labelled
line assumption implies that when tests lie on two different mechanism lines, they should
be matched with two different chromaticities. In this example, at threshold, a 135° test in
cone contrast space should produce a different color match than a 225° test since they are
detected by two separate mechanisms (Figure 2b).
Methods

Observers

The same three well-practiced observers who participated in the detection experiment (Shepard et al. 2016) also participated in the color matching experiment presented here. All had normal scores on the Farnsworth-Munsell 100 hue test (Farnsworth, 1943) and the Ishihara Plates. Northeastern University’s Institutional Review Board approved the research protocol; the procedures comply with the Declaration of Helsinki.

Apparatus

Stimuli on the main experimental monitor were created on a Power Macintosh computer and displayed on a SONY GDM-F520 CRT monitor running at 75hz, using an ATI Radeon 7500 video board with a driver verified to support the 10-bit digital-to-analog converters (DACs). Spectroradiometric calibration was performed at 4 nm intervals across the spectrum using a Photo Research PR-650 spectroradiometer, and the monitors were linearized with gamma correction lookup tables. The color matches were determined on a 15-inch MacBook Pro using an Intel HD 4000 graphics card that was placed off to the side; a piece of black cardboard was used to mask the screen, showing only the matching stimuli. All experiments were conducted in a dark room. Head position was stabilized with a chin and forehead rest.

Observers were corrected to normal visual acuity using a trial lens that was placed in front of their dominant eye; the other eye was patched. The trial lens was used to view the experimental monitor but was not used when the observer made the color matches on
the peripheral monitor (which was positioned close to the observer). Observer SAF is an emmetrope, CLM is slightly myopic (-1.00D, best sphere) and observer TGS’ correction is 3.25D (best sphere). Neither CLM nor TGS reported any difficulty when selecting a color match on the second monitor without refractive correction.

**Test and Noise Stimuli**

Stimuli were identical to those used by Shepard et al (2016). Test and noise stimuli are specified by their polar angle in the L,M plane of cone contrast space, with 0° and 90° representing L and M cone increments respectively. The bipolar noise has two angles, separated by 180°; the noise direction will be identified by the first pole (e.g. the noise direction 42°/222° will simply be called 42° noise).

The background of the test and color matching stimuli was the same mid-grey (CIE x=0.30, y=0.31). The tests were circular Gaussian blobs (σ=1°), presented with a rapid-start ‘sawtooth’ profile of 333ms total duration (Shepard et al., 2016, Fig. 2).

The masking noise consisted of horizontal lines that were superimposed on the test. The lines randomly and independently changed from one chromaticity to a symmetrically opposite chromaticity (on the opposite side of the white point). The noise contrast was always 90% of the maximum available at the noise direction; in cone contrast units, the noise cone contrast vector length |n| was 0.498, 0.414, and 0.267 for the 42°, 48°, and 64° noises, respectively. Further details of the test and noise stimuli may be found in Shepard et al, 2016).

**Procedure**

Each observer was presented the threshold-level tests, in the same noise conditions from the detection experiment: one no-noise, and up to three noise conditions
(i.e. $42^\circ$, $48^\circ$, and $64^\circ$ in cone contrast space). TGS and observer CLM participated in all four noise conditions and SAF participated in 3 noise conditions, since SAF only collected thresholds in three conditions in the detection experiment (Shepard et al. 2016).

For color matching, test angles were blocked by noise condition, and run in random order within that block; the observer did not know which test was being presented. For each test angle, the first step was to confirm that the test contrast was correctly set at threshold. The observer was shown the test, fixed at its measured threshold from the detection experiment (Shepard et al, 2016), in one of two temporal intervals, just as in the threshold measurement procedure. If the observer correctly selected the test interval, then the color match was performed. If the observer did not correctly select the test interval, the color match was not performed on that trial and the next trial in the block was presented. In each block, the observers selected the correct trial ~80% of the time, confirming that thresholds did not change between the collection of detection thresholds and the color matching experiment. After a correct response, the observer looked at the test, then turned to the laptop positioned off to the side and selected a matching chromaticity on a HSV color wheel (Figure 4). When satisfied, the observer confirmed their match by viewing a disc on the same peripheral display that was roughly the same size as the threshold level test. The disc was presented on a grey background with the same chromaticity coordinates as the main experimental monitor (Figure 4). When the observer was finally satisfied, the H.S.V. values of the matches were recorded, and a new test angle was selected. Five matches were made at each test angle, chosen in random order, in each noise condition; matches in a given noise condition were collected in at least two sessions on different days. Although observers
were free to vary the brightness of the matching stimulus, none of them did, so the analysis here is based only upon hue and saturation. After the experiment was completed, the (x, y, L) chromaticity coordinates of each of the five matching discs were measured on the laptop with an X-Rite i1Display Pro. These five chromaticities were averaged; the average of the five matches was then transformed to CIE (u’,v’) space. The averaged match coordinates of each test were used as the basic data in all further analysis of the color matches.

**Results and Discussion**

The first step in the analysis was to pool all the (u’, v’) values for a given observer (pooling across noise conditions), and subject the entire set of matches to a k-means cluster analysis without constraining the number of clusters. This analysis parallels the modeling in Shepard et al., (2016), in which the detection model was fit to the data from all noise conditions simultaneously (however, the color matches were measured after the threshold model was fit, and the two analyses are completely independent). For each observer, six clusters of color matches were found by the k-means analysis. In fact, the clusters were completely non-overlapping so the formal analysis was unnecessary. For each cluster considered separately, a 95% error ellipse around the centroid was computed based upon principal components analysis. For example, Figure 5 shows all of the color matches, along with the six ellipses, from the pooled analysis for observer TGS. In Figure 6-9 below, the matches for each noise condition are plotted in separate panels, along with the same six ellipses from the pooled analysis; these ellipses are colored to approximately represent the appearance of the stimuli in the cluster.
No-Noise condition

Figure 6, left column, shows the color matches in the no-noise condition, for each observer, as black discs. In each panel, the point labelled “WP” is the chromaticity of the test and matching monitor grey backgrounds. The matches were, as expected, unsaturated, as is evident by their close proximity to the white point of the monitor; the yellow and orange clusters were the least saturated for all observers.

First examine panel (e), at bottom left, for observer SAF. There are five points near the green error ellipse, representing matches for the five test stimuli in the no-noise condition that fall into this cluster; these test stimuli have color angles ranging from 64° to 195° (the values in square brackets near the ellipse). Four tests (ranging from 42° to 52°) fall into the yellow cluster, two into orange, four into red, four into purple, and only one in the blue cluster. There is one match in each of the red, yellow, and green clusters that is hidden by another match with very similar u’,v’ coordinates.

Panels (a) and (c) show the no-noise color matches for the other two observers, with the same conventions used for SAF in panel (e). Note that for both of these observers, unlike SAF, no matches fall into the orange or blue clusters; recall that the clusters are based upon pooling the data across noise conditions, and (as shown below) orange and blue matches were made by TGS and CLM to tests in the presence of noises, but not in the no-noise condition.

The right-hand panels in Figure 6 show the two-alternative forced-choice detection data from Shepard et al. (2016) as colored discs. In this L,M plane of cone contrast space, the origin represents the mean adapting field and the axes represent the modulation of the cones relative to the excitation produced by the mean adapting field.
The lines represent mechanism thresholds from the model-fit to the detection data, with up to six lines shown in each panel, and the solid contour shows the probability sum of the mechanisms. The threshold points have been colored to approximately represent the mean of the five color matches for each test angle. The mechanism threshold lines have also been colored, with a more saturated version of the hue of the matches falling along, or very near the lines.

Note that for TGS and CLM, the detection model places the O and B mechanism lines well outside of the data. Thus according to the detection model, these mechanisms do not contribute to thresholds in this condition, and the slope and placement of the mechanism lines depend only upon thresholds measured in the other noise conditions (see Shepard et al. 2016 for discussion). The lack of contribution of these mechanisms to threshold in the right-hand panels exactly parallels the lack of blue and orange color matches in the left hand panels, independently determined.

For all observers, the number of filled clusters of color matches and the number of detection threshold mechanisms agrees perfectly, despite these two types of data being collected independently (and weeks or even months apart). Tests angles within any given cluster generally align with only one mechanism line in the model fit (i.e. univariance) and each mechanism line is characterized by a different hue (i.e. labelled lines).

42° Noise

Figure 7 shows data from the 42° noise condition, in the same format as Figure 6. The theoretical constraints of the model require the mechanisms to have the same relative cone weights, and thus the same mechanism angles across all of the noise conditions (see Table 1 in Shepard et al, 2016 and Appendix for details); thus the mechanism threshold
lines (at 90° to the mechanism angles) have the same slope in all of the noise conditions, and their distance from the origin is constrained by the noise power and angle; see Shepard et al., 2016 and Wang et al. (2014).

Note that Figure 7 shows there is now an orange cluster in CLM’s matches (panel c), and that the detection model has the O mechanism now contributing to threshold (panel d). Also, the R mechanism lies outside the detection contour for CLM, and there are no red color matches. It is important to emphasize that the detection model was fit to the thresholds entirely independently of the color matches (in fact, the detection model was fit prior to measuring the color matches); the only change to the detection model from the color matching experiment is the assignment of the color. For TGS and SAF, thresholds along the long flanks are attributed to the R and G mechanisms (panels b and f), whereas for CLM (panel d) the model asserts that the R mechanism has been sufficiently masked that it does not contribute, and the lower flank is attributed to O. The color matches are consistent: TGS and SAF have reddish and greenish matches, whereas the flanks are seen as orangish and greenish for CLM.

In general, the agreement between the two procedures is excellent in this condition, as it is in the no-noise condition. However, a discrepancy exists for TGS (Fig. 7a and b) at the ends of the 42° noise detection contour. According to the detection model, the Y and P mechanisms lie somewhat outside the detection data, and the ends of the contour are served by the O and B mechanisms. However, the color matches show these thresholds fall into the purple and yellow clusters. This discrepancy will be discussed in a separate section below.
48° Noise

Figure 8 shows the matches (left) and thresholds (right) in the 48° noise condition, for CLM and TGS (SAF did not participate in this noise condition). As shown in Figure 8a, TGS’s color matching data clusters into five clusters (red, green, orange, purple, and blue, which has only one match). Similarly, five mechanisms are required to account for TGS’s test thresholds (panel b). The R and G mechanisms detect the majority of the tests along the long flanks, the P mechanism accounts for the decrement thresholds (i.e. end of the contour in quadrant III), the O and B mechanisms detect tests near ends of the contour (i.e. quadrant I and III), and the Y mechanism does not contribute to detection in this condition. Therefore, the detection model predicts there will be no color matches falling in the yellow cluster, and this is the case (Fig. 8a).

CLM’s color matches fall into four clusters (Figure 8c). Interestingly, the mechanisms that detect the majority of the tests along the flanks once again switch with the addition of the 48° noise (panel d). Tests along the upper flank are detected by the B mechanism, and tests along the bottom flank are detected by the R mechanism. Thus the same range of test angles that were detected by the R mechanism without noise (Fig. 6d) and have matches in the red cluster (Fig. 6c), are detected by the O mechanism in the 42° condition (Fig 7d) and have orange matches (Fig 7d); these angles then switch back to red in the 48° noise condition (Fig 8c and d). The detection model determines that a given test angle can be detected by different mechanisms in different noise conditions; the mechanism definitions (univariance and labelled lines) imply that the color matches should change in a consistent way with the mechanism switching, and they do.
Figure 9 shows the matches and thresholds in the presence of the 64° noise. TGS’s detection model (9b) has all six mechanisms contributing to the detection thresholds, and his matches fall into all of the six clusters. CLM’s thresholds (Fig. 9d) are determined by only four mechanisms, and her color matches fall into the corresponding four clusters (Fig. 9c); note that the noise has moved the G mechanism out of the way, leaving B to determine threshold, and correspondingly there are no green matches for this observer (compare with Figs 6c and d, 7c and d, and 8c and d). SAF’s thresholds require five mechanisms (Fig. 9f) and the color matches fall into the corresponding five clusters (Fig. 9e).

In general, the agreement between the two datasets in the presence of 64° noise is again excellent. However, there is one peculiar point in quadrant I for SAF. The 64° test angle (highest point in Figure 9f) is matched with a bluish hue which appears to suggest detection by the B mechanism. However, this point lies above the Y mechanism, which implies that this test should be detected by Y and produce a yellow match. This discrepancy, like the one discussed above in connection with Figures 7a and b, involves the Y (and P) mechanisms at the ends of the detection contour, and will be discussed below.

**Differential Masking**

There are a number of instances where the detection model determines that a given test angle is detected by different mechanisms in different noise conditions (due to differential masking) even though of course a given test angle has the same relative L and M cone excitations, irrespective of noise condition. A few examples of this finding were
noted above (see the discussion on CLM’s detection contour flanks, for example), but this point is so important it deserves additional emphasis.

The agreement between the differential effects of the noise on thresholds and on color matches shows that the two measurements do not just happen to be correlated, but are subject to the same manipulation, the variation in mechanism sensitivity produced by masking. Consider as an example the upper flank in Figure 7f. The color match, at 135°, consists of two cone contrasts that are of equal magnitude but opposite sign \[\frac{\Delta L}{L}/\frac{\Delta M}{M} = -1\], and a corresponding pair of ratios of the red, green, and blue primaries of the test monitor (r/g, g/b). At threshold in the presence of the 42° noise, this physical/physiological modulation is matched with a ‘greenish’ light on the other monitor, and falls into the green cluster in Figure 7e. In the presence of 64° noise, a stimulus consisting of the identical ratio of cone signals, and identical pair of primary ratios (r/g, g/b), is matched with a ‘bluish’ light by the same observer (Figure 9f). The appearance changes because the change in noise condition makes G less sensitive than B, according to the detection model, despite there being no change in the relative cone (or monitor primary) modulations.

The basic observation is not novel, of course. Similar effects can be found in, for example, the color matching results of Webster & Mollon, (1994), who used habituation rather than noise, and especially in (Giulianini & Eskew, 1998)–dashed line in Fig. 3a), who also related the change in appearance to a specific switch of detection mechanisms. Here, however, the correspondence between the detection model, fit only to thresholds, and the color matches, occurs in many cases throughout Figures 6-9, and is very strong.
evidence supporting the detection model and the univariance and labelled-line assumptions.

**Discrepancies involving Y and P**

In general, the correspondence between the mechanisms of the detection model and the independently-determined color matches is excellent. But, as mentioned above, there are three clear discrepancies, all of which involve tests at the ends of the detection contours. In the 42° noise condition for TGS (Fig. 7b), the detection model attributes detection of the highest thresholds (in QI and QIII) to the O and B mechanism, and the Y and P mechanism lines lie somewhat outside these extreme thresholds. However, the color matches show these thresholds fall into the yellow and purple clusters. Similarly, a single test (at 64°) in the 48° noise condition for SAF (uppermost point in Figure 9f) should be detected by Y, yet its color match falls in the blue cluster (Fig. 9e).

Both of the discrepant cases that involve Y are likely due to the difficulty of accurately estimating the relative cone weights (mechanism angle) of mechanisms at the ends of detection contours, where very few thresholds are available to constrain the mechanism estimates. For TGS (Fig. 7b), if the Y threshold line were slightly less steep (i.e., less L cone input), and thus slightly less sensitive to the masking effect of the noise, the result would be that these end thresholds would be detected by Y and would therefore agree with the color matches. For SAF (Figure 9f), increasing the L cone weight to Y (steepening the Y threshold line) just slightly would suffice. (Note also the analogous underestimate of Y sensitivity in Fig 7e, although these tests are correctly matched with yellow). A similar adjustment to the cone weights in P would make the highest
thresholds in QIII in Figure 7b (decrement end of the detection contour) detected by P, producing agreement between detection and color matching for those four test angles.

However, it is not likely that simply adjusting the weights of Y and P would alone suffice (consistent with trial calculations). The mechanisms in the detection model are all linear (except for rectification), and the model fits are the best estimates of those mechanisms given linear cone combinations (and the constraints of the EvN model). Because the few small discrepancies between detection and hue occur with the highest thresholds, it is quite likely that these mismatches involve nonlinearities in the mechanisms, nonlinearities that were not included in the detection model of Shepard et al. (2016). Unfortunately, the data do not provide sufficient constraints on the model to estimate such nonlinearities (see Shepard et al. 2016, for discussion).

**Single cone hues**

It has often been reported that, for some subjects and under some conditions, incremental M cone stimuli produces hues that are cyan – as more blue than green (e.g., Drum, 1989; Schirillo & Reeves, 2001). These observations were made with stimuli that were well above threshold. The present results show that, at threshold, the G and B mechanisms have similar sensitivity to $+\Delta M/M$ stimulation, with G more sensitive without noise and with $42^\circ$ noise, and, for some observers, B somewhat more sensitive with $48^\circ$ and $64^\circ$ noises. The model predicts that suprathreshold M cone increments should stimulate both G and B, with relative sensitivity depending upon conditions (e.g., perhaps with the size and duration of the stimulus).
Similarly, suprathreshold +ΔL/L stimuli would stimulate the R and O mechanisms, with the balance varying over conditions. Analogous points could be made with respect to cone decrements.

**Mechanism Hues**

As shown in the left panels of Figures 6-9, the chromaticities associated with the quasi-paired mechanisms of Figure 1 do not fall exactly on lines that go through the white point, i.e., they are not colorimetric complements. This is qualitatively consistent with the fact that the cone weights for R and G, for example, are not exactly of equal magnitudes (although they are of opposite sign; see Shepard et al., 2016).

The current results indicate a clear and consistent relationship between detection mechanisms and clusters of color matches. Modeling this relationship – predicting the actual color matches – is, however, not straightforward. Unlike the threshold tests, the matching stimuli were steadily viewed rather than flashed, were suprathreshold, and perhaps most importantly, were not silent for the S cones, unlike the L,M plane thresholds. In fact, calculations show that the centroids of the color matches represent variations in all three cone types, and thus cannot be predicted from combinations of L and M cones alone.

The color matches are *asymmetric* matches; had they been symmetric matches (threshold level, silent for S cones) they would in at least many cases been matches at the level of the cones, rather than postreceptoral metameric matches, and thus not informative about color appearance. Most importantly, the fact that S cones may contribute to the matching stimuli but not the test stimuli to which they are matched illustrates the principle that the hues are associated with the postreceptoral mechanisms,
not with particular cone types. A threshold response in a single, univariant labelled-line mechanism produces the same hue, regardless of whether that response is based upon signals in the L, M, or S cones.

It is puzzling why the tests near the corners of the detection contour, where the underlying mechanisms have the same sensitivity, do not produce color matches that are mixtures of the matches generated by the two mechanisms when isolated. For example, a test at 25° that is detected by both the R and Y mechanisms (as seen in Figure 6b) was only matched with a reddish hue (never a yellowish hue). Another example is seen in Figure 6f; a test at 64° is detected by the Y and G mechanisms but observer SAF only matched this threshold-level stimulus with a greenish hue. It is likely that the lack of mixtures is due to a combination of (a) the test not being exactly in the corner, and (b) the fact that the matches are memory matches, and possibly dominated therefore by one mechanism hue. In Chapter 5, I discuss a discrimination task that has little memory demand, and is also forced-choice with feedback, so that discriminability is driven towards the best possible performance rather than being dominated by one or the other mechanism as in the color matching task presented here.

**General Discussion and Conclusions**

We presented a chromatic detection model consisting of six linear mechanisms (Shepard et al 2016). We showed that six mechanisms are sufficient to account for selective masking when chromatic noise is placed near the corners of detection contours in the L,M plane. Here, in a separate experiment, a color-matching task provided insight into the subjective experience resulting from these mechanisms and also allowed us to
test our six-mechanism model. The matches correspond extremely well with the mechanism lines, except for three minor discrepancies at the ends of some detection contours, which allow us to formally apply a color label to each univariant, labelled-line mechanism.

The color matching experiment also helps rule out some alternative models. As discussed in Shepard et al (2016), a model containing only four mechanisms that have adaptive changes in cone weights in the presence of high contrast noises (Atick, Li, & Redlich, 1993; Zaidi & Shapiro, 1993) could fit the entire set of thresholds, simply by altering the cone contrast weights of the R and G mechanisms in each noise condition to align the long flanks of the detection contour approximately with the noise vector. Two other mechanisms are required (e.g. the Y and P mechanisms) to adequately fit the data (i.e. four mechanisms total). This model is nearly impossible to test using thresholds alone, without theoretical constraints on the adaptive changes in cone weights. However, the color matching data presented here indicate that the hues of the thresholds fall into six, not four, categories. An adaptive model with only four mechanisms could not easily account for the color matches presented here.

There is no doubt that a model with more than six mechanisms can fit the detection thresholds as well as the six mechanism model, but as discussed in Shepard et al (2016), these extra mechanisms do not improve the fit to the thresholds or improve the modeling of selective masking effects. Moreover, quantifying the hues associated with each mechanism, as done here, clearly shows that if there were additional mechanisms they would have to produce redundant hues. The color matches fall into six categories, not more (or less) than six.
None of the thresholds looked achromatic, for any of our observers in any of our conditions, and it therefore seems likely that at least one additional pair of mechanisms will eventually be required to account for modulations in three-dimensional cone space. Also, since in these experiments the S cones were not modulated, it is somewhat uncertain as to which of the present six mechanisms might get S cone input; the S cone input shown in Figure 1 is speculative. These are issues still to be settled, and it is of course possible, even likely, that more than six mechanisms will ultimately be required. However, the most parsimonious model to account for both detection and threshold-level color matches across a broad range of conditions in the L,M plane has six, and only six, mechanisms.
References


Figure 1: Six-mechanism model (Shepard et al., 2016). Each mechanism is half-wave rectified. Four of these mechanisms (R and G, and B and O) are ‘quasi-paired’, having nearly equal and opposite L and M cone weights. Two mechanisms (Y and P) have additive L and M cone inputs, and are asymmetric (unpaired). The S cone input to B and O is by analogy to the cardinal model; it was not studied in these experiments.
Figure 2: The lines are threshold responses of one (a) or two (b) detection mechanisms in the LM plane. The interpretation of the color matches is based on two properties of a mechanism. a) Univariance implies that physically different tests that are detected by a single mechanism should produce metameric color matches. For example, a M-cone increment test (greenish) may produce the same color match as a L-cone decrement test (greenish) because both test thresholds are detected by the same univariant mechanism. b) Labelled lines imply when tests lie on two different mechanism lines, they should be matched with two different hues. A 135° test (greenish) should produce a different color match than a 225° test (purplish) since they are detected by two separate mechanisms. The background is an exaggerated representation of the color of test at each point.
Figure 3: The test stimuli were circular Gaussian blobs ($\sigma=1^\circ$), the same test stimuli that were used to measure thresholds in Shepard et al, (2016).
Figure 4: HSV color wheel (left). Confirmatory match (right) was a colored disc of approximately the same size as the test stimulus displayed on a grey background that had the same chromaticity and approximately the same luminance as the experimental monitor.
Figure 5: Mean color matches pooled across all noise conditions (one mean per test angle per noise condition), along with the white point (white disc) of the experimental monitor, plotted in u’,v’ coordinates. The ellipses represent the 95% confidence region around the centroid, computed from principal components analysis. Example using observer TGS. For each observer, the ellipses determined from the pooled data are repeated in the plots for each noise condition (Figures 6-9, left panels).
Figure 6: No Noise condition. Left panels: Mean color matches for each test, along with the white point (white disc) of the experimental monitor, plotted in u’,v’ coordinates. The ellipses represent the 95% confidence region from the matches pooled across noise conditions (compare panel (a) with Figure 5). Right panels: Detection thresholds (colored discs) and model fits for observers TGS, CLM and SAF in cone contrast space. The color of each disc is an approximate representation of the color match for each test angle made in the color matching experiment (left panel). Colored lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms.
Figure 7: 42° Noise condition. Format as in Figure 6. Axis scales in the (u',v') plots (left panels) are the same as in Figure 6, but scales in the detection plots (right panels) are expanded due to the masking effects of the noise.
Figure 8: 48° Noise condition. Format as in Figure 6. Axis scales in the \((u',v')\) plots (left panels) are the same as in Figure 6, but scales in the detection plots (right panels) are expanded due to the masking effects of the noise. SAF did not participate in this noise condition.
Figure 9: 64° Noise condition. Format as in Figure 6. Axis scales in the (u’,v’) plots (left panels) are the same as in Figure 6, but scales in the detection plots (right panels) are expanded due to the masking effects of the noise.
Chapter 5: A Bayesian Classifier Model of Chromatic Discrimination in the (L,M) Plane
Abstract

Some models of chromatic detection make specific predictions about the discriminability of stimuli that are presented at detection threshold (Eskew, Newton, & Giulianini, 2001; Shepard et al., 2016). In this study, on a given trial, two threshold-level stimuli (one "standard" and one “test”) were presented in random order, and observers were asked to select the standard. Across trials, many different standards and tests in the (L,M) plane of cone contrast space were used, with and without chromatic masking noise (Shepard et al., 2016). Qualitatively, (i) if both stimuli are detected by the same chromatic mechanism, performance should be at chance level (50%); (ii) if the stimuli are detected by two different chromatic mechanisms, discrimination should be as good as detection (ca. 82%); (iii) if one or both stimuli are detected by multiple mechanisms, performance should be intermediate. Quantitatively, a Bayesian Color Classifier (R. Eskew et al., 2001) was used to combine the outputs of the six mechanisms in the detection model from Shepard et al., (2016); the classifier’s predictions are computed without any free parameters. In general, the classifier’s predictions for each observer and each condition are excellent, confirming the detection model.
Introduction

Recently we described a chromatic detection model that contains six mechanisms (Shepard et al., 2016a), but, unlike the cardinal model that also has six mechanisms (Krauskopf, Williams, & Heeley, 1982), accounts for selective masking in the (L,M) plane of color space (Hansen & Gegenfurtner, 2006). This model differs in one important way from the cardinal model: four of the six detection mechanisms have opposed L and M cone inputs, rather than just two as in the cardinal model. This is the critical feature that allows the model to account for selective masking when noise is placed near the ends of the detection contour in the (L,M) plane; noises that are nearly parallel to the long flanks of detection contours can cause different mechanisms to become most sensitive and thus determine threshold, tilting the overall detection contour (see figure 9 in Shepard et al., 2016a).

An outline of the detection model is shown in Figure 1. The top and bottom pairs of mechanisms take weighted differences of the L and M cone signals, with different weights for all four mechanisms (within each row, the weights are similar in magnitude but of opposite signs). The middle pair of mechanisms takes the sum of the L and M cone signals, again with weights that are similar in magnitude but of opposite sign for the two mechanisms in the row. An S cone input of opposite sign is shown to this mechanism; this input is speculative because the model has not yet been tested with S cone modulations. This model successfully predicts selective masking effects on 2AFC detection in the (L,M) plane.
The mechanisms in the model are assumed to be univariant labelled lines (Rushton, 1972; Watson & Robson, 1981), which implies that the hues of physically-different stimuli that are detected by a single mechanism should all be the same, and that the hues of two stimuli detected by different mechanisms should differ (with the additional assumption that the hues of the mechanisms are not redundant). In a later study, observers made asymmetric color matching of the threshold-level tests from Shepard et al. (2016a). Consistent with the model, the results showed that color matches fell into six clusters in CIE (u’, v’) space (across all the noise conditions), and these clusters correspond closely to the six mechanisms in the model. The capital letter names for the mechanisms in Figure 1 are mnemonics for the hues of the six colors generated by the mechanisms: (R)ed, (G)reen, (B)lue, (Y)ellow, (O)range, and (P)urple.

Although the separation of the color match clusters in the previous study suggests that stimuli detected by different mechanisms are likely to be discriminable at threshold, this implication was not tested. In the present study, using the same univariance and labelled-line assumptions, we further tested the six-mechanism model, with threshold level discrimination.

One issue in the color matching study was that, even for stimuli that apparently lie close to the corners of the detection contour where two mechanisms have equal sensitivity, the matches never were intermediate between the color clusters corresponding to the mechanisms. In other words, the matches were never ‘mixtures’ of the mechanism hues. As discussed in Chapter 4, the lack of mixtures is most likely due to a combination of (a) not being exactly in the corner, and (b) the fact that the matches are memory matches, and possibly dominated by one mechanism hue. The present study allows a test
for this idea, as described below, because the discrimination procedure does not have the memory component of the color matches.

The relationship between the detection thresholds and discrimination data are modeled using a Bayesian classifier, which was developed in (Eskew et al., 2001). This classifier takes, as its inputs, the outputs of the mechanisms that were fitted to the detection data (Shepard et al., 2016), and predicts the discriminability of any pair of threshold-level tests, under the conditions of chromatic masking noise used in the detection study. The model-fitting is only done on the detection thresholds, prior to the collection of the discrimination data; the Bayesian classifier has no free parameters.

**Methods**

**Observers**

Observers TGS and SAF, who participated in the detection and color matching experiment (Shepard et al. 2016) also participated in the discrimination experiment presented here. The observers were well practiced at the start of the discrimination study, several months after the detection study had been completed, but some additional practice with the discrimination task was given in conditions where discrimination was difficult. Both observers had normal scores on the Farnsworth-Munsell 100 hue test (Farnsworth, 1943) and the Ishihara Plates. Northeastern University’s Institutional Review Board approved the research protocol; the procedures comply with the Declaration of Helsinki.

**Apparatus**
Stimuli were created on a Power Macintosh computer and displayed on a SONY GDM-F520 CRT monitor running at 75hz, using an ATI Radeon 7500 video board with a driver verified to support the 10-bit digital-to-analog converters. Spectroradiometric calibration was performed at 4 nm intervals across the spectrum using a Photo Research PR-650 spectroradiometer, and the monitors were linearized with gamma correction lookup tables. Observers were corrected to normal visual acuity using a trial lens that was placed in front of their dominant eye; the other eye was patched. Head position was stabilized with a chin and forehead rest. All experiments were conducted in a dark room.

**Test and Noise Stimuli**

Test and noise stimuli were identical to those used by Shepard et al (2016). The tests were circular Gaussian blobs (σ=1°), presented with a rapid-start ‘sawtooth’ profile of 333ms total duration. The masking noise consisted of horizontal lines that were superimposed on the test (Shepard et al., 2016, figures 2 and 3). The lines randomly and independently changed from one chromaticity to a symmetrically opposite chromaticity (on the opposite side of the white point).

Stimuli are specified by their polar angle in the L,M plane of cone contrast space, with 0° and 90° representing L and M cone increments respectively. The bipolar noise has two chromatic angles separated by 180°, one for each pole; noise direction will be identified here by the first pole (e.g. the noise direction 42°/222° will be simply be called 42° noise). The noise conditions used were three of the four conditions used in Shepard et al., (2016a & 2016b); no-noise, 42°, and 64° noises. The noise contrast was always 90% of the maximum available at the noise direction; in cone contrast units, the noise
cone contrast vector length was 0.498 and 0.267 for the 42° and 64° noises, respectively. Further details of the stimuli may be found in Shepard et al, 2016a).

Procedure

In each trial, the observer was presented with two stimuli (a “standard” color angle along with a test color angle) in random order. Both of the stimuli were fixed at their individual 2AFC detection thresholds (Shepard et al., 2016). Observers were asked to select the interval in which the standard was presented. The observer learned which stimulus was the standard on the first 10 trials of each run; results from these trials were discarded. Each run consisted of 10 practice and 50 test trials; auditory feedback was provided on every trial (i.e. the higher pitch tone indicated a correct discrimination). Runs were blocked by standard chromaticity; test chromaticities were selected in random order within a block of runs. Percent correct discrimination was generally estimated from two to three pooled runs, collected on separate days.

In a few cases the test or standard was selected, not from actual thresholds, but from the model detection contour (the solid blue contour in figures 4-7 in (Shepard et al., 2016) – in other words, rather than using the measured threshold contrasts, in these cases we used predicted threshold contrasts. This was done in cases where a particular standard or test chromaticity of interest in the discrimination study was not used in the detection study, and so no measured threshold was available. These points are plotted as solid black squares in the figures.

Detection and Discrimination Models

The six-mechanism detection model was fit to the two-alternative forced-choice detection thresholds in the prior paper (Chapter 3), across all the noise conditions.
simultaneously. Figure 2a, for observer TGS in the no-noise condition (Shepard et al., 2016, figure 4); the discs represent the cone-contrast vector length thresholds. The colored lines represent the six mechanism threshold lines (mechanisms R, G, Y, B, O, P). The colors of the threshold points are based upon the results of (Shepard et al., 2016b; Chapter 4); they are exaggerated representations of the color matches made to these threshold-level tests, in a separate experiment. The mechanism lines are colored to be a more saturated representation of the matched colors of the tests detected by the mechanism that, according to the detection model, is responsible for detecting that test angle.

Figure 2b & c give hypothetical discrimination data represented in both a Cartesian and a polar plot, to illustrate the format of the subsequent figures. The black disc denotes the standard, at 225°, and the colored discs represent tests. In the Cartesian plot (Figure 2c), the horizontal axis denotes the color angle and the vertical axis represents the performance (i.e., % correct).

The polar plot (Figure 2b) has an angular coordinate representing the polar angle of the test color direction, the same angles as the polar angles of the stimuli in Figure 2a. Whereas in Figure 2a the distance of a colored disc from the origin represents detection threshold (cone contrast vector length), in Figure 2b the distance of a colored disc from the origin represents its discriminability from the standard color direction (225° in this case). In Figures 2b and 2c, the black line (six-mechanism model) indicates the performance predicted by the classifier for each possible test against the 225° standard, and the red dashed line indicates chance-level discrimination performance (50%).
In order to interpret the discrimination results, consider what would happen if two equally-detectable stimuli are compared in this discrimination procedure. If these stimuli are detected by a single univariant mechanism, they will be indiscriminable from one another at threshold (50% performance). The purple colored disc (i.e., 222°) in Figure 2b & 2c is an example of such a test—the predicted performance for this test is ~50%. This happens when two stimuli are detected by one and only one mechanism (i.e., (P) mechanism) even though they are physically different.

When two stimuli are detected by two different labelled-line mechanisms, they should be approximately as discriminable as they are detectable (Watson & Robson, 1981). An example is shown in Figure 2. The standard at 225° is detected by the P mechanism and the test (i.e., yellow disc) 180° away from this stimulus at 45° is detected by the Y mechanism—discrimination performance should be around 82%. In another example, a 315° test (i.e., red disc) that is detected by R mechanism (adjacent to the P mechanism) should be highly discriminable from the standard since they are detected by two different mechanisms.

If two stimuli are simultaneously detected by multiple mechanisms, near a corner of the detection contour, they will be imperfectly discriminable, because on some trials one mechanism will detect the test and another mechanism will detect the standard, allowing good discrimination on those trials. On other trials, the same mechanism will detect both the test and the standard, resulting in chance performance on those trials. For example, on some trials the test at 215° (purple/green disc) will be detected by the G mechanism and on some trials the same test will be detected by both the G and the P mechanisms. Interestingly, in the color matching experiment (see Chapter 4), found that
matches in these intermediate regions were not mixtures of mechanism hues; instead, they were consistently matched with one hue (i.e. purple).

The majority of the standards were selected to be near the corners of the detection contour where performance is likely influenced, not only by the sensitive flanking mechanisms, but also by other mechanisms (these are the test angles where we might have expected to get ‘mixture’ color matches; Shepard 2016; Chapter 4) Fewer tests were selected along the long flanking (e.g. R and G) mechanisms when the standard also fell along one of those mechanisms; the indiscriminability of lights that are detected by R or G has been shown in previous studies (Eskew et al., 2001);(Newton & Eskew, 2003) Also, few tests were selected near 180° away from the standard, since good discrimination in these cases is also well established (Eskew et al., 2001; Newton & Eskew, 2003; Gowdy, Stromeyer, & Kronauer, 1999); this good discrimination between ‘opponent pairs’ is a major reason that the model has unipolar mechanisms (see (R Eskew, 2009; Eskew, 2008, for discussion).

The discrimination data were quantitatively analyzed using a Bayesian classifier model (Eskew et al, 2001, Appendix and (Newton & Eskew, 2003). In overview: first, the detection model was taken from Shepard et al., (2016) (without adjustment of any kind). The Bayesian Classifier takes the model mechanism outputs, for all mechanisms, in response to the presented stimuli, and gives the probability of various response patterns, across all mechanisms, that are created by a pair of stimuli. The classifier then decides which stimulus was presented by picking the stimulus that was most likely to create the pattern of responses in the mechanisms.
In detail: The detection model is used to calculate the responses of all of the mechanisms to either the test (I) or the standard stimulus (II). That set of responses is denoted by the vector $\vec{\alpha}$. The likelihood that the classifier will categorize the stimulus as the standard is based upon Bayes’ rule (with equal priors, since the test and standard are presented equally often):

$$P(I \mid \vec{\alpha}) = \frac{P(\vec{\alpha} \mid I)}{P(\vec{\alpha} \mid I) + P(\vec{\alpha} \mid II)}$$

Each of the terms on the right-hand side is a joint psychometric function, the probability of a set of detection responses given stimulus ‘test’ I (colored discs) or ‘standard’ stimulus II (black disc) in Figure 2a-c. Again, note that this model was not fit to the discrimination data. Instead, we used the same model that had been fit to the detection data (Shepard et al., 2016a), to make a prediction of the performance in a completely different task, without any parameter estimation or adjustment. For comparison, other detection models (described later) were also used to make predictions.

The black solid lines in the discrimination plots show the performance of a Bayesian classifier. As discussed in Eskew et al., (2001), the inputs to the model are the test color angles and vector lengths which represent the test and standard stimuli. The model classifier chooses between test (‘I’) and the standard (‘II’) stimuli based upon the posterior probability:

“$I$” \quad $P(I \mid \vec{\Omega}_i) \geq P(II \mid \vec{\Omega}_i)$

“$II$” \quad otherwise
\( \Omega \) represents the set of binary decisions made by each mechanism during a single stimulus presentation. The classifier receives that set of mechanism detections \( \Omega \), and makes the ideal decision based upon those mechanism outputs by choosing the stimulus with the greatest probability.

Watson & Robson, (1981) claim that stimuli detected by two different labelled line mechanisms are discriminated as well as they are detected. However, in detail the level of discriminability depends upon the tasks in question. In our detection experiment, there was one stimulus per trial, presented in either the first or second interval. In the forced-choice discrimination task, there were two presentations in each trial—one in each interval. In the discrimination experiment, each stimulus can be used to make a correct discrimination. After correcting for the difference between a single presentation in the detection experiment and the double presentation in the (present) discrimination experiment, the best performance of our model in discriminating two stimuli that are exactly at 82% detection threshold is 86%. These and other details of the classifier model are described in the appendix of Eskew et al. (2001).

It is important to remember that the mechanism’s cone weights and noise sensitivity parameters were taken directly from the detection model (see Table 1 in Shepard et al., (2016a) and were used to predict the detection responses of each of the six mechanisms to the test and the standard. These responses were then combined, according to the classifier (see Eskew et al., 2001) and used to generate the discrimination performance predictions. This means that there were no free parameters in these model predictions as they are based on the detection model fits from a completely different experiment.
Results and Discussion

Observer No-Noise Discrimination

As described in the Methods, both the test and the standard were fixed at threshold contrast (82% 2AFC detection) as determined in a prior experiment (Shepard et al., 2016a). Figure 3 shows TGS’s model fit in the no-noise condition from that experiment. The threshold points are replotted from (Shepard et al., 2016; Chapter 4); the color of each disc is an approximate representation of the color match for each test angle made in the color matching experiment. Colored lines on the plots represent mechanism thresholds and the smooth closed contour is the probability sum of those mechanisms. The color of the line is a representation of the hue signaled by each mechanism (Shepard et al., 2016). Four mechanisms contribute to detection (R, G, Y, and P) in this condition and the model fits the data extremely well.

The data points in Figure 4a show TGS’s ability to discriminate threshold-level tests, in this no-noise condition, from a 0° standard stimulus (refer to Fig. 3). The figure shows that TGS could not discriminate between the standard and the 315° test (~50% correct). As seen in the model fit (Figure 3) the standard and test stimuli are detected by the same mechanism (R), so discrimination is at chance level, as required for a univariant mechanism. TGS was able to discriminate between the 0° standard and the 180° test since the standard is detected by the R mechanism and the test (at 180°) is detected by the G mechanism, as implied for two labelled lines; this result was also shown by Eskew et al., (2001). In fact, the best discrimination performance (ca. 82%) is observed between the 0° standard and any test between the range 45°-225°, since these tests and the standard are detected by different mechanisms. Discrimination between the 25° test and the 0°
standard is intermediate (~72% correct), since, according to TGS’s detection model, the 25° test is near the corner of the detection contour and detected by both the R and Y mechanisms. As previously discussed, on some trials the R mechanism detects the 25° test, on other trials the Y mechanism detects the 25° test. On those trials in which both the test and standard are detected by the same mechanism (i.e. R mechanism), the observer cannot discriminate them (chance level performance), but on the other trials the observer will be able to do so, resulting in an intermediate level of performance. This good correspondence between the corner angles of the detection contour and these imperfectly discriminable test angles is due to the exceptional fit of the detection model to the thresholds found in Shepard et al., (2016). Notably, the color of the 25° test, as measured by color matching and shown by the plotted point’s color, is reddish, not intermediate between red and yellow; this difference between color matching and discrimination will be discussed later.

The continuous black, dotted blue, and dotted magenta lines in the discrimination panels represent the Bayesian Classifier applied to three different detection models. The black line is for the six-mechanism model from Shepard et al., (2016), which is described in the Introduction and shown in Figure 1. The other two lines are for two alternative models that will be described in a later section.

TGS’s discrimination against the 45° standard is shown in Fig. 4b. The 42°, 45°, 48°, 52°, and 58° tests are indiscriminable (ca. 50% performance) from the 45° standard which corresponds well with the detection model—all of these stimuli are detected by the Y mechanism. These same five test angles were matched with a yellowish stimulus in Shepard et al., (2016). Tests between 90°-315°, and the 0° test, were perfectly
discriminable from the 45° standard; according to the model these stimuli should always be highly discriminable since they are detected by different mechanisms than the one detecting the standard. Tests at 15° and 25° (reddish matches) were imperfectly discriminable since they both lie near the upper right corner of the detection contour where multiple mechanisms exist (i.e., R and Y mechanisms).

Discrimination against a 315° standard is shown in Figure 4c. The measured discriminability of the 315° standard against itself is higher than chance, 55% correct; this is within the expected range due to binomial variability (the expected standard error is 5%). The detection model shows that the test that lie between 270°-0° are only detected by the R mechanism. Since they are detected by the same mechanism, discriminability between these tests and the standard should be at chance level, as they are (~50%). Each test selected from the range of 45°-225° is highly discriminable from the 315° standard, since they are detected by different mechanisms. Tests at 15° and 25° (i.e., reddish color matches) are imperfectly discriminated from the 315° standard (65%, and 73% correct, respectively) which correspond well with the location of these test in the corner of the detection contour in the model fit (Shepard et al., 2016).

SAF’s discrimination performance in the no-noise condition also corresponds well with her detection model (Figure 5). For example, SAF’s discrimination against the 45° standard is shown in Figure 6a. According to the model, the 45° standard is detected by the Y mechanism. Tests at 39°, 40°, 41°, 42°, 45°, and 55° are also detected by the Y mechanism which resulted in poor discrimination (47%-57% correct). Tests between 135°-315° are detected by other mechanisms so discrimination is ~80%. The transitional stimuli (64° and 65°) also correspond well to a corner of the detection contour, where on
some trials the Y mechanism detects these tests and on other trials the G mechanism
detects these tests.

Figure 6b and 6c show SAF’s ability to discriminate threshold-level tests from a
135° and a 225° standard, respectively. In both conditions, when the test and standard
were detected by the same mechanism according to the model fit, discrimination was
poor. When the test and standard are not detected by the same mechanism, (e.g. a 135°
standard and a 315° test; Figure 6b) discrimination was good, at ~82%. And last, when
the test is in the corner of the detection contour and shares a common mechanism with
the standard, the stimuli were imperfectly discriminable. All of these findings are
qualitatively consistent with the detection model.

The black line in Figure 4a is the predicted performance when we use the
parameters from the six-mechanism model (Shepard et al., 2016). The model predicts that
discriminability against the 0° standard should gradually rise as the test angle becomes
larger and moves away from the standard, and reaches an asymptote near 70°. The
observed data follow this pattern quite well. Another example is the prediction made in
Figure 4b, with a 45° standard. At 45°-58° the prediction falls along the red chance level
line and progressively rises as the test angle becomes larger to an asymptote at ~64°. The
prediction for discrimination is ~85% for the all tests between 64°-0° where the
prediction begins to fall back into the trough of the contour at 45°. Again, the pattern of
observed discrimination is very similar to the predictions made by the classifier.

Overall, the Bayesian Classifier does an excellent job of capturing the pattern of
discrimination for all of these standards. The model predicts the test angles around the
standard, which are indiscriminable from it, the transition through the tests with
intermediate discrimination up towards good discrimination, and the set of test angles that are well-discriminated from the standard. This is true for each standard and for both observers.

*Discrimination with Added Noise*

The noise conditions were identical to those used to collect the detection thresholds in Shepard et al., (2016a) and the color matches to those same thresholds in Shepard et al., (2016b). Discrimination was tested in two of the noise conditions for TGS (42° and 64°) and in one condition for observer SAF (64°). As shown below, both observers’ discrimination performance, with the added noise, corresponds well with the model fits that were generated in the previous experiment.

Figure 8 shows discrimination against a 45° standard in the 42° noise condition. In the detection model, four mechanisms contribute to detection (Figure 7), two along the flanks, and two at the ends of the detection contour. A 45° standard is detected by the O mechanism. Discrimination is at chance level for tests that are close to the standard (41°, 42°, and 48°)—since these stimuli are all detected by the O mechanism. Tests that are detected by the G, P, and R mechanisms (according to the detection model) are highly discriminable (70°-15°) from this 45° standard.

The Bayesian Classifier based upon the six-mechanism model (solid black lines) again does an excellent job in predicting the discrimination performance, without any free parameters. The predictions in Figure 8 is computed from the same detection model as used in Figure 4; the only difference is that the noise angle and noise contrast have been input to the detection model, which changes the sensitivities of the mechanisms. With no free parameters, the discrimination predictions made by the Bayesian Classifier
in the added-noise conditions now reflects the outputs of these masked mechanisms as its inputs. The classifier again does an outstanding job of predicting TGS’s discrimination performance (black line) across many tests in the 42° noise condition. The other lines in the figure are predictions made with the inputs of two alternative models (discussed below). Performance of the human observer falls somewhat below the highest performance of the ideal classifier; see discussion below and in (Eskew et al., 2001)

Figures 10 and 12 depict discrimination data for observers SAF and TGS in the 64° noise condition. Similar to the rest of the discrimination data that have already been presented, discrimination can generally be predicted by inspecting the detection model fits (Figures 9 and 11) and implementing the principles of a univariant-labelled line mechanism. Figure 10a-c, show discrimination against three different standards (0°, 48°, 90°) in the presence of 64° noise for observer TGS. For all three standards, discrimination is good when the standard and tests are detected by separate mechanisms (e.g. the 180° test in Figure 10a) and poor when the test is detected by the same mechanism (e.g. the 15° test in Figure 10a). Imperfect discrimination occurs when a test lies in the corner of the detection contour and shares a common mechanism that on some trials detects both the test and the standard (e.g. where the R and Y mechanism meet)—performance here is 61%-75% for tests between 35°-45°. Figure 10b and 10c show a similar pattern of results.

Figure 12a shows SAF’s discrimination results of a fairly small set of tests against a 315° standard. When the standard and the test are the same, performance is at 50%. When the test is detected by a separate mechanism than the 315° standard, performance is high (ca. 82%). When a test is detected, on some trials, by the same mechanism as the
standard (e.g. 0° and the 270° test), performance is intermediate. Discrimination performance in the current experiment strongly support SAF’s detection model fit to the detection thresholds, which were collected almost a year prior to this study.

Considering all the noise conditions and both observers, the classifier does an excellent job of predicting discriminability, as it has in prior applications (Eskew et al., 2001; Newton & Eskew, 2003), despite the lack of any model-fitting. For cases when discrimination is good, the classifier generally does slightly better than the human observer, as also found before. The classifier is not an ideal observer – it is an ideal classifier of sub-optimal representations of the cone signals, the mechanism responses (see discussion in Eskew et al., 2001). The fact that the human observers come close to the performance of the classifier suggests that, for the detection and discrimination tasks, the noise that limits performance is early, and that the decisions based upon the suboptimal chromatic detection mechanisms come fairly close to being optimal.

**Alternative models**

A major aim of Shepard et al., (2016a) was to determine whether a “higher order” model— with more than six mechanisms— was required to produce selective masking and account for the data along the detection contours. Therefore, the Bayesian Classifier prediction was recalculated for two alternative detection models, both of which contained eight mechanisms, that were first fit to the detection data (like the six-mechanism model); then those parameters were used as inputs to the Bayesian Classifier.

We began with a greatly simplified version of the Hansen & Gegenfurtner (2006) model. This first model contains eight equally spaced, equally sensitive mechanisms in the (L,M) plane of MBDKL space. The set of angles in this model includes the four
cardinal mechanism directions along with four intermediates (i.e. 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°). When fitting the measured thresholds, for half of the mechanisms – those in the right half plane of MBDKL space – the vector length |\(f|\) (Eq. 1 in appendix Shepard et al., 2016) were allowed to vary, as was the constant of proportionality \(b\). The other half of the mechanisms formed symmetric pairs with the first half. This model fit the data worse than the six-mechanism model where sensitivity and the L,M cone weight ratios were free to vary across observers (see Supplementary in Shepard et al., 2016a) for discussion of a similar model with 16 mechanisms).

The other model used here also has eight mechanisms, but they obey the same Energy vs. Noise relationship as the six-mechanism model. In other words, this model simply adds two mechanisms to our six-mechanism model, with the angles and sensitivities of those added mechanisms chosen for best fit to the detection data. The fits were not significantly better than the six-mechanism model—adding mechanisms was unnecessary to predict detection (see Supplementary material in Shepard et al., 2016a). The parameters from both of the alternative models were fed into the Bayesian classifier to generate discrimination predictions for both models.

The predictions are shown as the dashed blue (best 8 mechanism model) and dashed magenta (8 equally spaced mechanism model) lines in Figures 4, 6, 8, 10, and 12. In general, the classifier predictions based upon these alternative models do not differ substantially from those based upon our six-mechanism model. There are, however, two instances where the models do differ appreciably. In Figure 8a, the eight mechanism model (dashed blue line) performs worse than both the six and the eight equally-spaced in the range of 45°-225°. The predictions are different because the 45° standard is detected
by two mechanisms in the eight mechanism detection model fit (not shown). In this case, all of the tests along the upper flank are detected by the G mechanism but the standard is detected by both the Y and G mechanism, which leads to intermediate performance.

Another case is seen in figure 10b. The classifier’s performance with the parameters from the eight equally-spaced model (dashed magenta line), is much different from the other two predictions in the range of 270°-45°. This discrepancy is again due to location of the standard in the corner of the detection contour in the eight equally-spaced model fit. Ongoing experiments will add tests in these regions to further test the alternative models against the six-mechanism model. Overall, the alternative model predictions are no better than those made with the parameters from the six-mechanism model fit.

General Discussion

We presented a chromatic detection model consisting of six linear mechanisms (Shepard et al 2016). We showed that six univariant labelled line mechanisms are sufficient to account for selective masking when chromatic noise is placed near the corners of detection contours in the (L, M) plane. Here, in a separate experiment, a discrimination task with the threshold stimuli in the detection experiment allowed us to test our six-mechanism model. We consider mechanisms as univariant labelled lines and generally saw three types of behavior which were indicated by the observers’ discrimination performance: 1) when both stimuli were detected by the same chromatic mechanism, performance was at chance level (50%); 2) when the stimuli were detected by two different chromatic mechanisms, discrimination was as good as detection (ca. 82%); 3) when one or both stimuli were detected by multiple mechanisms, performance
was intermediate. Based on our definition of a mechanism, we could qualitatively predict discrimination performance for a variety of thresholds level tests against multiple standards and confirm the detection model.

We also used a Bayesian Classifier model (Eskew et al., 2001) to make quantitative predictions of performance. The classifier was used to optimally combine the outputs of the six mechanisms in the detection model from Shepard et al., (2016)—the classifier’s predictions are computed without any free parameters. The classifier does a very good job of predicting the pattern of discriminability for multiple standards in multiple noise conditions.

Performance in the discrimination task is bounded by 50% and 100%. Neither the human nor the Bayesian Classifier can do worse than chance (except for experimental error in the case of the human). However, as discussed in Eskew et al., (2001), the classifier performs optimally given the suboptimal inputs of the chromatic detection mechanisms, and thus at high levels of discriminability, the human’s absolute level of discrimination performance is likely to be worse than that of the Classifier (86% in this task; see Methods). The data in the present study generally conform to this expectation (as it did in Eskew et al., 2001 and Newton & Eskew, 2003), with the highest performance of the observers averaging closer to 82% than 86% (Figures 4, 6, 8, and 10). As noted above, the relatively high efficiency of the human in this task suggests that most of the limiting noise in this discrimination task is early, contributing to the detection but not to discrimination.

In the color matching study of Shepard et al., (2016; Chapter 4), a puzzle was why the tests near the corners of the detection contour, where the underlying mechanisms have
the same sensitivity, do not produce color matches that are mixtures of the matches generated by the two mechanisms when isolated. The hypothesis offered by Shepard et al., 2016b; Chapter 4), was that the lack of mixtures is most likely due to a combination of (a) not being exactly in the corner, and (b) the fact that the matches are memory matches, and therefore possibly dominated by one mechanism hue. If that hypothesis is correct, then a task that did not have the memory component of the matching task might indeed show intermediate discrimination – a performance analog of the mixture idea. The present results are consistent with this idea.

For example, Figure 4a shows TGS’s ability to discriminate threshold-level tests, in no-noise condition, from a 0° standard stimulus. The R mechanism that detects the 0° standard (as seen in Figure 3) signaled a reddish hue in the color matching experiment. Strangely, a test at 25° that is detected by both the R and Y mechanism (as seen in Figure 3) was only matched with a reddish hue (never a yellowish hue). However, discrimination performance for this test against the standard is intermediate (75% correct; Figure 4) which suggests that in fact two mechanisms must detect this test—on some trials, the test and standard are detected by the R mechanism (resulting in chance performance) and on other trials, the R mechanism detects the standard and the Y mechanism detects the test (good performance).

Another example is SAF’s ability to discriminate threshold-level tests, in no-noise condition, from a 45° standard stimulus (Figure 6a). The Y mechanism that detects the 45° standard (as seen in Figure 5) signaled a yellowish hue in the color matching experiment. A test at 64° is detected by the Y and G mechanism (as seen in Figure 5) but observer SAF only matched this threshold-level stimulus with a greenish hue in the
matching experiment. Discrimination of the 64° test against the standard was at 68% (Figure 6a). This transitional performance is likely due to the location of the 64° test in the corner of the detection contour of the six-mechanism model. On some trials the 64° test and 45° standard is detected by the Y mechanism (50% correct) and on other trials the G mechanism detected the 64° test and the Y mechanism consistently detected the 45° standard. If the 64° test was always detected by the Y mechanism, performance should be at chance level and we would expect to see a yellow match. If the test was only detected by the G mechanism, performance should be ~82% and signal a green match. However, performance was intermediate and matched with a greenish hue—not yellowish and greenish—suggesting that color memory matches were perhaps dominated by one mechanism hue. The discrimination task has little memory demand, and is also forced-choice with feedback, so that discriminability is driven towards the best possible performance rather than being dominated by one or the other mechanism as in the color matching task.

Taken together with the results of the color matching experiment, we provide further support for the six-mechanism model (Shepard et al., 2016). Although in general there are only a few minor differences in the Bayesian Classifier’s performance based upon the alternative detection models, where there are differences (Figure 8a from 45°-225°, Figure 10b from 270°-45°), the six-mechanism model produces discrimination performance that is more like the human observer, providing even more evidence for the six-mechanism model.
References


Figure 1: Six-mechanism model (Shepard et al., 2016). Each mechanism is half-wave rectified. Four of these mechanisms (R and G, and B and O) are ‘quasi-paired’, having nearly equal and opposite L and M cone weights. Two mechanisms (Y and P) have additive L and M cone inputs, and are asymmetric (unpaired). The S cone input to B and O is by analogy to the cardinal model; it was not studied in these experiments.
Figure 2: a) Model fit for TGS in No-Noise condition (Shepard et al., 2016). b) Hypothetical discrimination data plotted in polar coordinates (left) and Cartesian coordinates (right). The black disc indicates the ‘standard’ and the colored discs denote the test stimuli. The dashed red line indicates chance level performance. If two stimuli are detected by a single mechanism, they will be indiscriminable (50% performance) from one another at threshold (purple disc). If a test is simultaneously detected by multiple mechanisms, they will be imperfectly discriminable, because on some trials only one of the mechanisms will detect these stimuli, allowing discrimination on those trials and, on other trials, both mechanisms will detect the test which results in chance performance (green/purple disc). If the standard and the test are detected by different mechanisms, they will be as discriminable as they are detectable at ~82% (red discs).
Figure 3: No-Noise condition. Detection thresholds (colored discs) and model fit for observer TGS in Cone Contrast space. Colored lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms. The black arrows point to the three ‘standards’ in No-Noise discrimination. The red arrow points to a 25° test in the corner of the detection contour.
a) $r^2 = 0.9$

b) $r^2 = 0.87$

c) $r^2 = 0.91$
Figure 4: No-Noise condition for TGS. Discrimination data in polar coordinates (left panels). The angular coordinate is the test color angle, and the radial coordinate is discriminability from the standard (upper right corner), with chance (50%) discriminability at the circumference of the red-dashed disk. The same data are plotted in Cartesian coordinates (right panels). The colored lines indicate performance predicted by the classifier (black=six mechanism, blue=eight mechanism, magenta=eight equally-spaced mechanism model). TGS: a) 0° standard b) 45° standard c) 315° standard. The colored discs denote the test stimuli at the measured threshold. The black squares indicate tests that were selected from the contour of the six mechanism model.
Figure 5: No-Noise condition. Detection thresholds (colored discs) and model fit for observer SAF in Cone Contrast space. Colored lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms.
Figure 6: No-Noise condition for SAF. Format as in Figure 4. a) 45°standard b) 135°standard c) 225°standard
Figure 7: 42° Noise condition. Detection thresholds (colored discs) and model fit for observer TGS in Cone Contrast space. Colored lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms.
Figure 8: 42° Noise condition. Format as in Figure 4. TGS: a) 45° standard

$r^2 = 0.69$
Figure 9: 64° Noise condition. Detection thresholds (colored discs) and model fit for observer TGS in Cone Contrast space. Colored lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms.
Figure 10: 64° Noise condition. Format as in Figure 4. TGS: a) 0° standard b) 48° standard c) 90° standard
Figure 11: 64° Noise condition. Detection thresholds (colored discs) and model fit for observer SAF in Cone Contrast space. Colored lines represent mechanism thresholds and the smooth closed contour is the probability sum of these mechanisms.
Figure 12: 64° Noise condition. Format as in Figure 4. SAF: a) 315° standard

$\text{r}^2 = 0.95$
Chapter 6 General Discussion

The three experiments presented in this dissertation combined chromatic detection, color matching, and discrimination paradigms to search for higher order mechanisms and test a variety of models. It has been shown that a model can produce selective masking with only six linear mechanisms, the same number as the cardinal mechanism model—adding more than six mechanisms to the model never significantly improved the fit. In Experiment 1 (Chapter 3), we replicated an earlier finding of selective masking of tests in the (L, M) plane of cone space when the noise is placed near the corners of the detection contour and found that selective masking is not evidence for large numbers of mechanisms, which differs from the conclusion of Hansen and Gegenfurtner (2013). In fact, we showed that this model also predicts selective masking for noises across a range of angles near $45^\circ/225^\circ$ in cone contrast space and the much wider corresponding range of angles in threshold-scaled MBDKL space. The model also predicts much less selectivity for noises away from the ends of the detection contour (e.g., Fig. 8 & 9; Chapter 3), with the same six mechanisms. This conclusion does not only depend upon the use of the unipolar test stimulus. We have shown the same selective masking with Gabor patch tests, and a six-mechanism (necessarily symmetric) model could account for those selective masking results as well (Eskew & Shepard, 2013; Shepard, Swanson, & Eskew, 2013).

Our detection model is depicted again in Figure 1. Note that S cone input is shown in two of the mechanisms, to be as similar as possible to the cardinal model (Figure 1; Chapter 3). The assignment of this S cone input to these particular mechanisms is speculative here, since we did not modulate S cones in the experiments.
But as discussed, any linear postreceptoral mechanism(s) receiving S cone input will be active in the (L,M) plane (the sole exception being a mechanism that gets only S cone input, for which there is no evidence), and since some of the mechanisms in Figure 1 must get S cone input we have depicted it in the figure for completeness. In future studies, one may want to re-measure detection thresholds where the S-cones are also modulated and refit the data.

Importantly, although our model has six mechanisms like the elaborated cardinal model of Figure 1 in Chapter 3, our model differs in one important respect: four of our six detection mechanisms have opposed L and M cone inputs, rather than only two (top four detection mechanisms in Figure 1). This is the essential feature of our model that allows it to account for selective masking when noise is placed near the ends of the detection contour in the (L,M) plane; noises that are nearly parallel to the long flanking detection contours can cause different mechanisms to become most sensitive and thus determine threshold, tilting the overall detection contour (discussed in Chapter 3).

Figure 1: Six-mechanism model (Shepard et al., 2016). Each mechanism is half-wave rectified. Four of these mechanisms (R and G, and B and O) are ‘quasi-paired’, having nearly equal and opposite L and M cone weights. Two mechanisms (Y and P) have additive L and M cone inputs, and are asymmetric (unpaired). The S cone input to B and O is by analogy to the cardinal model; it was not studied in these experiments.
In Experiment 2 (Chapter 4), a color-matching task provided insight into the subjective experience resulting from these mechanisms and allowed us to further test our six mechanism model. We worked under the assumption that mechanisms are univariant labelled lines (Rushton, 1972; Watson & Robson, 1981). Thus, the colors of physically-different stimuli that are detected by a single mechanism should all be the same, but the colors of two stimuli detected by different mechanisms should be different. Physically-different stimuli that lie along one mechanism threshold line should therefore produce metameric color matches – post-receptoral metamers – while tests that lie on two different mechanism lines should be matched by different colors.

The results showed that color matches fall into no more than six clusters in CIE (u’, v’) space, across all the noise conditions (left panels in Figures 6-9; Chapter 4), and that these clusters correspond closely to the six mechanisms in the model (right panels in Figures 6-9; Chapter 4). Most importantly, there are a number of instances where the detection model determines that a given test angle is detected by different mechanisms in different noise conditions (due to differential masking) even though a given test angle has the same relative L and M cone excitations, irrespective of noise condition. The hue of that test angle also changes in a consistent way.

The color matching data essentially applied a color label to each of the mechanism threshold lines, confirmed the six-mechanism model, and quantified the hue signaled by each mechanism. Interestingly, none of the thresholds looked achromatic for any of our observers in any of our noise conditions. It therefore seems likely that at least one additional pair of mechanisms will eventually be required to account for modulations in three-dimensional cone space (modulating S-cones).
The color matching experiment also ruled out some alternative models with more or less than six mechanisms. As discussed in Chapter 3, a model containing only four mechanisms that has adaptive changes in cone weights in the presence of high contrast noises (Atick, Li, & Redlich, 1993; Zaidi & Shapiro, 1993) could fit the entire set of thresholds, simply by altering the cone contrast weights of the R and G mechanisms in each noise condition to align the long flanks of the detection contour approximately with the noise vector. Two other mechanisms are required (i.e. the Y and P mechanisms) to adequately fit the data (i.e. four mechanisms total). This model is nearly impossible to test using thresholds alone, without theoretical constraints on the adaptive changes in cone weights. However, the color matching data presented in Chapter 4 indicate that the hues of the thresholds fall into six, not four, categories. An adaptive model with only four mechanisms could not easily account for the color matches presented here.

We have also shown that a model with more than six mechanisms (e.g. 8 and 16 mechanisms) can fit the detection thresholds just as well as the 6 mechanism model, but as discussed in Chapter 3, these extra mechanisms do not significantly improve the fit to the thresholds or improve the modeling of selective masking effects. Moreover, quantifying the hues associated with each mechanism, as done here, clearly shows that if there were additional mechanisms they would have to produce redundant hues. The color matches fall into six categories, not more (or less) than six. The color matching experiment strongly supports the six mechanism detection model presented in Chapter 3.

In Experiment 3 (Chapter 5), on a given trial, two threshold-level stimuli from Experiment 1 (one "standard" and one “test”) were presented in random order, and observers were asked to select the standard angle. Threshold level standards and many
tests in the (L,M) plane of cone contrast space were used, with and without the chromatic masking noise used in Shepard et. al., (2016). Following the principles of univariance and labelled lines we were able to make some predictions about discrimination performance, based on the location of the tests and the mechanisms lines in the detection model fits from Experiment 1; the discrimination results confirmed those predictions: 1) when both stimuli were detected by the same chromatic mechanism, performance was at chance level (50%); 2) when the stimuli were detected by two different chromatic mechanisms, discrimination was as good as detection (ca. 82%); 3) when one or both stimuli were detected by multiple mechanisms, performance was intermediate (ca. 65%-75%).

We also used a Bayesian Classifier model (Eskew et al., 2001) to make quantitative predictions of performance in the discrimination task. The classifier was used to combine the outputs of the six mechanisms in the detection model from Shepard et al., (2016a). Note that the classifier’s predictions were computed without any free parameters. Overall, the classifier did an excellent job of predicting the pattern of discriminability for multiple standards in multiple noise conditions—which provides even more evidence for the detection model.

In summary, we developed a new chromatic detection model. With only six unipolar color mechanisms – the same number as the cardinal model – the new model accounts for the threshold contours across the different noise conditions, including the asymmetries and the selective effects of the noises. The key for producing selective noise masking in the (L,M) plane is having more than two mechanisms with opposed L and M cone inputs, in which case selective masking can be produced without large numbers of color mechanisms. Two subsequent experiments (i.e. color matching and
chromatic discrimination) directly tested this model and provide further evidence for this much simpler model of chromatic detection.

If we are to understand how we use color for high-level cognitive tasks in our daily lives, we must understand how the entire color visual system works (i.e. from the photoreceptors to the brain). To do so requires a strong psychophysical model that adequately explains behavior (e.g. color detection and discrimination). Regardless, it is difficult to relate any psychophysical threshold model to the activity of visual cortical neurons, in part because behavioral thresholds are likely dominated by the most sensitive subset of cells. Nonetheless, there are several possible points of comparison between our model and cortical neurons, at least those with foveal or near-foveal receptive fields. First, the mechanisms in our model are, at least to a good approximation, linear combinations of cone signals. Some cortical neurons respond to linear cone combinations, especially in V1, but others are nonlinear and thus more narrowly or more broadly tuned (Russell L. De Valois, Cottaris, Elfar, Mahon, & Wilson, 2000; Horwitz & Hass, 2012; Lennie, Krauskopf, & Sclar, 1990); it is not generally clear how sensitive the linear cells compared to the nonlinear ones. Second, there are only six mechanisms in our model, which might suggest that the especially sensitive cells would fall into six clusters in terms of their cone weights. Many physiological studies report cells with a large variety of chromatic tunings (Russell L. De Valois et al., 2000; Horwitz & Hass, 2012; Komatsu, 1998), but again it is not clear that these are highly-sensitive, nor even that they all actually have to do with color vision (e.g., Horwitz & Hass, 2012); the fact that they respond to chromatic stimuli might only be a result of irrelevant variation in cone connectivity (Conway, 2009). Third, based upon our model, most of the sensitive cells
should have opposed L and M cone inputs; L/M opponency is common in many cortical cells but it is not clear that it predominates (e.g., Horwitz & Hass, 2012). Fourth – and most optimistically – the cells might have cone contrast weights similar to those in Table 1 (Chapter 3). Of course, it is important to keep in mind that, even among cortical cells that actually serve color vision, there are likely to be neurons that do not satisfy the definition of psychophysical color mechanisms: univariant labelled lines with fixed relative spectral tuning (e.g. tuning may change with contrast; Horwitz & Hass, 2012; Namima, Yasuda, Banno, Okazawa, & Komatsu, 2014; Solomon & Lennie, 2005); (see R Eskew, 2009) for discussion).
References


General Appendix 1 - Gamma Correction

The color and intensity of a pixel on a computer screen is defined by three numbers: (R)ed, (G)reen and (B)lue. These three numbers represent the amount of voltage supplied to the red, green and blue CRTs (cathodes) for that pixel. R, G, and B can range in value from 0 to 1023 for any display system, such as ours, with 10 bits of resolution in the digital to analog converters (DACs). Before we can correct the gamma, we must first measure the relationship between the R, G, and B guns and luminance for a pixel—this relationship is non-linear. We determine the relationship between these values and define it as a function. Applying the inverse of that function to these R, G, and B values produces a linear relationship that is used to create the stimuli desired for a given stimulus. The gamma correction of our experimental monitor was achieved via the four-step process described below.

Step 1: Measuring gamma functions

Initial gamma functions for the experimental monitor were measured for each of the red, blue and green guns separately. There are many ways in which these measurements can be made. For the measurement of the red gun, one could set the values of the green and blue gun to zero, effectively eliminating the contribution of the green and blue guns to the luminance measurement. The red value could then be incremented across the full range of possible values while measurements are taken (0-1023). However, this method is not used here for two reasons. First, light meters (including the PR-650 used here) often have minimum light requirements making measurements very low light levels impossible. The second reason is that it is important to make measurements under conditions that are as similar to the experimental conditions as possible. All of the experiments in this
dissertation were completed on a mean grey background field (outside of the noise area) so we instead used a modulation method where we measured the gamma of the Red gun and the values of Green and Blue were kept at half of the luminance that they are capable of producing. If the gamma were linear, this value would be 512 (i.e. \( \frac{1}{2} \) of 1023), so we used this value for the two guns that were not being measured. While, the values of G and B were kept at 512, R is incremented along its full range in a central test patch. Figure 1 shows the R, G, and B values as measured with the bipolar method—note that the three curves cross at 512—they must since all three values are the same at this point.

![Figure 1: Red, green and blue lines represent the luminance of that color gun given that the other two guns are set at 512.](image)

**Step 2: Spectroradiometer (PR-650) Measurements**

A PhotoResearch PR-650 spectroradiometer was used to make all of the measurements. This device was connected to the computer that controls the display through a serial port connection. A software program was used to communicate with the PR-650. The program
varied the properties of the stimulus on the display and took a luminance measurement for each increment. The test patch was a 150x150 pixel square in the center of the screen, where the test stimulus in each of the experiments was displayed. The PR-650 was placed 32 inches from the monitor in a darkened room. The monitor was allowed to warm up for at least an hour before collecting any measurements. Each gun was individually incremented in steps of 8 from 0 through 1023 and a measurement was made at each step.

**Step 3: Analysis and linearization**

Since our measurement increments were greater than 1, interpolation is required to get values for each R, G, and B value between 0 and 1023. To linearize our gamma we create a 1024x3 table of all R, G and B values scaled to a maximum of 1. Our linearization routine takes [R G B] values and writes different values to the video card based on this table. For example, if a patch on the screen with values [0 512 512] is desired, the linearization routine calculates 0/1023, 512/1023 and 512/1023 and then looks for these values in the table, and uses the corresponding [R G B] instead of [0 512 512]. In this case, the calculations turn out to be [0 0.5 0.5] so the values that correspond to 0, 0.5, and 0.5 of the maximum luminance for the R, G and B guns respectively are used.

**Step 4: Gamma Check**

Once the linearization is complete, we make the same measurements again for R, G, and B, this time using the new linearized values. The gamma check will check the accuracy of the gamma correction. Rather than reflecting the underlying gamma function (i.e. non-linear), the new values should produce three straight lines that cross at R=G=B=512, as shown in figure 2. The Gamma check is normally performed every couple of months to
ensure that the monitor and computer setting have not changed. If the three lines are not straight, then the gamma must be corrected again (repeat process).

Figure 2: Red, green and blue lines represent red, blue and green guns respectively. The lines are linear showing that the gamma correction has been corrected for the native non-linear relationship between the guns.