

Do common mechanisms of adaptation mediate color discrimination and appearance? Uniform backgrounds

James M. Hillis and David H. Brainard

*University of Pennsylvania, Department of Psychology, 3401 Walnut Street 302C,
Philadelphia, Pennsylvania 19104*

Received February 8, 2005; revised manuscript received April 14, 2005; accepted April 18, 2005

Color vision is useful for detecting surface boundaries and identifying objects. Are the signals used to perform these two functions processed by common mechanisms, or has the visual system optimized its processing separately for each task? We measured the effect of mean chromaticity and luminance on color discriminability and on color appearance under well-matched stimulus conditions. In the discrimination experiments, a pedestal spot was presented in one interval and a pedestal + test in a second. Observers indicated which interval contained the test. In the appearance experiments, observers matched the appearance of test spots across a change in background. We analyzed the data using a variant of Fechner's proposal, that the rate of apparent stimulus change is proportional to visual sensitivity. We found that saturating visual response functions together with a model of adaptation that included multiplicative gain control and a subtractive term accounted for data from both tasks. This result suggests that effects of the contexts we studied on color appearance and discriminability are controlled by the same underlying mechanism. © 2005 Optical Society of America © 2005 Optical Society of America

OCIS codes: 330.1690, 330.4060, 330.7720.

1. INTRODUCTION

Color signals support two important visual functions. First, color is useful for discriminating objects from their background (i.e., for segmenting the image into regions that correspond to physically distinct surfaces). Second, as the perceptual correlate of surface reflectance, color is useful for identifying objects. These functions are conceptually distinct, and each function may place different demands on visual processing. In this paper, we ask whether color discrimination and color appearance are mediated by the same mechanisms of adaptation. We refer to this proposition as the common mechanism hypothesis.

Both sensitivity to color differences and color appearance depend on the context in which the stimulus is placed. Color difference sensitivity is revealed by threshold measurements. Thresholds for detecting a monochromatic spot presented against a spatially uniform monochromatic background vary with the wavelength and intensity of the background.¹ The dependence of threshold on context reveals a form of chromatic adaptation, wherein the processing of a focal stimulus changes with context. There is an extensive literature examining how chromatic adaptation affects visual sensitivity (for reviews see Refs. 2–6).

Chromatic adaptation is also revealed by experiments that assess color appearance. For example, the judged appearance of a test spot varies with the spectral composition of the background against which it is seen.⁷ The effect of adaptation on the appearance of stimuli is also the subject of an extensive literature found under the rubrics of both color appearance^{8–10} and color constancy.^{11,12}

Research examining effects of context on color discrimination and color appearance has often proceeded along separate paths. It is relatively rare for a single empirical paper to report results from both paradigms. Nonetheless results from discrimination and appearance experiments are typically accounted for within the same general model of color processing and adaptation.^{13,14} In broad outline, this model incorporates encoding of light by three classes of cone photoreceptors, recombination of cone signals into a luminance and two color-opponent channels, and sensitivity regulation at sites both preceding (first-site) and following (second-site) the signal recombination.^{15,16}

Numerous variants of the general model have been proposed to account for specific features of both discrimination and appearance data.^{3,17–22} It is not clear whether the variations reflect fundamental differences between visual processing for discrimination and appearance, differences in stimulus conditions studied, or differences in the way individual investigators have treated their data. We have thus undertaken a systematic investigation of the relation between mechanisms of visual processing revealed by discrimination and appearance.

We are, of course, not the first to consider the relationship between color discrimination and appearance. A few studies have explicitly tested the idea that specific aspects of color discriminability and appearance are controlled by a common set of mechanisms.^{23–25} If we broaden attention to intensity discrimination and lightness/brightness perception, additional studies have tested Fechner's hypothesis²⁶ that the rate of apparent stimulus change is inversely proportional to discrimination thresholds. This hypothesis may be used to derive a

visual response function from discrimination data;^{27–32} for review see Ref. 33. This response function relates an inferred perceptual dimension to stimulus magnitude and can be used to predict appearance. We adopted this general approach. Our formulation is very close to that developed by Heineman,³⁴ but subsequent advances in models of early visual processing and numerical data fitting procedures allow us additional leverage.

This paper describes the experimental paradigms we used and develops our formal analysis. Here we report the results of investigations for adaptation to changes in the luminance and chromaticity of uniform backgrounds.

A. Methods Overview and Formal Model

1. Experimental Paradigms

We conducted two types of experiments: color discrimination and asymmetric matching. In the color discrimination experiments, test stimuli were excursions from the background color that isolated either the L and M cones or the S cones. The test stimuli may be represented by the expression $[L_{bg} M_{bg} S_{bg}] + I[L_{test} M_{test} S_{test}]$ where the vectors indicate LMS cone coordinates.³⁵ More specifically $[L_{bg} M_{bg} S_{bg}]$ indicates the cone coordinates of the background and $[L_{test} M_{test} S_{test}]$ represents the color direction of the excursion. For example, $[L_{test} M_{test} S_{test}] = [0 \ 0 \ 1]$ represents an S-cone isolating test. The value of I specifies the intensity of the stimulus excursion in units yoked to $[L_{test} M_{test} S_{test}]$, and we refer to it as the test intensity.

We used a two-interval-forced-choice (2IFC) task to measure the probability that observers could discriminate two stimuli with intensity parameters I and $I + \Delta$. These data provide the probability that two stimuli can be discriminated as a function of the value of the pedestal I and intensity difference Δ . Measurements were made for tests presented against a number of backgrounds, and observers' performance as a function of the test direction, pedestal intensity, and intensity difference were used to characterize the effect of chromatic adaptation on color discrimination.

In the asymmetric matching experiments, observers adjusted a matching spot against one background to appear the same as a test presented against another. Differences between the LMS coordinates of observers' matches and those of the corresponding tests were used to characterize the effect of chromatic adaptation on color appearance.

2. Model

The core idea underlying the model of mechanisms controlling appearance and discrimination judgments is illustrated in Fig. 1 and described more completely with Fig. 2. The two curves in Fig. 1 are stimulus-response functions of the same mechanism in two different contexts. Each curve defines the relationship between stimulus intensity and response magnitude for one state of adaptation. Given these response functions, we can predict appearance matches across the context change and discrimination thresholds measured within each context. Equality of appearance across the context change is predicted by equality of mechanism response. In Fig. 1, the stimuli indicated by the downward arrows are predicted

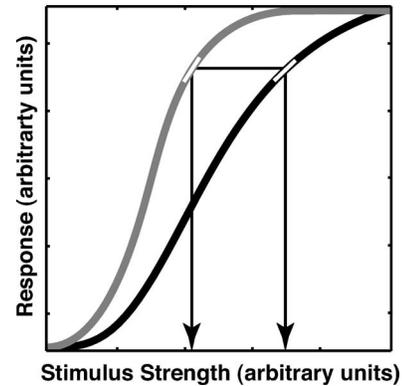


Fig. 1. Two hypothetical response functions. Each function corresponds to a different context (e.g., different chromaticity of background) and plots a mechanism's response as a function of stimulus strength (e.g., test spot contrast). Equality of appearance across the context change is predicted by equality of mechanism response, so that the stimuli indicated by the downward arrows would be predicted to match in appearance. Discrimination thresholds (i.e., test increment required to discriminate between a pedestal presented alone and a pedestal plus the test) are predicted to be inversely proportional to response function slope.

to match in appearance across the change in context. Discrimination thresholds are predicted to be inversely proportional to response function slope. For example, given the two stimuli indicated by the downward arrows, discrimination thresholds would be higher for the response function on the right (black curve) because equal changes in the stimulus intensity yield smaller expected response changes (i.e., the slope of the white line segment superimposed on the black response function is lower than the slope of the white line segment superimposed on the gray response function). As will be detailed below, smaller expected response changes translate into a diminished ability to discriminate I and $I + \Delta$. In our work, we make both types of measurement and ask whether they may be accounted for by the same pair of response functions.

To link data from the 2IFC discrimination and asymmetric matching tasks, we relied on Fechner's assumption²⁶ that sensitivity to changes in stimulus intensity is proportional to the rate of apparent stimulus change. Figure 2D shows two hypothetical psychometric curves from the discrimination task for a single choice of background. To stage ideas tangibly, imagine that this background appears green and that the test spots are an excursion along a color direction that modulates the L and M cones together. The two curves show performance for two different pedestals I_1 and I_2 (indicated above the panel). Consider the left curve, which corresponds to the less intense pedestal I_1 . The x axis on the graph shows test intensities, while the y axis is probability of correct discrimination. Each point on the curve represents the probability that the observer correctly discriminates a pair of tests, one with intensity I_1 and the other with intensity $I_1 + \Delta$. The value of $I_1 + \Delta$ is specified by the position on the x axis. Similarly, the right curve is for pedestal I_2 . This curve has a shallower slope, indicating that discrimination performance declines with the increase of pedestal intensity.

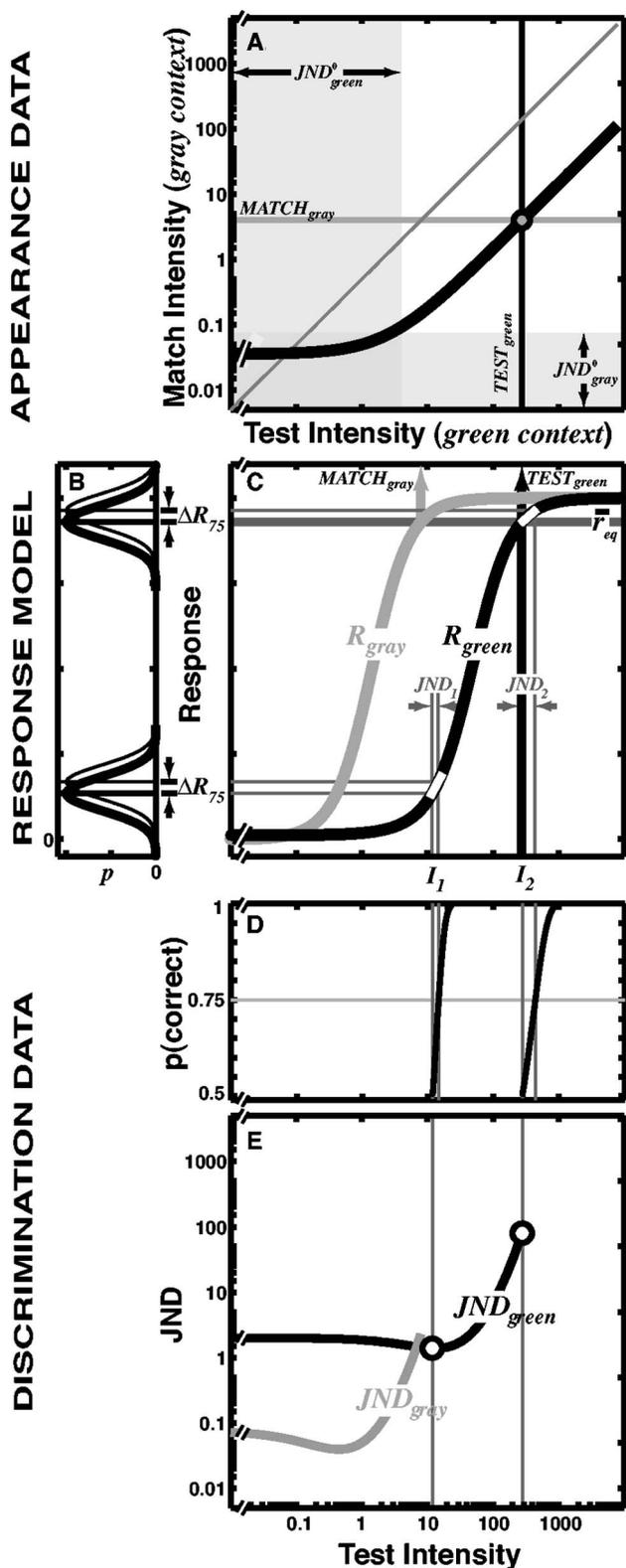


Fig. 2. Intensity response model with expected asymmetric matches and discrimination performance. The x axes in panels A,C,D, and E test stimulus intensity. Panel C shows the response model [Eq. (2)]. The y axis is the magnitude of the response. The two curves in Panel C are the response functions for one mechanism in two adapted states. The gray curve represents expected response in a *gray* context and the black curve represents expected response in a *green* context. The difference between the two curves is a difference in the gain-and-subtractive parameters [g and s in Eq. (2)]. Panel B shows the Gaussian response distributions for four test intensities ($I_1, I_1 + \Delta I_{1,75}, I_2, I_2 + \Delta I_{2,75}$ where $\Delta I_{i,75}$ is the incremental intensity that yields 75% correct performance) presented in the *green* context. The x and y axes are probability and response magnitude, respectively. Panel D shows two psychometric curves for pedestals I_1 and I_2 . These curves are the expected discrimination performance derived from the model characterized in panels B and C for the response function adapted to the *green* context. Panel E shows the increments expected to yield 75% correct (JNDs) as a function of pedestal intensity derived from the response model in B and C. Finally, Panel A shows the expected performance in an asymmetric matching task with the tests presented in the *green* context and the matches set in the *gray* context. The height of the shaded rectangle running the width of the x axis represents absolute threshold for a spot presented against a gray background (JND_{gray}^0 , which is equivalent to the point where the JND_{gray} curve in Panel E intersects the y axis). The width of the shaded rectangle running the height of the y axis represents absolute threshold for a spot presented against the green background JND_{green}^0 . Any test-match pairs in this gray shaded region would be extremely difficult to obtain.

Conventionally, one point from the psychometric curves shown in Fig. 2D is selected and identified as a just-noticeable difference (JND). The 75% JNDs from the two psychometric curves in Fig. 2D are plotted as open circles in Fig. 2E. In this plot, it is clear that discrimination performance is worse at pedestal I_2 than at pedestal I_1 . More

generally the black line in Fig. 2E shows JNDs as a function of pedestal intensity, measured against the green background. Such curves are typically called tvi (threshold versus intensity) curves. The tvi curve shown has the “dipper” shape often found in this type of experiment.^{29,31,36,37} The gray curve in Fig. 2E shows a

hypothetical tvi curve for tests presented against a second background, one that appears gray. The change in background induces a change in the tvi curves.

What do such data reveal about the mechanisms that underlie performance? We followed earlier authors^{28–32} and adopted a functional equation solution²⁸ to Fechner's proposal that JNDs along a single stimulus dimension represent equal changes in response. Equivalently, in signal detection theoretic terms, we assumed that performance at all intensity levels is limited by additive Gaussian noise of fixed variance. Along with the assumption that a single mechanism mediates performance for all pedestals, differences in JNDs at I_1 and I_2 reflect differences in the slope of the response function at these two intensities. Thus if we know the response function and noise standard deviation, we can predict JNDs for any pedestal. More generally, we can predict the probability that any pair of tests I and $I + \Delta$ will be discriminated.

Figures 2B and 2C show this logic graphically. The y axis of both panels represents inferred response magnitude. The x axis in Fig. 2C is stimulus intensity and the x axis in Fig. 2B is probability (of obtaining a particular response magnitude). The distribution of responses to pedestal I_1 is represented by the lowest Gaussian distribution (thick black line) in Fig. 2B. For this noise distribution, we can calculate the probability that any Δ added to I_1 will give a higher response. These probabilities correspond directly to the y axis of Fig. 2D: the probability that an increment will be detected. The probability that I_1 yields a smaller response than $I_1 + \Delta$ is

$$\begin{aligned} p(R_{\text{green}}(I_1) < R_{\text{green}}(I_1 + \Delta)) & \\ &= p(R_{\text{green}}(I_1) - R_{\text{green}}(I_1 + \Delta) < 0) \\ &= p(n(\bar{r}_{\text{green}1}, \sigma^2) - n(\bar{r}_{\text{green}1+\Delta}, \sigma^2) < 0) \\ &= p(n(\bar{r}_{\text{green}1} - \bar{r}_{\text{green}1+\Delta}, 2\sigma^2) < 0) \\ &= \int_{-\infty}^0 n(\bar{r}_{\text{green}1} - \bar{r}_{\text{green}1+\Delta}, 2\sigma^2), \quad (1) \end{aligned}$$

where R_{green} represents the response function when tests are presented against a green background, n represents the normal distribution function, $\bar{r}_{\text{green}i}$ is the mean response to increment i , and σ^2 is response variance. The last line of this equation is a cumulative Gaussian with mean $\bar{r}_{\text{green}1} - \bar{r}_{\text{green}1+\Delta}$ and variance $2\sigma^2$. When integrated to 0, this yields the probability that I_1 yields a smaller response than $I_1 + \Delta$.

The above logic tells how to predict discrimination data given a response function. Our interest is in going the other way: using discrimination data to infer the response function.^{26,28–32} In Fig. 2C, JND_1 and JND_2 represent test intensities (Δ) that yield, on average, a fixed probability (0.75 in this example) that the observer discriminates I_1 from $I_1 + \text{JND}_1$ and $I_2 + \text{JND}_2$. These JNDs correspond to equal average changes in the neural response and therefore reflect the local slope of the response function (white line segments through the intersections of the lines representing ΔR_{75} and $\text{JND}_{1\&2}$ in Fig. 2C show the slopes of the response curves that would be inferred at pedestals I_1 and I_2). Thus using discrimination performance, we can

piece together local estimates of response function slopes to infer the shape of the stimulus-response function.

Since in practice discrimination data sample the intensity dimension fairly coarsely, inferring the response function is done most effectively using a parametric description. We chose a parametric form previously shown to account for both psychophysical and physiological data:^{29,31,38,39}

$$R = M \frac{(gI + s)^p}{(gI + s)^q + 1}. \quad (2)$$

Here M controls the height of the curve, p and q determine the rate of expansion at low intensities and the rate of saturation at high intensities, g is a gain parameter, and s is a subtractive term. Given this form, numerical search can be employed to find parameter values that best account for observed discrimination performance. The parameter M trades off perfectly with noise standard deviation, so we fixed noise standard deviation at unity throughout.

The response curve labeled R_{green} in Fig. 2C represents the response function [Eq. (2)] that would be inferred from the green background tvi curve in Fig. 2E. Similarly, the response curve labeled R_{gray} in Fig. 2C represents the response function [Eq. (2)] that would be inferred from the gray background tvi curve.

Inferred response functions allow us to predict asymmetric matching data. We use the linking hypothesis that two tests will appear the same when they produce the same mechanism response. Consider again Fig. 2C. When a test stimulus, $\text{TEST}_{\text{green}}$ (indicated on the top axis of Fig. 2C; note that the subscript refers to the color of the background), is presented against the green background it yields an expected response of \bar{r}_{eq} (shown on the right axis of Fig. 2C). Under the linking hypothesis stated above, observers will set the intensity of a spot against a gray background to yield the same response. This expected intensity may be found from Fig. 2C by finding the appropriate point on the response curve R_{gray} and is indicated by $\text{MATCH}_{\text{gray}}$.

Figure 2A plots the intensities of spots that match across a change in background. The gray circle represents the expected $\text{TEST}_{\text{green}} : \text{MATCH}_{\text{gray}}$ pair as determined by the response functions in Fig. 2C. The thick black line represents the locus of $\text{TEST} : \text{MATCH}$ pairs obtained by repeating this procedure for different response levels. The key point is that the inferred response functions provide a quantitative link between discrimination data (Figs. 2D and 2E) and asymmetric matching data (Fig. 2A). We used this link to determine whether the mechanisms controlling adaptation to background chromaticity are common to both discrimination and appearance judgments. This analysis assumes that all tests are processed by the same mechanism. Our stimuli were chosen to make this assumption reasonable.

2. METHODS

A. Observers

Two observers participated. Observer JMH was an author and observer QRS was a paid volunteer. Observer QRS

was unaware of the experimental hypotheses and had little previous experience in psychophysical experiments. Both observers had normal color vision as assessed by an Ishihara test for color blindness.

B. Apparatus

Stimuli were presented on a calibrated RGB monitor (HP p1110) with 14-bit intensity resolution for each channel (provided by a Cambridge Research System BITS++ device) operating at a 75 Hz refresh rate. Gamma functions and phosphor spectral power distribution were measured for each CRT phosphor with a Photo Research 650

spectra-radiometer. At the 40 cm viewing distance, the monitor subtended 45 deg \times 37 deg (each pixel subtended \sim 0.041 deg). The observer's head position was stabilized with a chin rest.

C. Stimuli

Test stimuli were spots (1.5 deg diameter) convolved with a 2D Gaussian to produce a smooth intensity ramp at the edge (as shown in the top panel of Fig. 3). They were located 3 deg, on average, horizontally from a central fixation point. One test spot was located to the left of fixation and the other to the right. To avoid adaptation to the tests, their locations were perturbed from trial to trial. Perturbations were selected from a uniform square distribution, and the same perturbations were used for left and right tests on each trial, so that the locations were always symmetric with respect to a vertical axis. The xy position of the left and right tests were thus $[-3\ 0] + [\epsilon_x\ \epsilon_y]$ and $[3\ 0] + [-\epsilon_x\ \epsilon_y]$, respectively, where ϵ_x and ϵ_y were random draws from a uniform distribution:

$$P(\epsilon) = \begin{cases} 0 & \text{for } \epsilon < -1.5 \\ \frac{1}{3} & \text{for } -1.5 < \epsilon < 1.5 \\ 0 & \text{for } \epsilon > 1.5 \end{cases}$$

On every trial, test locations were indicated by square frames composed of sparse black points. The frame had 8 points per side, including the corner points. Each point was 0.041 deg in diameter and the points were separated by 0.23 deg. The frames appeared 13 ms before each trial and were extinguished 13 ms after the test appeared. Intensity of the test spots was ramped on and off gradually. The ramp was a cumulative Gaussian with standard deviation = 40 ms (3 frames) such that the total duration of the ramp was 133 ms (11 frames). The test intensity was fixed for 200 ms (15 frames) between the on and off ramps. The bottom panel of Fig. 3 shows spatial and temporal profiles of the test spots.

The test spots were modulations relative to the background. That is, test modulations were defined as $[L_{bg}\ M_{bg}\ S_{bg}] + I[L_{test}\ M_{test}\ S_{test}]$. We used the Smith-Pokorny estimates⁴⁰ of cone spectral sensitivities and standard methods⁴¹ to convert between cone coordinate specifications and video DAC settings.

Test spots were presented in various contexts defined by a uniform background. We used five backgrounds in the experiments reported here. Table 1 provides the CIE chromaticity and LMS cone coordinates of the backgrounds, along with a descriptive color name for each. The gray background served as a reference with respect to which the other backgrounds were defined. These differ

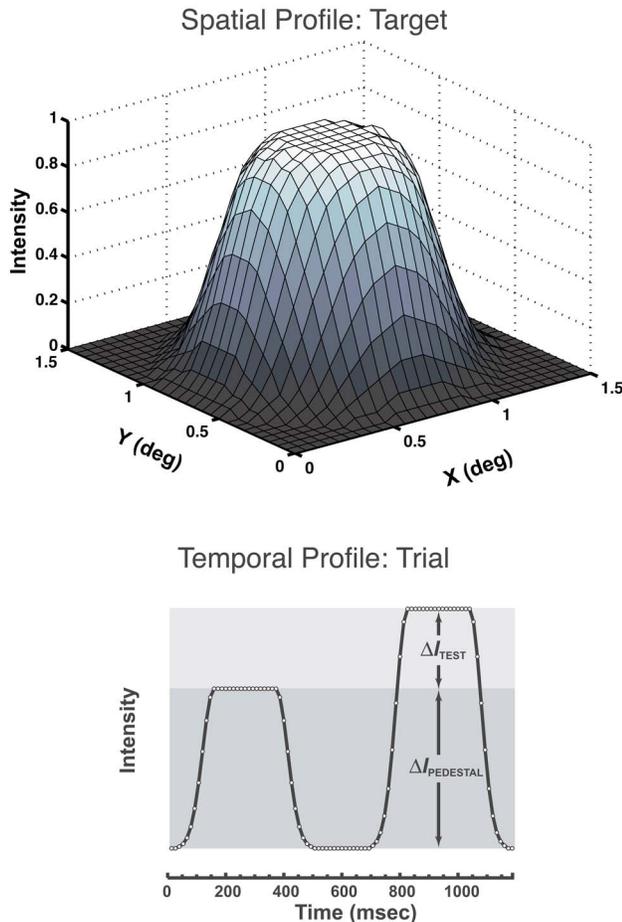


Fig. 3. Spatial and temporal profiles of test spots. The top panel shows the spatial profile of test spots in the discrimination and matching experiments. The bottom panel shows the temporal parameters of a trial in the discrimination experiment. The small white spots indicate frame timing (vertical blanking). The temporal profile in the matching experiment was the same except that only one interval was used.

Table 1. Properties of Test Backgrounds

Condition Name	Color Name	CIE Coordinates			Cone Isomerizations ($\times 10^6$)		
		x	y	Y	L	M	S
GRAY	Gray	0.310	0.310	22.5	2.72	1.01	0.03
Gray+LM	Pale green	0.313	0.368	31.1	4.03	1.59	0.03
Gray-LM	Pale red	0.306	0.225	13.3	1.42	0.44	0.03
Gray+S	Blue-purple	0.280	0.240	22.5	2.72	1.01	0.01
Gray-S	Brownish-yellow	0.365	0.437	22.5	2.72	1.01	0.05

from the gray background by steps that modulate the LM or S cones as specified in the table. LMS units in the table are the expected number of isomerizations from the background light for the area and duration of test stimuli. Test intensities will be reported as the expected number of isomerizations independent of the background. Thus the background LMS coordinates in Table 1 can be used to compute a contrast for the test intensities reported in the results:

$$C = \frac{\text{test isomerizations}}{\text{background isomerizations}}.$$

Test spots were increments or decrements modulated in two color directions. One direction modulated the L and M cones together and in equal amounts, with S cone stimulation held constant. That is, $[L_{\text{test}} M_{\text{test}} S_{\text{test}}] = I[1 \ 1 \ 0]$ such that L and M isomerization rates for individual cones varied by the same amount. The total number of L and M cone isomerizations will differ when there are different numbers of each cone type in the stimulus area. To compute isomerization totals, we used an L:M cone ratio of 2:1, which means that for this test direction there would be twice as many L-cone isomerizations as M-cone. L- and M-cone contrasts for this test direction are

$$C_L = \frac{2}{3} I_{\text{test}} / I_{\text{bg,L}}, \quad C_M = \frac{1}{3} I_{\text{test}} / I_{\text{bg,M}},$$

where $I_{\text{bg,L}}$ and $I_{\text{bg,M}}$ are listed in Table 1 and I_{test} are test isomerization totals. The other direction modulated the S cones while the L- and M-cone stimulation was constant. That is, $[L_{\text{test}} M_{\text{test}} S_{\text{test}}] = I[0 \ 0 \ 1]$.

D. Procedure: Discrimination

We used a pedestal + test paradigm to measure threshold versus pedestal intensity (tvi) curves for spots modulated in different color directions against different backgrounds. The definition of the pedestal and test is shown graphically in the lower panel of Fig. 3. On the trial represented, the pedestal alone was shown in the first interval and the pedestal + test was shown in the second interval. The test and pedestal were always in the same color direction. Observers were instructed to select the interval in which they saw the more “intense color.” Before experimental trials began, observers were given practice trials with auditory feedback (used throughout the experiment) where the test was clearly visible. Observers were instructed to run practice trials until they were certain of the apparent color change that corresponded to the pedestal + test. Once they were certain, they pressed a button on a game pad that initiated a 90 s adaptation period. There was a minimum of 2 s between each trial. One pedestal direction and intensity were selected for each experimental session. The number of pedestals tested for each background and test/pedestal color direction depended on limits imposed by the gamut of the monitor.

Of particular concern was matching the stimulus conditions to those used in the asymmetric matching task (described below) as closely as possible. This introduces one uncommon aspect of the procedure: there were two spots presented on each stimulus interval. In one interval

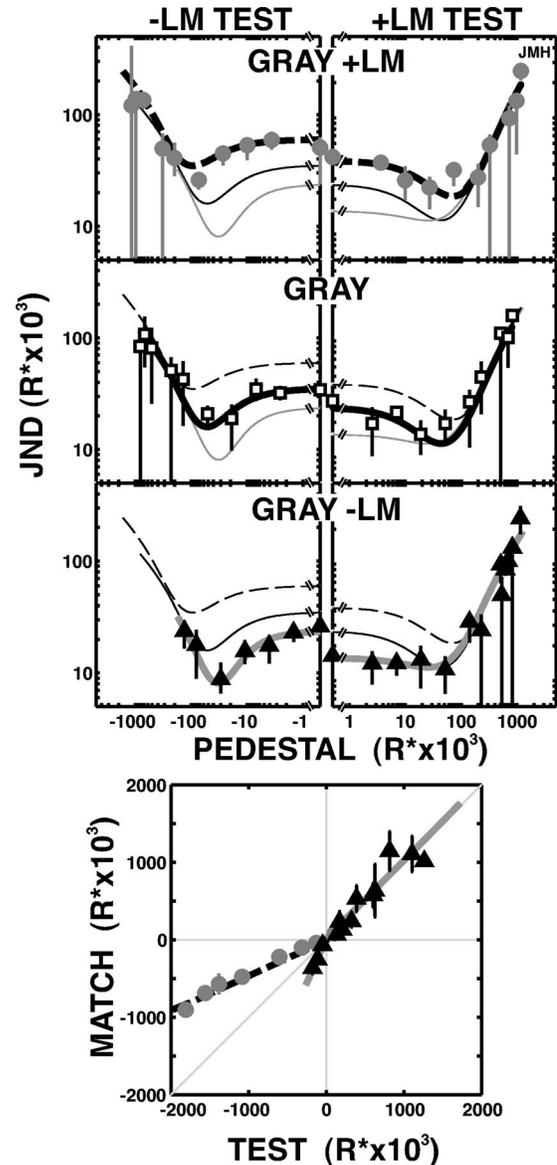


Fig. 4. JMH's results from discrimination and matching experiments for LM tests and backgrounds that varied in their LM input. The top six panels are discrimination thresholds (JNDs) plotted as a function of pedestal intensity. The three left panels are JNDs for decrements and the three right panels are JNDs for increments presented on, from top to bottom, the Gray+LM, Gray and Gray-LM backgrounds. Error bars are 95% confidence intervals. The bottom panel shows asymmetric matching data for the same set of conditions. The x axis is the test excursion (i.e., the number of isomerizations expected from the test independent of the background) against the Gray+LM (circles) and the Gray-LM (triangles) backgrounds. The y axis is the match excursion from the gray background against which the matches were set. Error bars are standard error of the mean. Dashed and solid lines are results of model fits with parameters selected on the basis of both discrimination and matching data. These are fits for the gain-and-subtractive model of adaptation. Increments and decrements were fit independently but the p , q , and M parameters in Eq. (2) were yoked across adapting conditions. The raw psychometric and matching data underlying the points and model fits shown in this figure and Figs. 4-7 can be obtained at http://color.psych.upenn.edu/supplements/com_uniform/.

the pedestal was presented alone in the left and right target locations. In the other interval, the pedestal was pre-

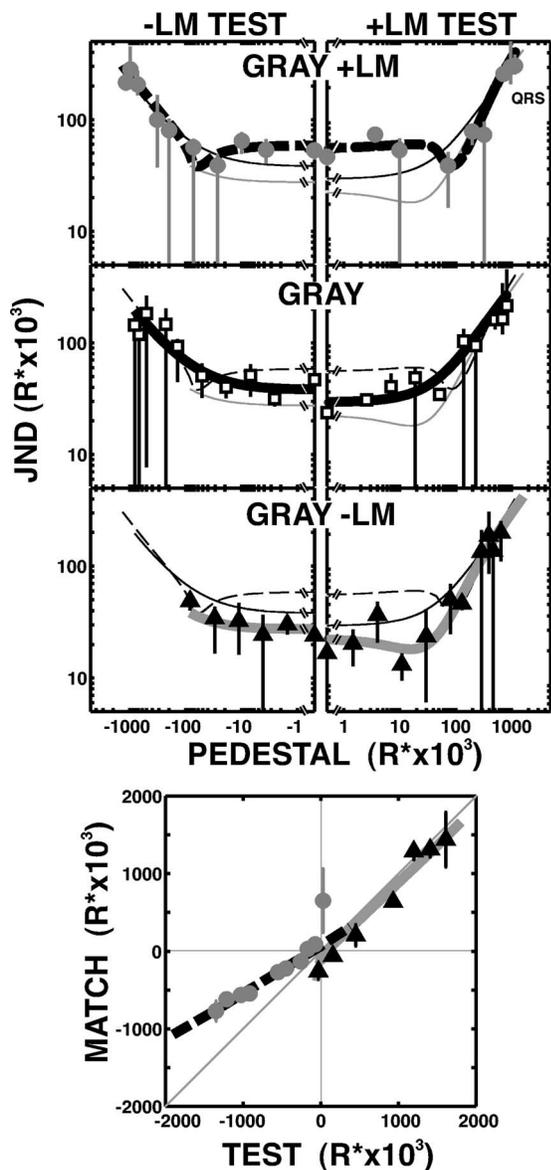


Fig. 5. QRS's results from discrimination and matching experiments for LM tests and adapting fields that varied in their LM input. Results are plotted in the same format as Fig. 4.

sented alone at one target location and the pedestal + test was presented in the other interval. Observers indicated with a key press which interval they believed contained the test, independent of which side it appeared on. A feedback tone indicated when observers selected the incorrect interval. The intensity of the test was controlled by staircase procedures. Four randomly interleaved staircases were used in each experimental session: two for tests shown on the right and two for tests on the left. We used 2-down-1-up and 3-down-1-up staircase rules to ensure comprehensive sampling of psychometric functions for each pedestal. Typically, one session was run for each pedestal intensity. In cases where the data were particularly noisy, an additional session was run.

To reduce the number of sessions, we put the same context on both the left and right halves of the display. To keep the same split-field conditions as in the matching ex-

periment would require that each observer run in twice the number of sessions per context (because half of the data would always be against the neutral context). A control experiment, presented below, compared discrimination performance for split-field and uniform-field contexts for a subset of conditions. The results indicate no differences that would affect the analyses presented in this paper.

E. Procedure: Appearance

We used an asymmetric matching task to measure the effect of context on color appearance. Observers adjusted the color of one of two simultaneously displayed test spots until they appeared identical. We gave observers full control over the three-dimensional LMS coordinates of the adjustable match spot. One spot appeared on the left and the other on the right of fixation. One half (either left or right; 22.5 deg x 37 deg) of the monitor was filled with the

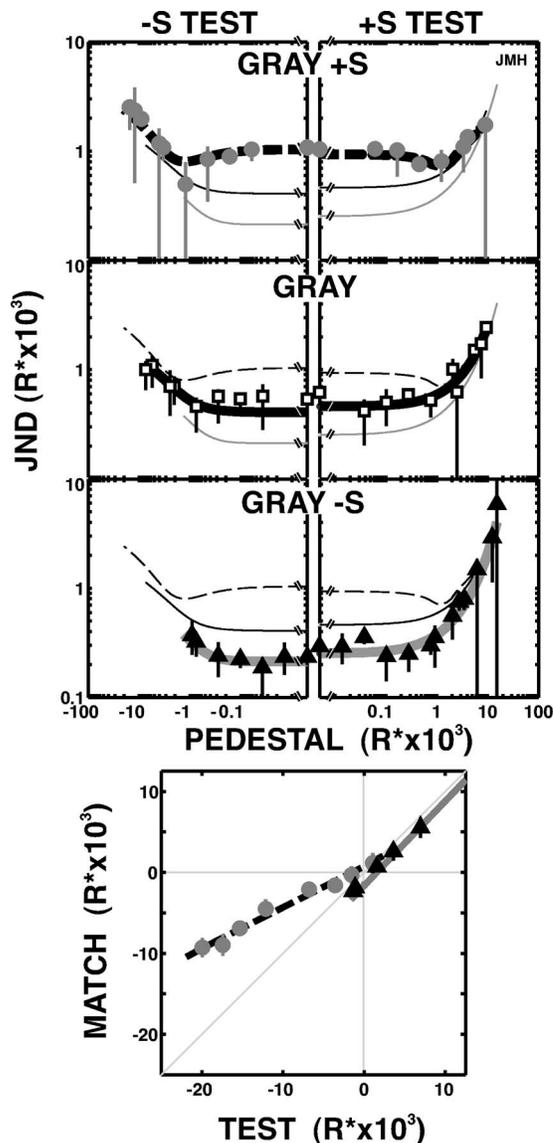


Fig. 6. JMH's results from discrimination and matching experiments for S-cone tests and adapting fields that varied in their S-cone input. Results are plotted in the same format as Fig. 4 except that values correspond to expected S-cone isomerizations.

gray background. The other half had the adapting background. Observers were instructed to fixate a small black point at the border of the two backgrounds throughout the experiment. There was an initial 90 s adaptation period prior to the first time the test spots were displayed.

Each time the test spots were displayed, the fixed test spot, called the standard, appeared against the adapting background and the adjustable spot, called the match, appeared against the gray background. The spatial and temporal profiles of the test spots were the same as in the discrimination experiment, with the exception that the spots were flashed once per exposure rather than twice. Observers used a game controller to adjust the match in the CIELAB $L^*a^*b^*$ coordinates. Approximately red–green (a^*) and blue–yellow (b^*) adjustments were made by pressing correspondingly colored buttons on the game controller. Luminance (L^*) adjustments were made with a joystick. The standard and match were displayed after each adjustment with a minimum interstimulus interval of 2 s. The standard and match could also be displayed without making an adjustment by pressing an appropriate key on the game pad. Observers could, at any time, choose any one of four step sizes for the adjustments. After completing a match, observers rated the quality of the match with a value between 0 (couldn't make the match) and 3 (perfect match).

Four matches were completed in a single experimental session (typically lasting 20 min). Each match was made to a different standard, but the standards used within session were always selected from the same color direction. For each direction in LMS color space, we measured match settings for eight standard tests. Two to four total matches were made for each standard color for each of the five contexts. To the extent possible, stimuli were left–right counterbalanced across sessions.

3. RESULTS

Figure 4 shows one of JMH's discrimination and matching results for LM tests presented on the Gray, Gray+LM, and Gray–LM backgrounds. The top six panels plot discrimination thresholds as a function of pedestal intensity. The left column shows thresholds for decrements (–LM) and the right column shows thresholds for increments (+LM). Note, therefore, that points intersecting the y axis are detection thresholds for decrements (left column) and increments (right column). The three rows, from top to bottom, correspond to three adapting conditions: Gray+LM, Gray, and Gray–LM. Units on abscissa and ordinate are the expected number of isomerizations for the pedestal and test, respectively (i.e., the x axes represent the expected number of isomerizations attributable to the pedestal independent of the background, and the y axis represents the expected number of isomerizations attributable to the test independent of the background and pedestal). The quantities and calculations used to estimate isomerizations are detailed in Appendix A. Contrast units can be computed using the background isomerizations in Table 1 as specified in the methods.

Data points (circles, squares, and triangles for the Gray+LM, Gray, and Gray–LM backgrounds, respectively) are JNDs determined by fitting the raw psycho-

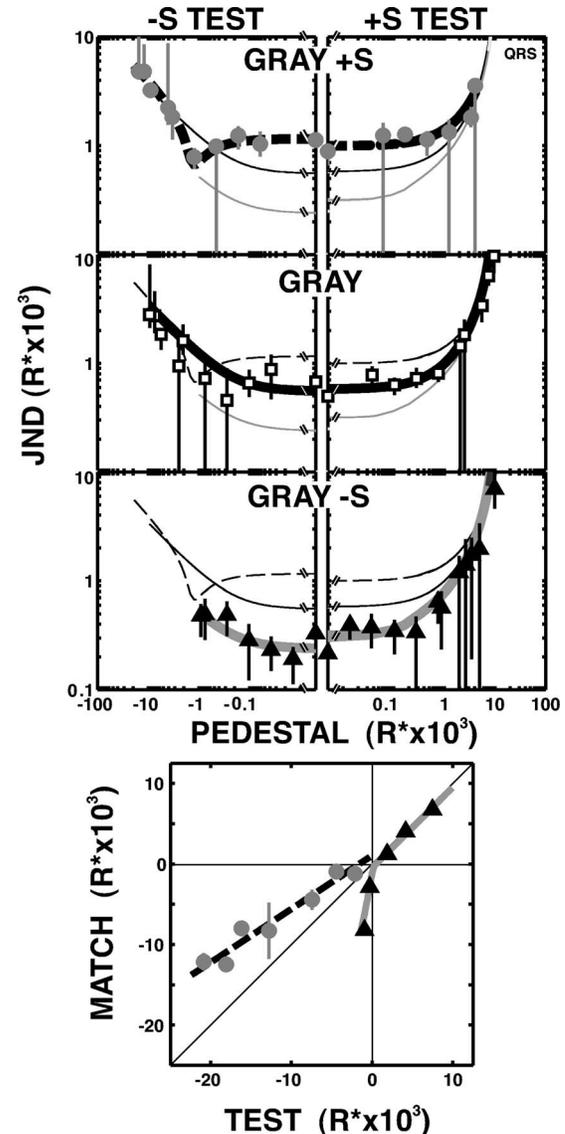


Fig. 7. QRS's results from discrimination and matching experiments for S-cone tests and adapting fields that varied in their S-cone input. Results are plotted in the same format as Figs. 4–6.

metric data with separate cumulative normal functions. Parameters were selected by a maximum-likelihood criterion. Data points shown in the top six panels are test intensities that yield 75% correct on the fitted curves (i.e., JNDs). Error bars are 95% confidence intervals determined by a bootstrapping of the maximum-likelihood parameters with trial number and test intensities set by the psychometric data.⁴² The increase in the size of the error bars with pedestal intensity may result from suboptimal sampling of points on the psychometric function by our staircase algorithm.

The curves shown in Fig. 4 are a result of selecting the parameters for the response function [Eq. (2)] that led to the best account of the data. The exact optimization criteria are described in Subsection 3.A below. The curves in these six panels are the 75% correct points (JNDs) inferred from the fitted response model (see Fig. 2). In Fig. 4, the dashed black curve, solid black curve, and solid

gray curves are JNDs against the Gray+LM, Gray, and Gray-LM backgrounds, respectively. Response model JND curves for each adapting condition are repeated in each panel to provide a common reference; the curves that correspond to the adapting condition represented by each panel are thickened.

There are three notable trends in the discrimination data. First, consistent with many studies of contrast discrimination,^{29,37,43} there is a dip in JNDs at low (sub-threshold) pedestal values and a subsequent increase in JNDs as pedestal intensity increases. Under the assumption that fixed-variance additive Gaussian noise limits discriminability, the dip reflects an accelerating nonlinearity at low contrasts^{29,37,43} while JND increases at high pedestals reflect response saturation (Fig. 2C). Second, there is a clear effect of background color on JNDs, particularly at low pedestals. JNDs for the LM tests increase as the intensity of the LM background is increased. Third, there is good agreement between JNDs determined by the conventional method of fitting cumulative normals to the psychometric data and the JNDs determined by the response model fit.

The response model parameters that determine the JND curves in the top six panels of Fig. 4 were also used to derive performance for the corresponding asymmetric matching data. The bottom panel of Fig. 4 shows the results of the asymmetric matching task for the same adapting conditions represented by the top six panels of discrimination data. The x axis represents intensity of the standard presented against either the Gray+LM or Gray-LM background. The y axis represents intensity of the match set against the gray background (circles, Gray+LM background; triangles, Gray-LM background).⁴⁴ Error bars are 1 standard error of the mean. Observers' ratings of the quality of the match (between zero and four) were used to reject trials where the observer could not find an adequate setting. Any setting rated 0 or 1 was rejected from the data set (no trials were rejected for observer JMH and 6 out of 120 trials were rejected for QRS). The dashed black curve and solid gray curve are predictions derived from the response model. The correspondence between predictions and data is excellent.

Figure 5 shows data for QRS for the same conditions shown in Fig. 4. Figures 6 and 7 show, respectively, JMH's and QRS's discrimination and matching data for S-cone tests presented against adapting backgrounds that varied only in their S-cone input. Each figure shows data from one observer in the same format as Figs. 4 and 5. The bottom panel again shows results from the asymmetric matching task for S-cone tests. The agreement between the model and data in Figs. 4-7 is good: Almost all systematic trends are well predicted by the inferred response functions. The raw psychometric and matching data used to obtain points shown in Figs. 4-7 can be downloaded at http://color.psych.upenn.edu/supplements/com_uniform/.

A. Error Trade-Off Analysis

Our broad goal is to determine whether both discrimination and appearance are mediated by the same mechanisms of adaptation. The fits in Figs. 4-7 suggest an affirmative answer for the conditions studied here. These

fits were obtained by leveraging both discrimination and appearance data simultaneously to determine the response function parameters. It remains possible, therefore, that considerably better fits to each data set could have been obtained had we fitted the two separately. Here we investigate this by fitting the response function with varying emphasis on discrimination and matching data.

Figure 8 illustrates the analysis for JMH's -LM test data presented against Gray and Gray+LM backgrounds for observer JMH. The center panel shows model fit error for discrimination and matching data. The x axis is the normalized negative log-likelihood of the model parameters given the discrimination data. The y axis is the normalized sum-of-squared error between the model and the asymmetric matching data.

The normalizing term for the negative log-likelihood of the discrimination data was the likelihood of observing the data if the probability p of every binomial variable in the psychometric data was equal to the proportion of trials the observer got correct for that test intensity, or

$$L = \prod_{i=1}^{N_I \times N_{\text{ped}}} \binom{N_i^{\text{trials}}}{N_i^{\text{correct}}} p_i^{N_i^{\text{correct}}} (1-p_i)^{N_i^{\text{trials}} - N_i^{\text{correct}}},$$

where N_i^{trials} is the number of trials for a given test intensity, N_i^{correct} is the number of 2 IFC trials the observer got correct, N_I is the number of test intensities for each pedestal, N_{ped} is the number of pedestals, and $p_i = N_i^{\text{correct}}/N_i^{\text{trials}}$. To calculate likelihoods for models of the discrimination data, p is replaced by the probability determined from the model.

The normalizing term for the squared error of the asymmetric matching data was the sum-of-squared difference between the mean of the observer's settings and their actual settings: $Err = \sum_{i=1}^{N_I} \sum_{j=1}^{N_{\text{trials}}} (\text{match}_{j,i} - \mu_i)^2$, where $\text{match}_{j,i}$ is the observer's setting on the j th trial for test intensity I , μ_i is the mean of settings for test intensity I , N_{trials} is the number of trials completed for test intensity I , and N_I is the number of test intensities. To calculate the sum of squared errors for the model fits, μ_i is replaced by the value of the match determined by the model for test intensity I .

For these error metrics, no model could do better than the values of the normalizing terms. Thus the normalized likelihoods and normalized sum of squared errors for any model will be greater than or equal to 1.

Consider the normalized likelihoods for different model fits to the discrimination data only. The data points (JNDs) in Figs. 4-7 were determined from cumulative normal fits to the discrimination data. Characterizing psychometric data with cumulative normal parameters is common and assumes, as our model does, that performance is limited by additive Gaussian noise. Such characterization of psychometric data makes no assumptions about the relationship between data across pedestals and adapting conditions. We therefore consider it a baseline against which we can compare the quality of other model fits.

The thick vertical gray line in the center panel of Fig. 8 is the normalized negative log-likelihood for the cumulative normal fits to the psychometric data (i.e., a 40-parameter model given the two parameters for each of the 20 -LM test psychometric functions from the Gray and Gray+LM backgrounds). The shaded (lighter) gray region bordering this line is the 95% confidence interval of the negative log-likelihood for these cumulative normal fits. We determined these confidence intervals by a bootstrapping procedure using the maximum-likelihood parameters (i.e., binomial probabilities for data resampling were determined by model fits, not the data trial numbers, and test intensities from the original data set. We performed maximum-likelihood fits to each of 1000 resampled data sets to determine the range of likelihoods indicated by the shaded regions. This region therefore gives us an idea of what range of likelihoods we could reasonably expect for this baseline fitting method.

The thick, vertical, black, dashed line is the normalized negative log-likelihood for the response model where all five parameters in Eq. (2) were allowed to vary to fit the discrimination data for the -LM tests in the Gray and Gray+LM backgrounds (i.e., a 10-parameter model with five for the Gray and five for the Gray+LM). Finally, the gray star with the black outline is the normalized nega-

tive log-likelihood for the response model fits with parameters M, p , and q yoked across Gray and with Gray+LM adapting conditions and parameters g and s chosen separately for each adapting condition (i.e., a 7-parameter model that we refer to as the gain-and-subtractive model of adaptation).

There is little cost in fitting the data with the response model rather than independent cumulative normals; likelihood of the data given the response model parameters for both unyoked and yoked fits falls within the confidence range of likelihoods for unconstrained cumulative normal fits. There is also little cost in yoking the M, p , and q parameters across adapting conditions (as indicated by the negligible horizontal shift in the star relative to the vertical dashed line).

The two panels on the left show the results of the yoked gain-and-subtractive model fits when the parameters were fitted to the discrimination data. The fit to the discrimination data is good, but here the model does not do well in predicting the matching data. This is reflected in the relatively high, normalized sum-of-squared error that results from these model parameters (star in the center panel).

Now consider the fits to the matching data only. Model predictions for the asymmetric matching data require pa-

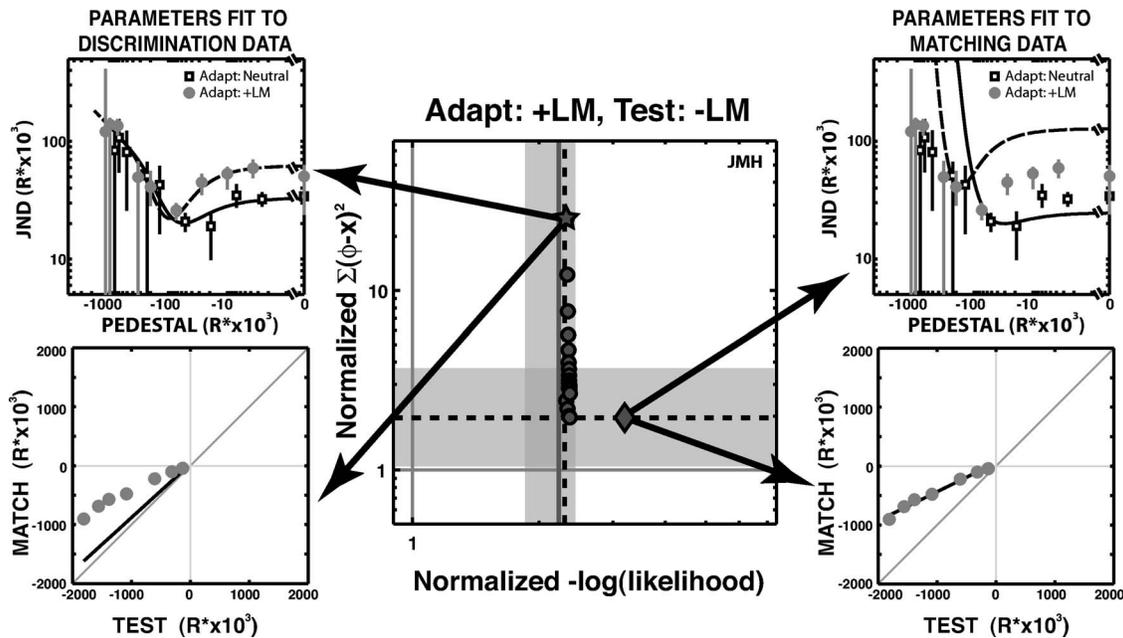


Fig. 8. Error trade-off analysis for JMH's -LM test data in the Gray and Gray+LM adapting conditions. Central panel are results of the error trade-off analysis described in the text for the gain-and-subtractive model of adaptation. The x axis is the normalized negative log likelihood (-LL) of model parameters given the full complement of discrimination data from the Gray and Gray+LM adapting conditions (we plot the negative log likelihood so small values correspond to better fits, consistent with the sum-of-squared error metric used as a criterion for the matching data). The y axis is the normalized least-squared error (LSE) of model fits to the matching data in the same adapting conditions. The two left panels show the data that underlie the analysis presented in the central panel and are replotted from Fig. 4. The two right panels show the same data. The model fits in the two left panels are fits where model parameters were determined exclusively by the discrimination data. The gray star in the central error trade-off panel is the -LL, LSE combination corresponding to these fits. The gray diamond in the central error trade-off panel is the -LL, LSE combination corresponding to these fits. The filled gray circles in the central panel are -LL, LSE combinations where both data sets were used to determine the model parameters. Each gray circle represents a -LL, LSE combination for a specific combination of weights to the matching and discrimination error. Higher points in the graph are from fits where more weight was given to maximize the likelihood of the parameters given the discrimination data than to minimize the sum-of-squared error for the model parameters given the matching data. Similarly, the more rightward points are from fits where more weight was given to minimize the sum-of-squared error for the model parameters given the matching data than to maximizing the likelihood of the parameters given the discrimination data.

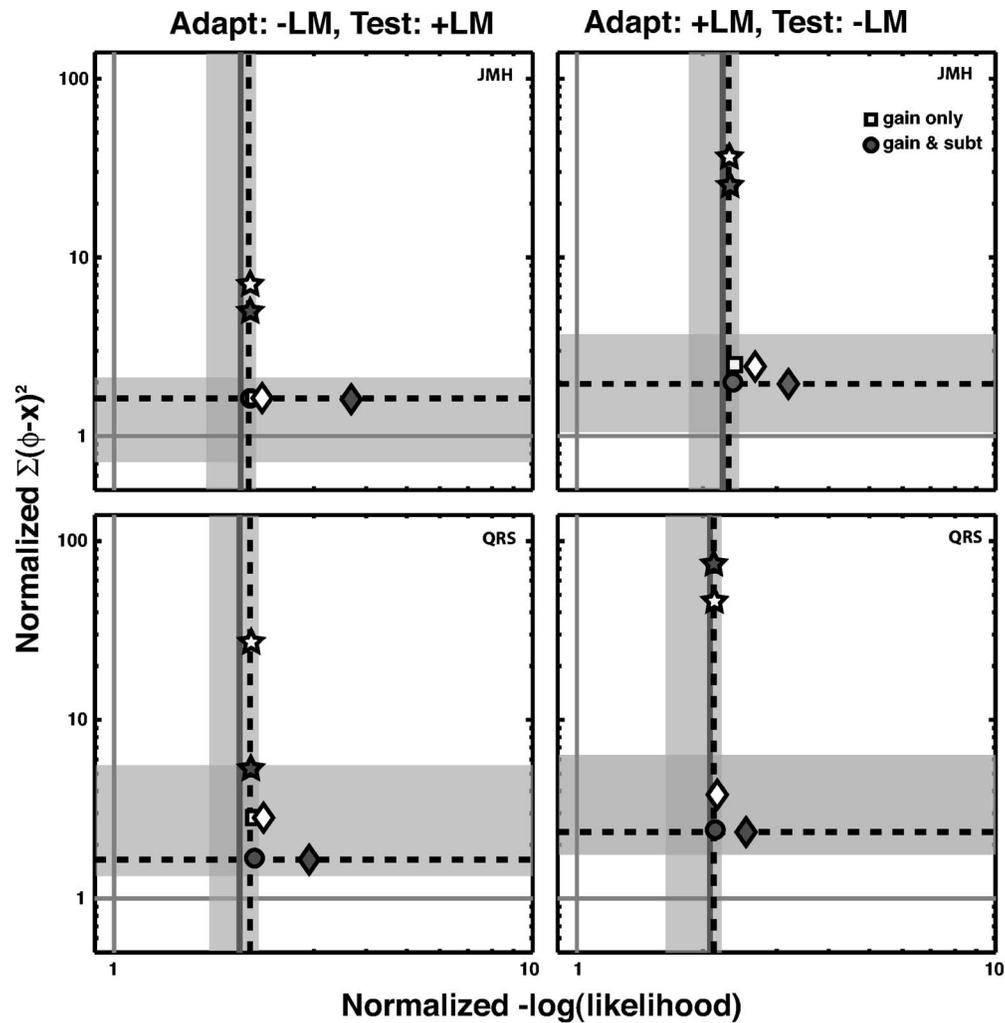


Fig. 9. Error trade-off analysis for LM tests and adapting fields that varied only in their LM component. The top two panels are, from left to right, JMH's results from +LM tests on Gray and Gray-LM adapting fields and -LM tests on Gray and Gray+LM adapting fields. The bottom two panels are from the same conditions for QRS. Plotting conventions are the same as those for the central panel in Fig. 8. We have included results of the error trade-off analysis for the gain-only (open white symbols) model as well as the gain-and-subtractive (gray symbols) model.

rameters for both backgrounds. When no parameters are yoked, there are 10 model parameters for every adapting condition (five for Gray, five for Gray \pm X). These parameters are vastly underconstrained by the matching data (bottom panels of Figs. 4–7). However, the error for fits obtained using this many parameters provides a good baseline. The horizontal black dashed line in the center panel of Fig. 8 is the normalized sum of squared error for this 10-parameter fit to the matching data. We generated simulated data from these model parameters to determine the range of sum-of-squared errors we could expect if those model parameters accurately characterized the underlying response function. The gray shaded region around the horizontal dashed line is the 95% confidence interval for the sum-of-squared error derived from these simulations. The gray diamond with black outline represents errors from fits of the gain-and-subtractive adaptation model to the matching data (when M , p , and q are yoked across adapting conditions). Again there is little cost to yoking parameters M , p , and q across adapting

conditions. The two panels on the right show the results of these yoked gain-and-subtractive model fits. With the parameters selected on the basis of the matching data, the model provides a good fit to the matching data and a poor fit to the discrimination data.

Finally consider the error trade-offs possible when data from both tasks are used to constrain the model fits. The black-outlined gray star in Fig. 8 represents the lower bound for the negative likelihoods and the upper bound for the squared error for the gain-and-subtractive adaptation model. Similarly, the black-outlined gray diamond represents the lower bound for the matching squared-error and the upper bound for the negative likelihoods for the gain-and-subtractive adaptation model. If performance in matching and discrimination tasks is controlled by common mechanisms of adaptation then there should be little cost in terms of the likelihoods of the discrimination data and squared errors for the matching data when both data sets are used to constrain the model parameters. That is, the likelihoods and squared errors for the

weighted fits should follow the contour of the horizontal and vertical dashed black lines with some points close to the intersection of these two lines. The black-outlined gray circles are likelihoods and squared errors for weighted fits with the gain-and-subtractive adaptation model. Using both data sets, we found model parameters that gave errors close to both lower bounds. Further the likelihood and squared error for these fits fell within the intersection of the confidence intervals established for each error metric. These results establish, for this data set, that both matching and discrimination data can be accounted for by common mechanisms of adaptation.

We also examined the quality of the fits by eye, comparing fits when both data sets were factored into the parameter selection and fits where only one or the other data set was employed. Differences between the fits with errors in the lower left corner of the error trade-off plots and those determined by each data set in isolation were vanishingly small (i.e., the fits determined by appearance and discrimination data independently essentially looked the same as the yoked fits shown in Figs. 4–7).

Figures 9 and 10 show the results for all the data sets that could be subjected to this error trade-off analysis.

The conventions in Figs. 9 and 10 are the same as those established in Fig. 8 with two exceptions. First, we added results of the analysis for a simpler model of adaptation that allows only the gain [g in Eq. (2)] to vary across contexts. For this gain-only model (shown as open symbols), the subtractive term [s in Eq. (2)] was set to zero. Second, we plot only the extremes of the fits (fits to either the discrimination or matching data) and the one result of the weighted fits that came closest to the lower bounds established by the fits to each data set in isolation. The top two panels in Fig. 9 are results of this error-trade-off analysis for JMH for $-LM$ tests presented against the Gray+LM background (left panel) and $+LM$ tests presented on the Gray-LM background (right panel, same as Fig. 8). The bottom two panels are results from QRS for the same conditions. The white stars are from fits to the discrimination data of the gain-only adaptation and the gray stars are from the gain-and-subtractive model of adaptation. The white diamonds are from fits to the matching data of the gain-only adaptation, and the gray diamonds are from the gain-and-subtractive model of adaptation. The white square and gray circle are the best weighed fits for the gain-only and gain-and-subtractive adaptation models,

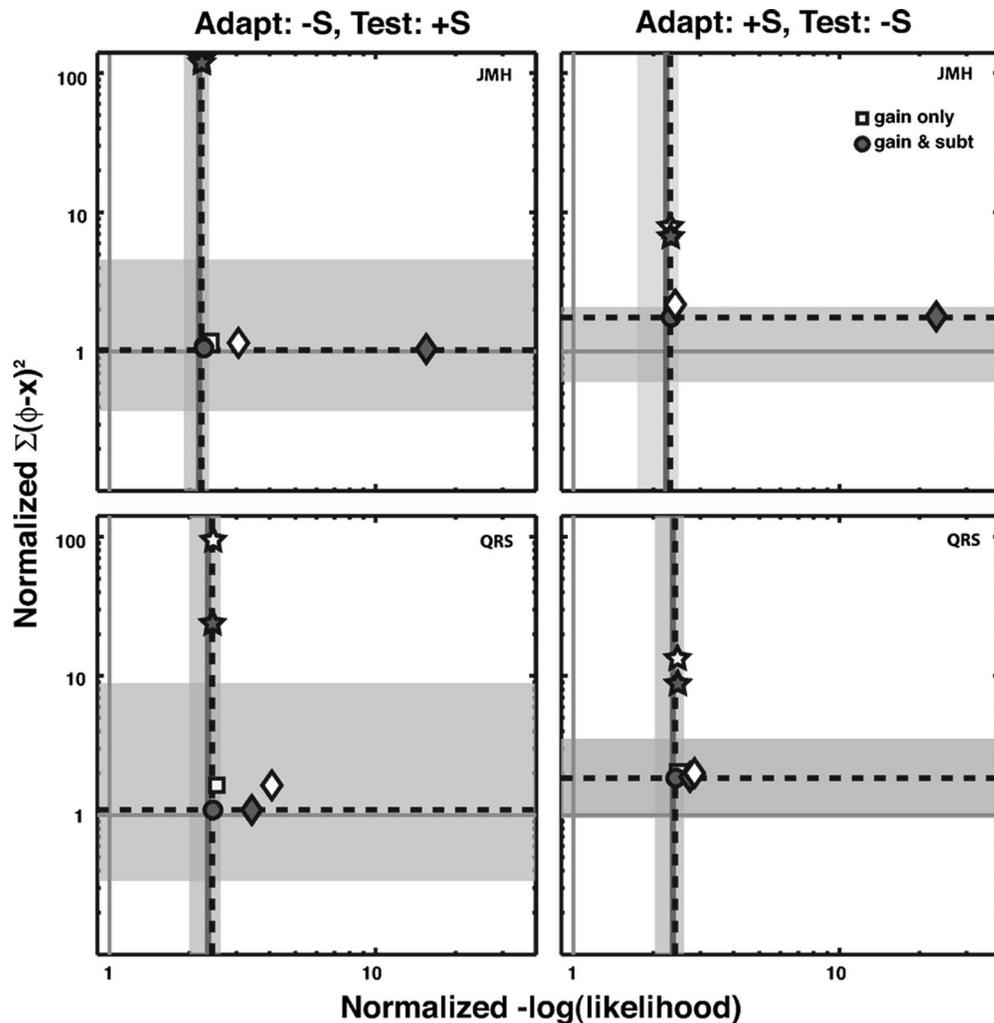


Fig. 10. Same as Fig. 8 except for S-cone tests and adapting fields that varied only in their S-cone component.

respectively. Figure 10 shows results for the S-cone tests and adapting conditions using the same conventions as Fig. 9. The parameters that determined the error values for the gray circles (gain-and-subtractive adaptation model) in Figs. 9 and 10 are the parameters used in the model fits shown in Figs. 4–7.

Three important trends are revealed through the analysis depicted in Figs. 8 and 9: (1) in all cases, the weighted fits come very close to the lower bounds of the error established by considering each data set in isolation; (2) there is essentially no cost in terms of the model error to yoking parameters M , p , and q across the relevant adapting conditions for either the discrimination or the matching data; and (3) there are several cases where the gain-and-subtractive model provides a clear reduction in the model error relative to the gain-only model (though this error reduction is observed primarily in the matching data).

Our error tradeoff analysis confirms the impression conveyed by Figs. 3–6: Effects of adaptation on both discrimination and appearance can be explained by a change in response function common to both. Before discussing the results and their implications further, we briefly present two control experiments.

4. CONTROL EXPERIMENT 1: DISCRIMINATION PERFORMANCE ON SPLIT FIELDS

In the asymmetric matching experiment, test spots were presented against a split field, one on either side of the color discontinuity. To obtain the discrimination data, on the other hand, tests were presented against a field that had the same color across the entire display. We used a completely uniform field in the discrimination task because, with this arrangement, each experimental session yielded twice as much data for the given adapting condition. The control data presented here indicate that there were no unexpected interactions between the two sides of the split-field context.

Methods were identical to the main discrimination experiment except that tests and pedestals were presented on a split field. One half of the field was Gray and the other half was Gray-LM. Pedestals and tests were +LM. One pedestal intensity was tested in each experimental session. One observer was tested with four staircases for each background-pedestal combination.

For our purposes, it is sufficient to show that the ratio of JNDs between the split versus uniform field adapting conditions is constant. In Appendix B we show that such a uniform shift would not have affected any fitting parameters in Eq. (2) except for M , and that a change in M does not affect the predictions for the asymmetric matches.

The ratio of JNDs from the first experiment and this control experiment are shown in Fig. 11. The x axis is the expected number of isomerizations for the +LM pedestal. The y axis is the ratio $JND_{\text{uniform}}/JND_{\text{split}}$. Different symbols denote different adapting conditions: Open circles are from the Gray background and filled circles are data from the Gray-LM background. Error bars are 95% confidence intervals determined by a bootstrapping procedure. There is no systematic difference in the JND ratios for these adapting conditions.

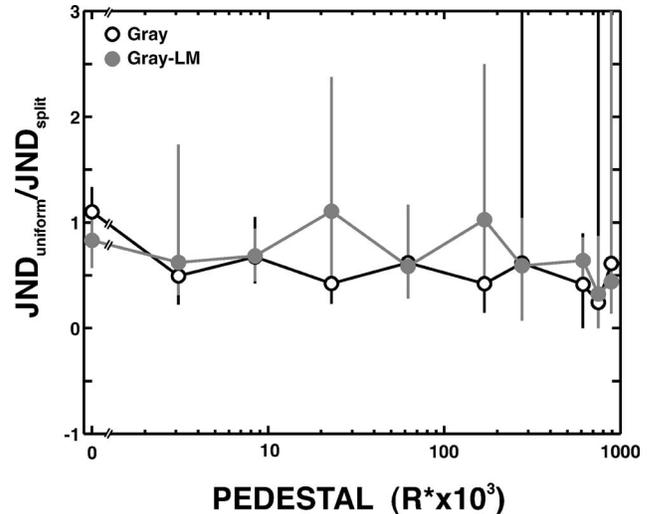


Fig. 11. JMH's JND ratios for the split field versus uniform field conditions for LM increments presented against Gray and Gray-LM backgrounds. The x axis is isomerizations of the pedestal (same as lower right and middle right panels in Fig. 4). The y axis is the JND ratio for discrimination data collected on a uniform and split field. Open circles are JND ratios from the Gray adapting field and filled gray circles are ratios from the Gray-LM adapting field. Error bars are 95% confidence intervals determined by a bootstrap analysis.

5. CONTROL EXPERIMENT 2: TESTING THE VON KRIES HYPOTHESIS

In the main experiment, tests and adapting field shifts were in the same color direction. We also examined effects of LM background shifts on the discriminability and appearance of S tests, and vice versa. This tests the von Kries hypothesis that each cone type adapts independently for both appearance and discrimination.

Methods were identical to those of experiment 1. We measured detection thresholds and asymmetric matches for S and LM tests against Gray \pm LM and Gray \pm S adapting fields, respectively. That is, S tests were presented on backgrounds that varied in their L- and M-cone coordinates and LM tests were presented on backgrounds that varied only in the S-cone coordinates. The same two observers who participated in the first experiment, along with one other observer, participated in this experiment.

Detection thresholds and asymmetric matches for one observer are shown in Fig. 12. The top left panel shows detection thresholds for \pm LM tests as a function of S background intensity. The filled black circles are test intensities that gave 75% correct for a cumulative normal fit to the data by a maximum-likelihood criterion. The solid gray lines provide a reference for no changes in thresholds as a function of background intensity. Similarly, the filled black circles in the top right panel are absolute thresholds for \pm S tests as a function of LM background intensity.

The bottom two panels of Fig. 12 are results from the asymmetric matching experiment: In the left panel are data for LM tests when the S-cone component of the background was varied, and in the right panel are data for S-cone tests when the LM component of the background was varied. The filled black squares are a symmetric matching condition (both the fixed test and adjustable

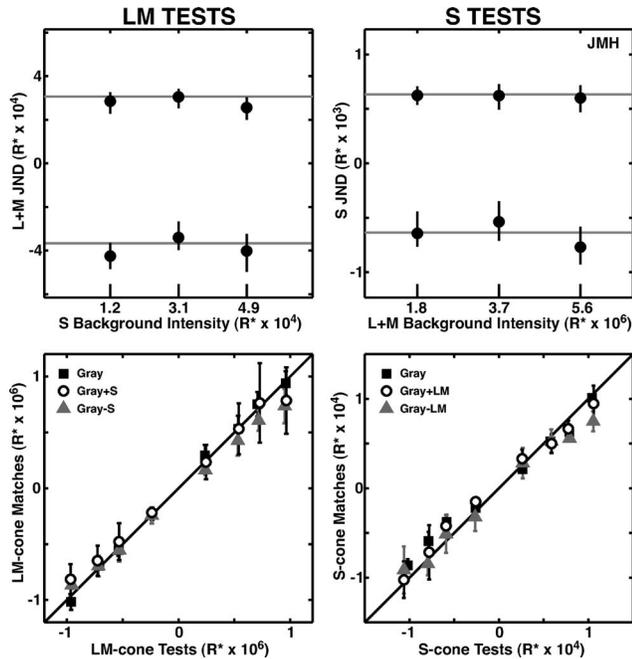


Fig. 12. JMH's detection and matching results for LM-cone tests presented on adapting fields that varied only in their S component and S-cone tests presented on adapting fields that varied only in their LM component. The top two panels are detection thresholds plotted as a function of background intensity. The bottom two panels are results of the asymmetric matching task. The two left panels are results from LM-cone tests presented on backgrounds that varied only in the S-cone components. The two right panels are results from S-cone tests presented on backgrounds that varied only in their LM-cone component. The x axis in the top left panel is the expected number of S-cone isomerizations from the background light for the same area and temporal interval as the test stimuli, and the y axis is the expected number of isomerizations for an LM-cone test. Similarly, the x axis in the top right panel represents background LM-cone isomerizations and the y axis S-cone test isomerizations. Data points in these top two panels are the 75% thresholds determined by fitting the detection data with a cumulative normal. The x axis in the bottom left panel is the expected number of isomerizations for fixed LM-cone tests presented on either the Gray+S, Gray, or Gray-S adapting fields. The y axis is S-cone isomerizations for JMH's match settings against the gray background. Filled squares are from the symmetric matching conditions (where both the fixed test and adjustable match were presented on the Gray background). Open circles and filled diamonds are conditions where the fixed tests were presented against the Gray+LM and Gray-LM conditions, respectively. Error bars are standard errors of the mean. The convention for the bottom right panel is the same as for the bottom left panel except that the axes correspond to the expected isomerizations of S-cone tests.

match were presented against the Gray background) and provide an indication of any bias in the observer's settings. The slight systematic trend for the +LM matches and +S matches to be lower than the baseline settings in the Gray-S and Gray-LM conditions, respectively, was not evident for two other observers.

Results from both the asymmetric matching and detection task are generally consistent with the hypothesis that, for uniform fields, cones adapt independently. Results from two other observers (not presented) were similar. The lack of interaction between cones observed in both discrimination and appearance judgments is consistent with the results of several previous studies.⁴⁵⁻⁴⁸

6. DISCUSSION

We have tested the assumption that effects of background chromaticity and luminance on color discriminability and appearance are mediated by mechanisms common to both judgments. This common mechanism hypothesis holds for the conditions we examined. Our general conclusion agrees with that of earlier authors who have studied adaptation of discrimination and appearance with respect to changes in uniform backgrounds.^{24,34,49-52} Our work extends previous studies in several ways.

First, we measured full tvi curves and used these to learn how adaptation affects discriminability across a wide range of stimulus intensities. In this regard, we share much with Heinemann.³⁴ Most earlier work, however, characterized adaptation with changes in detection threshold and relied on gain-control models of adaptation to extrapolate predictions to the intensity range where appearance was studied. Use of detection thresholds is adequate only for stimulus conditions where the gain-control model is valid. Some of our conditions required inclusion of a subtractive term in the adaptation model. Had we relied only on detection thresholds, we would have drawn erroneous conclusions about the common mechanism hypothesis. We expect this consideration to become increasingly important as we extend our work to other adapting contexts that include contrast and more complex spatial structure. There is good evidence that richer models of adaptation are required to account for adaptation effects for such contexts.^{22,36,53}

A second novel feature of our work is the error tradeoff analysis presented in Figs. 8-10. This analysis shows that fitting a model to discrimination data alone leads to poor predictions of appearance and vice versa. Such a result could arise either because the common mechanism hypothesis is false, or because measurement variability propagates to the model fits in a manner that leads to poor extrapolation across tasks. These two possibilities are distinguished by the error tradeoff analysis. For adaptation to uniform fields, the case studied here, the common mechanism hypothesis holds. Under other conditions we might expect the common mechanism hypothesis to fail. For example, Neger *et al.*⁵⁴ found that "filling-in" of retinally stabilized images affected appearance judgments but not detectability. While this result suggests different mechanisms mediating detection and appearance, such a conclusion would be premature for two reasons. First, our results show the importance of measuring the tvi function and not just detection thresholds when comparing threshold and suprathreshold performance. Second, the common mechanism hypothesis should be subjected to the kind of statistical test provided by the error tradeoff analysis presented here. Further, error tradeoff is important to bear in mind when evaluating conclusions drawn from comparisons of data that were collected and analyzed in different laboratories.

A third important feature of our work was the attempt to match as carefully as possible the stimulus conditions used in the discrimination and appearance experiments. Thus the same observers viewed spots flashed with the same temporal and spatial profiles. Careful matching of stimulus conditions rules out the possibility that differences between discrimination and appearance occur sim-

ply because the visual system is in a different state of adaptation across the two paradigms. The one major difference that persisted in our experiments was the use of split backgrounds in the appearance experiments and uniform backgrounds in the discrimination experiments. Control measurements, however, indicated that this difference was not of consequence for our comparisons.

At the core of our modeling is an inferred response function common to both discrimination and appearance.^{34,49} This function explains discrimination (tvi) data through the linking hypothesis that sensitivity is proportional to its slope.⁵⁵ The same function explains appearance through the linking hypothesis that two stimuli appear the same when they lead to the same response. We chose a particular parametric form for the response function [Eq. (2)]. This function is consistent with previous studies of both discrimination^{29–32} and appearance.^{51,57} A parametric response function sets the form of the common mechanism hypothesis and provides a way to aggregate data collected at discrete intensity levels. For the conditions studied here, Eq. (2) allowed an excellent description of both tvi and matching data. Under the assumption that the common mechanism hypothesis holds, the parameters of the response function are better determined using both discrimination and appearance data than by using either data set alone.

In addition to testing the common mechanism hypothesis, our data also speak to models of adaptation. The most general model we considered was one in which all of the parameters of the response function [Eq. (2)] were allowed to vary across adapting conditions. More restrictive models yoke some of the model parameters across conditions and thus account for adaptation through changes of a subset of the parameters. We considered two specific subset models, one in which only the gain (parameter g) was allowed to change and one in which subtraction was also allowed (both parameters g and s allowed to vary). We found cases where the gain-only model failed to account for the data while the gain-and-subtractive model worked well. Previous authors^{25,30,32,58,59} have also argued that models of adaptation must include a subtractive term⁶¹ (but see Ref. 60).

There are obvious practical reasons why testing the

common mechanism hypothesis is important: (1) Modeling adaptation is simplified for conditions where data of both discrimination and appearance experiments reveal the same mechanisms, and (2) predictions about the behavior of the neural substrate from the psychophysics are more straightforward under these conditions. More broadly, however, the status of this hypothesis is of interest as we consider the utility of color vision. Cases where the hypothesis fails are cases where separate mechanisms mediate the role of color in scene segmentation and object identification. Such cases would provide important clues to differences in the information processing requirements of these two tasks.

The data presented in this paper do not reveal failures of the common mechanism hypothesis, but represent only a limited set of stimulus conditions: We consider contexts consisting only of uniform backgrounds. In addition, we choose the color directions of the tests to maximize the probability that they would isolate individual mechanisms, as indicated by a large literature on how the visual system processes color.^{3,11,16,22} Clearly it will be of interest to determine how the common mechanism hypothesis generalizes. Of particular interest to us are conditions that include contrast adaptation,^{17,19,62,63} intermediate test color directions, and spatial structure in the contexts.⁶⁴ We believe that the basic methods and analysis presented here will allow sharp tests of the hypothesis for this wider range of conditions.

APPENDIX A: ESTIMATING CONE PHOTOPIGMENT ISOMERIZATIONS

Table 2 lists the conventions we used to estimate the isomerizations for the data presented in this paper. The software used to perform isomerization rate calculations is available at www.psychtoolbox.org. Test intensities were presented in terms of the total expected number of isomerizations independent of the background. To calculate the total expected number of isomerizations, isomerization rates were integrated for the spatial and temporal profiles of the test (Fig. 3). Specifically, $R_{\text{total}}^* = \sum_x \sum_y \sum_t R_{\text{rate}}^*(x, y, t)$, where f is the function defining the

Table 2. Conventions Used to Estimate Isomerizations

Property	Value	Source
Posterior nodal point distance	16.1 mm	Ref. 65
Pupil diameter	3.52 mm ^a	Ref. 66
Lens density	—	Ref. 67
Macular pigment optical density	—	Ref. 68; scaled by estimates of relative density at 3 deg eccentricity from Ref. 69
Photoreceptor density	1.6×10^4 per mm ²	Ref. 70
L cone: M-cone ratio	2:1	—
(L+M) cone: S-cone ratio	20:1	—
Inner segment length	2.9 mm	Ref. 65
Outer segment diameter	33 μ m	Ref. 65
Photopigment axial density	0.5	Ref. 65
Photopigment spectral sensitivity	—	Ref. 40
Quantal efficiency	2/3	Ref. 65

^aFor the purposes of calculating isomerizations, pupil diameter was fixed to its expected value for the Gray background.

stimulus, and x and y correspond to the assumed cone spacing. As we noted in Section 2 the conversion from isomerization rates (per cone per second) to isomerization totals incorporates differences in the cone ratios. Thus if there are more L than M cones, there will be proportionally more total L-cone isomerizations when the isomerization rates are the same. The cone densities and ratios used to calculate these totals are included in Table 2.

APPENDIX B: EFFECTS OF UNIFORM SHIFTS IN JNDS ON MODEL PARAMETERS

Here we show that effects on JNDs we observed in the split field versus uniform field conditions will not substantively affect model parameter estimation. Recall that JNDs are inversely proportional to the slope of the response function (i.e., $JND \propto 1/R'$). How would a shift in JNDs by a factor common to all pedestals and adapting conditions, as we observed, affect parameter estimation? First we note that

$$JND \propto \frac{1}{R'} = \frac{1}{M g \left\{ (gI + S)^{p-1} [(gI + S)^q (p - q) + p] \right\}^2}$$

Thus any multiplicative shift in JNDs common across all conditions would be reflected by a common change in M . What is important is, such a change does not affect the match predictions. Recall that we assume matches are made when the responses of the relevant mechanisms are equated, that is, when

$$R_i^A = R_i^B$$

$$M_A \frac{(g_A I + s_A)^p}{(g_A I + s_A)^q + 1} = M_B \frac{(g_B I + s_B)^p}{(g_B I + s_B)^q + 1}$$

Because both sides of this equation can be divided by a common change in M_A and M_B without affecting the equality, the magnitude of this change does not influence match predictions.

ACKNOWLEDGMENTS

We thank Jack Nachmias for helpful comments and discussion. This research was supported by NIH through grant R01 EY10016.

Corresponding author J. M. Hillis can be reached by e-mail at jmhillis@psych.upenn.edu.

REFERENCES AND NOTES

1. W. S. Stiles, "Color vision: the approach through increment threshold sensitivity," *Proc. Natl. Acad. Sci. U.S.A.* **45**, 100–114 (1959).
2. G. Wyszecki and W. S. Stiles, *Color Science—Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. (Wiley, 1982).
3. R. T. Eskew, J. S. McLellan, and F. Giulian, "Chromatic detection and discrimination," in *Color Vision: From Molecular Genetics to Perception*, K. Gegenfurtner and L. T. Sharpe, eds. (Cambridge U. Press, 1999), pp. 345–368.
4. D. C. Hood and M. A. Finkelstein, "Sensitivity to Light," in *Handbook of Perception and Human Performance: Sensory Processes and Perception*, K. R. Boff, L. Kaufman, and J. P. Thomas, eds. (Wiley, 1986).
5. P. Lennie and M. D'Zmura, "Mechanisms of color vision," *Crit. Rev. Neurobiol.* **3**, 333–400 (1988).
6. P. K. Kaiser and R. M. Boynton, *Human Color Vision*, 2nd ed. (Optical Society of America, 1996).
7. R. W. Burnham, R. M. Evans, and S. M. Newhall, "Prediction of color appearance with different adaptation illuminations," *J. Opt. Soc. Am.* **47**, 35–42 (1957).
8. S. K. Shevell, "Color appearance," in *The Science of Color*, 2nd ed., S. K. Shevell, ed. (Optical Society of America, 2003).
9. G. Wyszecki, "Color appearance," in *Handbook of Perception and Human Performance: Sensory Processes and Perception*, K. R. Boff, L. Kaufman, and J. P. Thomas, eds. (Wiley, 1986), pp. 9.1–9.56.
10. Q. Zaidi, "Color and brightness inductions: from Mach bands to three-dimensional configurations," in *Color Vision: From Genes to Perception*, K. R. Gegenfurtner and L. T. Sharpe, eds. (Cambridge U. Press, 1999), pp. 317–344.
11. D. H. Brainard, "Color Constancy," in *The Visual Neurosciences*, L. Chalupa and J. Werner, eds. (MIT Press, 2004), pp. 948–961.
12. D. H. Foster, "Does colour constancy exist?" *Trends Cogn. Sci.* **7**, 493–443 (2003).
13. L. M. Hurvich and D. Jameson, "An opponent-process theory of color vision," *Psychol. Rev.* **64**, 384–404 (1957).
14. W. S. Stiles, "Mechanism concepts in colour theory," *J. Colour Group* **11**, 106–123 (1967).
15. D. H. Brainard, "Color vision theory," in *International Encyclopedia of the Social and Behavioral Sciences*, N. J. Smelser and B. P. Baltas, eds. (Elsevier, 2001), pp. 2256–2263.
16. M. A. Webster, "Human colour perception and its adaptation," *Network Comput. Neural Syst.* **7**, 587–634 (1996).
17. J. Krauskopf, D. R. Williams, and D. W. Heeley, "Cardinal directions of color space," *Vision Res.* **22**, 1123–1131 (1982).
18. J. Krauskopf, D. R. Williams, M. B. Mandler, and A. M. Brown, "Higher order color mechanisms," *Vision Res.* **26**, 23–32 (1986).
19. M. A. Webster and J. D. Mollon, "Changes in colour appearance following post-receptoral adaptation," *Nature (London)* **349**, 235–238 (1991).
20. C. C. Chen, J. M. Foley, and D. H. Brainard, "Detection of chromoluminance patterns on chromoluminance pedestals II: Model," *Vision Res.* **40**, 789–803 (2000).
21. P. B. Delahunt and D. H. Brainard, "Control of chromatic adaptation: signals from separate cone classes interact," *Vision Res.* **40**, 2885–2903 (2000).
22. M. D'Zmura and B. Singer, "Contrast gain control," in *Color Vision: From Molecular Genetics to Perception*, K. Gegenfurtner and L. T. Sharpe, eds. (Cambridge U. Press, 1999), pp. 369–385.
23. J. M. Loomis and T. Berger, "Effects of chromatic adaptation on color discrimination and color appearance," *Vision Res.* **19**, 891–901 (1979).
24. E. N. Pugh and J. Larimer, "Test of the identity of the site of blue/yellow hue cancellation and the site of chromatic antagonism in the π_1 pathway," *Vision Res.* **20**, 779–788 (1980).
25. J. Walraven, "Perceived color under conditions of chromatic adaptation: Evidence for gain control by π -mechanisms," *Vision Res.* **21**, 611–620 (1981).
26. G. T. Fechner, *Elements of Psychophysics*, Henry Holt Edition in Psychology (Holt, Rinehart & Winston, 1966).
27. D. H. Krantz, "Integration of just-noticeable differences," *J. Math. Psychol.* **8**, 591–599 (1971).
28. R. D. Luce and W. Edwards, "The derivation of subjective scales from just noticeable differences," *Psychol. Rev.* **65**, 222–236 (1958).
29. G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Am.* **70**, 1458–1471 (1980).

30. E. H. Adelson, "Saturation and adaptation in the rod system," *Vision Res.* **22**, 1299–1312 (1982).
31. J. M. Foley, "Human luminance pattern-vision mechanisms: masking experiments require a new model," *J. Opt. Soc. Am. A* **11**, 1710–1719 (1994).
32. W. S. Geisler, "Mechanisms of visual sensitivity: backgrounds and early dark adaptation," *Vision Res.* **23**, 1423–1432 (1983).
33. D. Laming, *The Measurement of Sensation*, 1st ed., Vol. 30 of Oxford Psychology Series (Oxford U. Press, 1997), p. 262.
34. E. G. Heinemann, "The relation of apparent brightness to the threshold for differences in luminance," *J. Exp. Psychol.* **61**, 389–399 (1961).
35. D. H. Brainard, "Cone contrast and opponent modulation color spaces," in *Human Color Vision*, 2nd ed., P. K. Kaiser and R. M. Boynton, eds. (Optical Society of America, 1996), pp. 563–579.
36. C. C. Chen, J. M. Foley, and D. H. Brainard, "Detection of chromoluminance patterns on chromoluminance pedestals I: Threshold measurements," *Vision Res.* **40**, 773–788 (2000).
37. J. Nachmias and R. V. Sansbury, "Grating contrast: discrimination may be better than detection," *Vision Res.* **14**, 1039–1042 (1974).
38. K. I. Naka and W. A. Rushton, "S-potentials from colour units in the retina of fish (Cyprinidae)," *J. Physiol. (London)* **185**, 536–555 (1966).
39. M. J. Valetton and D. v. Norren, "Light adaptation of primate cones: an analysis based on extracellular data," *Vision Res.* **23**, 1539–1547 (1983).
40. V. Smith and J. Pokorny, "Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm," *Vision Res.* **15**, 161–171 (1975).
41. D. H. Brainard, D. G. Pelli, and T. Robson, "Display characterization," in *Encyclopedia of Imaging Science and Technology*, J. Hornak, ed. (Wiley, 2002), pp. 72–188.
42. F. W. Wichmann and N. J. Hill, "The psychometric function: II. Bootstrap-based confidence intervals and sampling," *Percept. Psychophys.* **63**, 1314–1329 (2001).
43. J. M. Foley and G. E. Legge, "Contrast detection and near-threshold discrimination in human vision," *Vision Res.* **21**, 1041–1053 (1981).
44. Note that matches varied along only the color direction of the test for this and other conditions examined in this paper. Thus, plotting intensity is sufficient.
45. J. Krauskopf and K. Gegenfurtner, "Color discrimination and adaptation," *Vision Res.* **32**, 2165–2175 (1992).
46. E. J. Chichilnisky and B. A. Wandell, "Photoreceptor sensitivity changes explain color appearance shifts induced by large uniform backgrounds in dichoptic matching," *Vision Res.* **35**, 239–254 (1995).
47. S. M. Wuerger, "Color appearance changes resulting from iso-luminant chromatic adaptation," *Vision Res.* **36**, 3107–3118 (1996).
48. A. R. Wade and B. A. Wandell, "Chromatic light adaptation measured using functional magnetic resonance imaging," *J. Neurosci.* **22**, 8148–8157 (2002).
49. P. Whittle and P. D. C. Challands, "The effect of background luminance on the brightness of flashes," *Vision Res.* **9**, 1095–1110 (1969).
50. P. Whittle, "Increments and decrements: luminance discrimination," *Vision Res.* **26**, 1677–1691 (1986).
51. P. Whittle, "Brightness, discriminability and the 'crispness effect'," *Vision Res.* **32**, 1493–1507 (1992).
52. O. Rinner and K. R. Gegenfurtner, "Time course of chromatic adaptation for color appearance and discrimination," *Vision Res.* **40**, 1813–1826 (2000).
53. A. G. Shapiro and Q. Zaidi, "The effects of prolonged temporal modulation on the differential response of color mechanisms," *Vision Res.* **32**, 2065–2075 (1992).
54. J. L. Nerger, T. P. Piantanida, and J. Larimer, "Color appearance of filled-in backgrounds affects hue cancellation, but not detection thresholds," *Vision Res.* **33**, 165–172 (1993).
55. This linking hypothesis is equivalent to assuming that performance is limited by additive noise of fixed variance following the nonlinearity. An alternative model includes signal-dependent noise. For a subset of the data, we examined parametric fits to the discrimination and appearance data that included signal-dependent noise (with variance proportional to the expected response). The precise shape of the inferred nonlinearity is different for a signal-dependent noise model, but the quality of the fits to the discrimination and appearance data was not affected (see Ref. 56 for an analysis of the power of discrimination data to test these alternative noise models).
56. G. E. Legge, D. Kerstein, and A. E. Burgess, "Contrast discrimination in noise," *J. Opt. Soc. Am. A* **4**, 391–404 (1987).
57. H. Takasaki, "Lightness change of grays induced by change in reflectance of gray background," *J. Opt. Soc. Am.* **56**, 504–509 (1966).
58. D. Jameson and L. M. Hurvich, "Color adaptation: Sensitivity, contrast, and afterimages," in *Handbook of Sensory Physiology*, L. M. Hurvich and D. Jameson, eds. (Springer, 1972), pp. 568–581.
59. J. Walraven, "Discounting the background: The missing link in the explanation of chromatic induction," *Vision Res.* **16**, 289–295 (1976).
60. S. K. Shevell, "Unambiguous evidence for the additive effect in chromatic adaptation," *Vision Res.* **20**, 637–639 (1980).
61. In the literature on subtractive adaptation, some authors apply the subtractive term to the incremental/decremental stimulus, as we have done here. Others apply it to the sum of the background and the increment/decrement. These two formulations are equivalent except in the interpretation of what a subtractive term of zero means.
62. Q. Zaidi and A. G. Shapiro, "Adaptive orthogonalization of opponent-color signals," *Biol. Cybern.* **69**, 415–428 (1993).
63. R. O. Brown and D. I. A. MacLeod, "Color appearance depends on the variance of surround colors," *Curr. Biol.* **7**, 844–849 (1997).
64. J. M. Hillis and D. H. Brainard, "A shadowy dissociation between discriminability and identity," *J. Vision* **4**, 56a (2004).
65. R. W. Rodieck, *The First Steps In Seeing* (Sinauer, 1998).
66. J. Pokorny and V. C. Smith, "How much light enters the retina?" in *Colour Vision Deficiencies XIII*, C. R. Cavonius, ed. (Kluwer Academic, 1997), pp. 491–511.
67. A. Stockman, L. T. Sharpe, and C. C. Fach, "The spectral sensitivity of the human short-wavelength cones," *Vision Res.* **39**, 2901–2927 (1999).
68. R. A. Bone, J. T. Landrum, and A. Cains, "Optical density spectra of the macular pigment in vivo and in vitro," *Vision Res.* **32**, 105–110 (1992).
69. A. G. Robson, J. D. Moreland, D. Pauleikhoff, T. Morrissey, G. E. Holder, F. W. Fitzke, A. C. Bird, and F. J. van Kuijk, "Macular pigment density and distribution: comparison of fundus autofluorescence with minimum motion photometry," *Vision Res.* **43**, 1765–1775 (2003).
70. C. A. Curcio, K. R. Sloan, Jr., O. Packer, A. E. Hendrickson, and R. E. Kalina, "Distribution of cones in human and monkey retina: individual variability and radial symmetry," *Science* **236**, 597–582 (1987).