

# Luminosity thresholds: effects of test chromaticity and ambient illumination

Jon M. Speigle and David H. Brainard

*Department of Psychology, University of California, Santa Barbara, California 93106*

Received June 7, 1995; revised manuscript received September 14, 1995; accepted September 22, 1995

Color constancy is often modeled on the assumption that color appearance in natural scenes is a function of the visual system's estimates of surface reflectance. Some stimuli, however, do not look like illuminated surfaces. Instead, they appear to be self-luminous. We hypothesized that the appearance of luminosity occurs when the visual system estimates a reflectance spectrum that is outside the gamut of physically realizable surfaces. To test this idea, we measured luminosity thresholds as a function of stimulus chromaticity and illuminant spectral power distribution. Observers adjusted the luminance of a test patch until it just appeared self-luminous. The test patch was spot illuminated by a computer-controlled projection colorimeter viewed in an experimental room lit diffusely by computer-controlled theater lamps. Luminosity thresholds were determined for a number of test patch chromaticities under five experimental illuminants. The luminosity thresholds define a surface in color space. The shape of this surface depends on the illuminant. We were able to describe much of the luminosity threshold variation with a simple model whose parameters define an equivalent illuminant. In the context of our model, the equivalent illuminant may be interpreted as the illuminant perceived by the observer. As part of our model calculations we generalized the classic notion of optimal stimuli by incorporating linear-model constraints. Given the equivalent illuminant, the model predicts that a patch will appear self-luminous when it is not consistent with any physically realizable surface seen under that illuminant. In addition, we show that much of the variation of the equivalent illuminant with the physical illuminant can be modeled with a simple linearity principle. The fact that our model provides a good account of our data extends the physics-based approach to judgments of self-luminosity. This in turn might be taken as support for the notion that the visual system has internalized the physics of reflectance.

© 1996 Optical Society of America

## 1. INTRODUCTION

Color constancy can be modeled on the assumption that color appearance in natural scenes is a function of the visual system's estimates of surface reflectance.<sup>1</sup> Some stimuli, however, do not look like illuminated surfaces. Instead, they appear to be self-luminous.<sup>2-12</sup> Indeed, for any viewing context, colored stimuli at most chromaticities can be varied in luminance so that they appear either as surfaces (at low luminances) or as self-luminous objects (at high luminances). We refer to the luminance at which the mode of appearance changes as the luminosity threshold.<sup>13</sup>

To connect the appearance of luminosity to the idea that color appearance is mediated by the visual system's estimates of surfaces and illuminants, we hypothesized that stimuli will appear self-luminous when the visual system estimates a reflectance spectrum that is outside the gamut of physically realizable surfaces. This general notion has been proposed previously but has not been extensively explored.<sup>2-4,9,11</sup>

We measured luminosity thresholds as a function of stimulus chromaticity and illuminant spectral power distribution. Observers adjusted the luminance of a test patch until it just appeared self-luminous. The test patch was spot illuminated by a computer-controlled projection colorimeter viewed in an experimental room lit diffusely by computer-controlled theater lamps. Luminosity thresholds were determined for a number of test patch chromaticities under five experimental illuminants.

The luminosity thresholds define a surface in color

space. We find that the shape of this surface depends on the illuminant. We were able to describe the luminosity threshold surfaces with a simple model whose parameters define an equivalent illuminant. In the context of our model, the equivalent illuminant may be interpreted as the illuminant perceived by the observer. Given the equivalent illuminant, the model predicts that a patch will appear self-luminous when it is not consistent with any physically realizable surface seen under that illuminant. In addition, we show that much of the variation of the equivalent illuminant with the physical illuminant can be modeled with a simple linearity principle.

The fact that our model provides a good account of our data extends the physics-based approach to understanding color vision to handle judgments of self-luminosity. This in turn might be taken as support for the notion that the visual system has internalized the physics of reflectance, an idea that has been suggested in the context of color constancy.<sup>1,16</sup>

## 2. METHODS

### A. Overview

In our experiments we measured luminosity thresholds.<sup>4,10</sup> For a test patch of fixed chromaticity, we defined the luminosity threshold as the luminance at which the appearance of the test patch changes from that of a surface to that of a self-luminous object. We used the method of adjustment to determine thresholds for test patches of many chromaticities under five different conditions of ambient illumination.

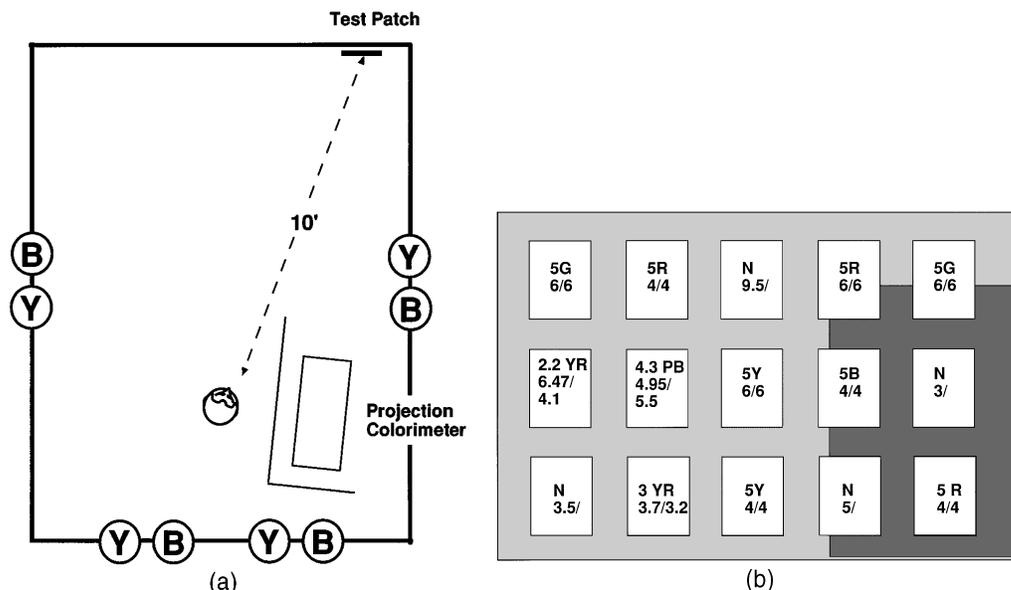


Fig. 1. (a) Top view of the experimental room. The ambient illumination was provided by eight theater stage lamps arranged in four pairs, B and Y. The observer viewed a test patch located on the far wall of the room, which was spot illuminated by a projection colorimeter. This apparatus provided experimental control of the luminance and chromaticity of the light reaching the observer from the test patch. (b) Schematic representation of the observer's view of the far wall of the experimental room. The far wall contained an array of 8.5-in.  $\times$  11-in. matte Munsell papers; each panel here is labeled with its Munsell designation. The test patch, Munsell designation N 3/, was located in the middle row and right-hand column of the array and was placed on a dark-gray piece of matte cardboard. Other objects in the room were visible to the observer, including a brown metal bookcase and an off-white table.

Our apparatus consisted of an entire experimental room, shown schematically in Fig. 1(a). The spectral power distribution of the ambient illumination in the room was produced by theater stage lamps and was under computer control. The observer judged the appearance of a test patch located on the far wall of the room. The light reflected from the test patch to the observer consisted of two components. The first was from the ambient illumination. The second was generated by a computer-controlled projection colorimeter and was spatially coincident with the test patch. The use of the colorimeter allowed us to vary the luminance of the light reaching the observer from the test patch while holding its chromaticity constant. The observer's task was to adjust the luminance of the test patch to the luminosity threshold. We describe our apparatus and experimental procedure in more detail in the sections that follow.

### B. Experimental Room

The walls and ceiling of the experimental room were painted a matte gray of roughly 50% reflectance. The far wall contained an array of 8.5-in.  $\times$  11-in. Munsell papers as shown in Fig. 1(b). Each panel was surrounded by a 0.25-in. black felt border. The test patch was located on the right-hand edge of the middle row of the array. From the observer's vantage point, the test patch subtended  $4.5^\circ \times 5.5^\circ$  of visual angle. A matte gray paper of 25% reflectance subtending  $16^\circ \times 20^\circ$  of visual angle was placed behind the test patch and was also partially occluded by several of the surrounding Munsell papers.

The ambient illumination of the room was controlled by two sets of theater stage lamps (SLD Lighting, 6-in. Fresnel #3053, BTL 500-W bulb), as shown in Fig. 1(a). The set labeled B had broadband blue filters (Roscolux #65); the set labeled Y had broadband yellow filters

(Roscolux #08). Each set of lights provided diffuse and roughly uniform ambient illumination. We controlled the lamps from software by varying the voltage across the bulbs (NSI 5600 Dimmer Packs, NSI OPT-232 interface card, 100 voltage quantization levels). We yoked the voltages of all lights within a set (B or Y) together. By varying the intensities of the two sets of lamps, we varied the spectral power distribution of the ambient illumination. Control software (see below) corrected for spectral shifts introduced when the voltage to the bulbs was varied.

The chromaticity and luminance of the test patch were controlled by the projection colorimeter. The colorimeter consisted of three slide projectors (Kodak 4400) racked in a vertical arrangement. Each projector beam was masked so that it illuminated only the test patch. The light from each projector passed through a red, green, or blue dichroic filter so that we had three independent primaries. The intensity of each primary was controlled by adjusting the voltage supplied to the corresponding projector lamp (NSI 5600 Dimmer Packs, NSI OPT-232 interface card, factory modified to provide 255 voltage quantization levels). Control software (see below) compensated for the ambient illumination reflected to the observer from the test patch and corrected for spectral shifts introduced when the voltage to the bulbs was varied. For the experiments reported here, we used the projection colorimeter to hold the chromaticity of the test patch constant while we varied its luminance.

### C. Colorimetric Characterization and Control

Colorimetric characterization and control of the experimental room presented some interesting challenges. We wanted software control of the spectral power distribu-

tion of the ambient illumination and, for any ambient illumination, of the chromaticity and luminance of the light reaching the observer from the test patch. We adopted a two-step approach to the characterization and control problem. The first step was to characterize as accurately as possible the relation between device settings and stimuli for both the theater lamps and the projection colorimeter. We used this characterization to control the stimuli during the experiments. At the end of each experimental session, however, we replayed the observer's luminosity thresholds and made direct spectral measurements. The data we report are direct measurements of the ambient illuminants and the observer's settings. This second step compensated for any small inaccuracies in the apparatus characterization.

For each set (B or Y) of lamps, we measured the full spectral power distribution at the test patch for 25 evenly spaced control voltages. The measurements were made with a spectracolorimeter (PhotoResearch PR-650) that imaged a reflectance standard (PhotoResearch RS-2, 99% reflectance  $\pm 1\%$  from 370 to 780 nm) placed directly in front of the test patch. These and all spectral measurements were made at 4-nm increments between 380 and 780 nm but interpolated with a cubic spline to the CIE-recommended wavelength sampling of 5-nm increments between 380 and 780 nm.<sup>17</sup> We fitted the measured spectra with a two-dimensional linear model.<sup>18</sup> Two-dimensional linear models provided good fits to the measured spectral data, accounting for essentially all of the measured variance. For each set the individual spectra could thus be represented with two coefficients. The first coefficient specified the contribution of the principal component of the entire data set to the spectrum. The second component specified a smaller correction that took account of shifts in the lamp spectra when the driving voltage was varied. We interpolated the relation between control voltage and linear-model coefficients to provide a complete 100-entry table of coefficients. Together with the spectra of the linear-model basis functions, this table allowed us to relate control voltage accurately to illuminant spectra.

Our control software used the characterization table and basis functions described above. By neglecting the higher-order term in the linear model, we could treat each lamp set as a single primary with a fixed relative spectral power distribution. Given this, we used multiple regression to find the principal linear coefficients (one for B and one for Y) to approximate any desired ambient illumination (in a least-squares sense). We then searched our device characterization table to find the appropriate device settings. This control procedure ignores the higher-order components of the linear model. Given the control voltages, however, we could predict which illuminant was actually produced by using all components of the model. Although in principle an iterative method could be used to converge on the closest approximation to the desired illuminant, here we simply accepted the one-step approximation. We characterized and controlled the projection colorimeter in a similar fashion. The colorimeter control software compensated for the ambient illumination reflected to the observer from the test patch.

The interested reader may obtain a sense of the precision of our characterization and control procedure from

Fig. 2 below. The figure shows both the desired and the actual test patch chromaticities used for one subject under one condition. The deviations of the actual chromaticities from the desired chromaticities are due to variance unaccounted for by our procedures. It is important to keep in mind, however, that our data and analyses are based on the actual measured values.

#### D. Experimental Procedure

The observer's task in our experiment was to adjust the luminance of the test patch until it just appeared self-luminous. To convey to observers what was meant by self-luminous, we chose a single lighting condition and presented a series of bright and dim stimuli at different chromaticities. All observers agreed that some of the stimuli looked like illuminated surfaces while others appeared to glow. The glowing appearance was much like that of a psychedelic poster viewed under black light. We demonstrated cases in which changing the luminance of the test patch while holding its chromaticity constant caused the test patch to change from one mode of appearance to the other. We explained that we wanted to study this transition and instructed the observer that his task was to adjust the intensity of the test patch to the transition point between the two modes of appearance. At the transition point that we studied, the stimulus retained surface qualities but did not look natural.

At the beginning of each experimental session the ambient lighting was set and the observer adapted for 1 min. The observer adjusted the test patch luminance by using a single knob. At the completion of the adjustment the observer indicated whether or not the result was satisfactory by pressing one of two buttons. Providing this choice was necessary because device gamut limitations occasionally made it impossible to achieve the transition luminance. Each session consisted of two blocks. Within each block, all test chromaticities were adjusted once in random order.

Immediately following each session, mean luminosity thresholds were computed. The resulting threshold stimuli were then displayed and measured directly. The ambient illumination was also measured.

#### E. Observers, Stimuli, and Details

Three male observers participated in the experiments reported here: the authors (JMS and DHB) and a paid graduate student (KI). All observers had normal color vision (as determined by anomaloscope for JMS and DHB and by pseudoisochromatic plates for KI). Subject KI was a graduate student generally interested in color vision. He was naïve as to our specific hypotheses at the time he observed in the experiments. Subject KI wore contact lenses (Menicon EX) while observing in the experiments and when he viewed the pseudoisochromatic plates. The spectral transmission of these lenses varied between 94% and 100% over the wavelength range 360–800 nm.

Each observer set luminosity thresholds under five different illumination conditions. We refer to the five corresponding illuminants as B1, Y1, B2, Y2, and BY. Illuminants B1 and Y1 had luminances of  $\sim 40$  cd/m<sup>2</sup> and differed in their chromaticities. Illuminant B1 was bluish and illuminant Y1 was yellowish. Illuminants B2 and Y2 had chromaticities similar to B1 and Y1, respec-

**Table 1. Illuminant 1931 CIE  $x$ ,  $y$  Chromaticities and Luminances for the Five Experimental Conditions**

Illuminant Condition	Chromaticity		Luminance Y (cd/m <sup>2</sup> )
	$x$	$y$	
Y1	0.49	0.42	43
B1	0.33	0.35	39
Y2	0.50	0.42	20
B2	0.35	0.37	16
BY	0.43	0.40	34

tively, but were dimmer, roughly 20 cd/m<sup>2</sup>. Illuminant BY approximated the admixture of B2 and Y2. Table 1 provides the illuminant chromaticities and luminances.

Because of gamut limitations, only certain chromaticities could be adjusted to threshold. These chromaticities varied with the ambient illumination and the observer. We ran preliminary experiments to determine the approximate gamuts for which luminosity thresholds could be set. During the actual experiments observers indicated whether each threshold setting was acceptable. We report results only for those test patch chromaticities for which acceptable thresholds were set. The decision regarding acceptability was made by the subject during data collection and before we had conducted any data analysis. The chromaticities of the test patches for each observer/illuminant condition are provided as part of Tables 2–4 below.

#### F. Methodological Discussion

We generally found the task of setting luminosity thresholds to be natural. For the chromaticities within gamut, observers were confident that they saw stimuli on either side of the threshold. Some observers exhibited considerable session-to-session variability of the criterion they used to define the threshold setting, a point that we consider in detail below.

The judgment we required of observers is similar to judgments studied by Buhler,<sup>2</sup> Evans,<sup>4,5</sup> Evans and Swenholt,<sup>6–8</sup> Ullman,<sup>9</sup> Bonato and Gilchrist,<sup>10</sup> and Ikeda and colleagues.<sup>11,12</sup> All of these authors required observers to distinguish between stimuli that appeared as illuminated surfaces and those that appeared in some way different. All agree that the judgment is reasonable. Evans and Swenholt distinguished two separate but related judgments. In one, observers judged whether stimuli appeared to be self-luminous light sources. In the other, observers judged whether stimuli appeared to be fluorescent (appear as fluorescent surfaces) or (equivalently in their view) to have zero gray content. Although our instructions do not match theirs, we suspect that the judgment made by our observers was similar to their fluorescence judgment. This is because in our experiment, the stimuli retained surfacelike qualities at the luminosity threshold, a feature that Evans regards as crucial to the distinction. Like us, the later authors<sup>9–12</sup> do not differentiate precisely between the two tasks. In retrospect, more-precise instructions to our observers might have reduced criterion shifts.

It is possible to study luminosity thresholds with a

yes–no procedure.<sup>10</sup> In such a procedure the observer is presented with a series of experimentally controlled test patches and simply judges whether each appears luminous or not. In pilot experiments<sup>19</sup> we compared a yes–no staircase procedure with an adjustment procedure. We did not observe any systematic differences between the results obtained with the two procedures and hence adopted the faster adjustment procedure for the main experiments.

One important feature of our study is that our observers viewed the test patch in a nearly natural viewing context. Because our modeling approach will be based on the idea that the visual system recovers (perhaps implicitly) an estimate of the illuminant, we felt that it was important to provide as many cues as possible. Computational analyses of color constancy<sup>20,21</sup> have shown that when an isolated patch is viewed against a uniform background under spatially homogeneous illumination, it is not possible to estimate the illuminant accurately.

### 3. RESULTS

Tables 2–4 tabulate our data. For each illuminant condition each observer set thresholds in either two or three experimental sessions. The tables provide the test patch chromaticities and mean luminosity thresholds for each observer/illuminant condition. For any observer/illuminant condition, the experiment determines luminosity thresholds as a function of test patch chromaticity. Figure 2 shows the sample chromaticities used for one observer/illuminant condition. The sampling density is typical for all of our conditions, although the exact chromaticities vary from condition to condition. Figure 3 shows, for two observers, contour plots of data collected under two illuminants. Note that the luminosity thresholds vary as the illuminant changes. This is seen most clearly in the data for observer DHB (note the shift in the peak of the luminosity threshold surface). For observer JMS the change is less apparent. This is primarily because the stimulus gamuts for which JMS could make settings did not overlap substantially and because neither surface shows a marked peak. Nonetheless, careful examination of the plot indicates that the gradient of the threshold surface is oriented differently under the two illuminants. The data for other illuminant pairs for these

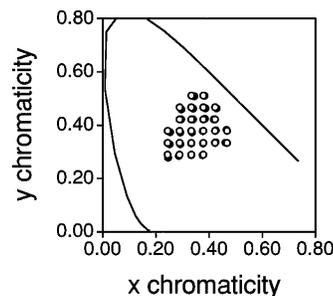


Fig. 2. Stimulus chromaticities for illuminant condition B2, observer DHB. Open symbols, desired chromaticities; filled symbols, chromaticities measured during the postcalibration procedure. Spectrum locus, 380–700 nm, sampled at 10-nm intervals. The density for other observers and conditions was comparable.

**Table 2. Luminosity Threshold Data for Subject JMS<sup>a</sup>**

Chromaticity		Threshold Luminance (cd/m <sup>2</sup> )			Mean T	Standard Deviation T
x	y	T1	T2	T3		
<b>Illuminant Condition Y1</b>						
0.38	0.37	3.62	3.43	3.61	3.55	0.11
0.38	0.42	4.12	4.52	5.22	4.62	0.55
0.42	0.38	4.22	3.33	3.76	3.77	0.44
0.43	0.42	5.00	4.64	6.11	5.25	0.76
0.43	0.46	6.14	5.53	6.38	6.01	0.44
0.43	0.51	5.92	5.30	5.92	5.71	0.36
0.47	0.38	5.07	4.35	4.84	4.75	0.37
0.47	0.42	6.33	6.11	6.40	6.28	0.15
0.47	0.46	7.11	6.86	7.97	7.31	0.58
0.51	0.38	6.56	5.82	6.40	6.26	0.39
0.51	0.42	7.94	7.19	7.82	7.65	0.40
0.56	0.38	6.17	5.56	6.24	5.99	0.37
<b>Illuminant Condition B1</b>						
0.25	0.24	3.23	3.49	NA	3.36	0.18
0.25	0.28	4.23	4.27	NA	4.25	0.03
0.25	0.33	6.14	5.71	NA	5.92	0.31
0.29	0.29	4.41	4.93	NA	4.67	0.37
0.30	0.33	5.27	6.15	NA	5.71	0.62
0.30	0.37	7.51	6.63	NA	7.07	0.63
0.30	0.42	7.72	7.40	NA	7.56	0.23
0.30	0.46	7.95	7.20	NA	7.57	0.53
0.34	0.29	4.68	4.85	NA	4.77	0.12
0.34	0.33	6.26	6.72	NA	6.49	0.33
0.34	0.37	7.12	7.04	NA	7.08	0.06
0.34	0.41	8.35	6.85	NA	7.60	1.06
0.34	0.46	7.91	8.13	NA	8.02	0.15
0.34	0.51	7.75	8.10	NA	7.92	0.25
0.38	0.33	6.70	6.92	NA	6.81	0.16
0.39	0.38	7.15	7.23	NA	7.19	0.05
0.39	0.41	8.91	7.71	NA	8.31	0.85
0.39	0.46	8.93	8.55	NA	8.74	0.27
0.42	0.33	5.56	6.42	NA	5.99	0.61
0.42	0.38	7.56	6.99	NA	7.27	0.40
0.43	0.42	7.65	8.00	NA	7.83	0.25
0.47	0.33	5.93	5.58	NA	5.75	0.25
0.47	0.38	6.25	6.72	NA	6.48	0.33
<b>Illuminant Condition Y2</b>						
0.37	0.33	1.40	1.70	NA	1.55	0.21
0.38	0.37	1.62	2.23	NA	1.92	0.43
0.38	0.42	2.51	2.56	NA	2.53	0.04
0.38	0.46	2.87	3.02	NA	2.95	0.10
0.38	0.50	3.22	2.97	NA	3.09	0.17
0.42	0.34	1.40	2.00	NA	1.70	0.43
0.42	0.38	2.06	2.12	NA	2.09	0.04
0.43	0.42	2.66	3.33	NA	2.99	0.47
0.43	0.46	3.39	3.46	NA	3.43	0.05
0.43	0.50	3.03	3.47	NA	3.25	0.31
0.47	0.38	3.00	2.99	NA	3.00	0.01
0.46	0.42	3.44	3.30	NA	3.37	0.10
0.47	0.46	3.87	3.99	NA	3.93	0.09
0.51	0.38	3.62	3.47	NA	3.55	0.10
0.51	0.42	4.45	3.99	NA	4.22	0.32
0.56	0.38	3.60	3.54	NA	3.57	0.04
0.56	0.42	4.79	4.41	NA	4.60	0.27
<b>Illuminant Condition B2</b>						
0.25	0.28	1.80	2.05	NA	1.93	0.18
0.25	0.32	2.28	2.66	NA	2.47	0.27
0.29	0.29	1.69	2.11	NA	1.90	0.29
0.29	0.32	2.46	2.49	NA	2.48	0.02
0.30	0.37	2.72	3.15	NA	2.94	0.30

**Table 2. (Continued)**

Chromaticity		Threshold Luminance (cd/m <sup>2</sup> )			Mean T	Standard Deviation T
x	y	T1	T2	T3		
<b>Illuminant Condition B2 (continued)</b>						
0.30	0.42	3.16	3.27	NA	3.21	0.08
0.30	0.46	3.13	3.55	NA	3.34	0.30
0.33	0.29	1.82	2.34	NA	2.08	0.37
0.33	0.33	2.50	3.19	NA	2.84	0.48
0.34	0.37	3.23	3.32	NA	3.27	0.07
0.34	0.42	3.35	3.67	NA	3.51	0.23
0.34	0.46	3.49	3.88	NA	3.68	0.28
0.34	0.50	3.12	3.78	NA	3.45	0.47
0.38	0.33	2.90	3.34	NA	3.12	0.31
0.38	0.38	3.08	3.53	NA	3.31	0.32
0.39	0.42	3.35	3.86	NA	3.60	0.36
0.39	0.46	3.54	4.05	NA	3.80	0.36
0.39	0.50	3.41	4.10	NA	3.76	0.49
0.43	0.33	2.79	3.41	NA	3.10	0.44
0.43	0.38	3.17	3.66	NA	3.41	0.35
0.42	0.42	3.14	3.67	NA	3.40	0.38
0.43	0.46	3.76	4.25	NA	4.01	0.34
0.46	0.33	2.38	3.16	NA	2.77	0.55
0.47	0.38	2.83	3.28	NA	3.06	0.32
0.47	0.42	2.96	3.58	NA	3.27	0.43
0.52	0.38	2.52	3.12	NA	2.82	0.42
<b>Illuminant Condition BY</b>						
0.29	0.33	3.56	3.21	NA	3.38	0.24
0.34	0.33	2.90	3.35	NA	3.13	0.32
0.34	0.37	4.81	4.50	NA	4.66	0.22
0.34	0.42	5.13	5.04	NA	5.09	0.06
0.34	0.46	5.97	5.81	NA	5.89	0.11
0.38	0.33	3.69	3.45	NA	3.57	0.16
0.38	0.37	4.58	5.10	NA	4.84	0.37
0.38	0.42	6.02	5.98	NA	6.00	0.03
0.39	0.46	6.25	6.23	NA	6.24	0.01
0.39	0.51	6.04	6.22	NA	6.13	0.13
0.42	0.33	3.94	4.89	NA	4.41	0.68
0.43	0.38	5.64	6.16	NA	5.90	0.36
0.43	0.42	6.23	6.57	NA	6.40	0.24
0.43	0.46	7.30	6.85	NA	7.08	0.32
0.46	0.33	4.89	4.98	NA	4.94	0.06
0.46	0.38	6.52	6.00	NA	6.26	0.37
0.47	0.42	6.35	6.38	NA	6.37	0.02
0.47	0.46	7.73	7.48	NA	7.61	0.17
0.51	0.38	5.89	6.23	NA	6.06	0.24
0.51	0.42	6.73	6.54	NA	6.63	0.13
0.56	0.38	5.22	5.72	NA	5.47	0.35

<sup>a</sup>The data set includes only those stimuli for which luminosity thresholds were set for all stimulus presentations. NA, not applicable.

two observers and for observer KI were qualitatively similar. The data shown represent the two extrema in terms of the apparent effect of changing the illumination.

Figures 4 and 5 show another view of the data. Here luminosity thresholds are plotted as a function of CIE *x* chromaticity. For observer DHB (Fig. 4) we have separated the data into two panels for clarity. For observer JMS (Fig. 5) the plot is clearer with data from both illuminants presented in a single panel. These plots also reveal that the luminosity thresholds vary with test patch chromaticity. The error bars in both figures indicate the standard error of the mean computed across sessions. We argue below that the large error bars for observer DHB are due to criterion shifts.

**Table 3. Luminosity Threshold Data for Subject DHB**

Chromaticity		Threshold Luminance (cd/m <sup>2</sup> )			Standard	
x	y	T1	T2	Mean T	Deviation T	
<b>Illuminant Condition Y1</b>						
0.33	0.37	10.68	4.64	7.66	4.27	
0.37	0.37	14.89	5.10	9.99	6.92	
0.38	0.42	12.13	5.82	8.98	4.46	
0.37	0.46	10.64	7.58	9.11	2.17	
0.41	0.33	9.26	6.54	7.90	1.92	
0.42	0.38	15.17	9.52	12.34	3.99	
0.42	0.42	20.59	9.76	15.17	7.65	
0.42	0.46	13.38	7.85	10.62	3.91	
0.42	0.51	10.56	6.42	8.49	2.93	
0.46	0.33	11.85	6.33	9.09	3.90	
0.46	0.38	13.30	8.94	11.12	3.09	
0.46	0.42	16.86	13.53	15.19	2.35	
0.47	0.47	13.58	7.83	10.70	4.07	
0.50	0.38	12.46	10.64	11.55	1.29	
0.51	0.42	12.35	11.17	11.76	0.83	
0.56	0.38	8.01	5.94	6.98	1.47	
<b>Illuminant Condition B1</b>						
0.25	0.28	8.26	4.01	6.14	3.00	
0.25	0.33	10.29	6.78	8.53	2.48	
0.25	0.37	11.76	8.67	10.21	2.19	
0.29	0.28	8.88	7.48	8.18	0.99	
0.29	0.32	11.48	10.89	11.18	0.42	
0.29	0.37	15.68	10.56	13.12	3.63	
0.29	0.42	16.65	13.13	14.89	2.49	
0.29	0.46	13.82	11.99	12.91	1.29	
0.33	0.33	13.05	11.70	12.37	0.96	
0.34	0.42	17.88	11.46	14.67	4.54	
0.34	0.46	17.53	11.56	14.55	4.23	
0.34	0.51	13.53	10.99	12.26	1.80	
0.37	0.28	10.03	8.40	9.21	1.15	
0.37	0.33	13.36	11.84	12.60	1.08	
0.38	0.38	18.40	9.46	13.93	6.32	
0.38	0.41	20.85	15.80	18.32	3.57	
0.38	0.46	22.14	12.30	17.22	6.96	
0.41	0.33	12.63	9.58	11.10	2.16	
0.42	0.38	17.46	10.86	14.16	4.66	
0.42	0.42	14.55	11.87	13.21	1.89	
0.42	0.46	20.40	15.79	18.10	3.26	
0.47	0.33	10.07	6.01	8.04	2.87	
0.46	0.38	12.39	8.82	10.60	2.52	
0.47	0.43	11.17	10.24	10.71	0.66	
<b>Illuminant Condition Y2</b>						
0.33	0.42	5.10	2.91	4.00	1.55	
0.38	0.38	9.82	2.05	5.94	5.49	
0.38	0.42	11.15	3.30	7.22	5.55	
0.38	0.46	7.33	4.34	5.83	2.12	
0.38	0.51	6.19	3.32	4.76	2.03	
0.42	0.38	14.95	4.03	9.49	7.72	
0.42	0.42	12.78	4.10	8.44	6.14	
0.42	0.47	7.98	4.45	6.22	2.50	
0.43	0.51	7.62	3.08	5.35	3.20	
0.46	0.33	7.97	2.80	5.39	3.66	
0.46	0.38	10.88	3.94	7.41	4.91	
0.46	0.42	16.84	6.50	11.67	7.31	
0.47	0.47	10.31	4.21	7.26	4.32	
0.50	0.33	4.04	3.64	3.84	0.28	
0.50	0.38	10.92	5.69	8.30	3.69	
0.51	0.42	11.79	6.00	8.89	4.10	
0.55	0.38	7.68	2.89	5.28	3.39	
0.55	0.43	8.69	4.52	6.61	2.94	

**Table 3. (Continued)**

Chromaticity		Threshold Luminance (cd/m <sup>2</sup> )			Standard	
x	y	T1	T2	Mean T	Deviation T	
<b>Illuminant Condition B2</b>						
0.25	0.28	2.76	2.28	2.52	0.34	
0.25	0.33	3.60	2.45	3.03	0.81	
0.25	0.37	5.19	3.56	4.38	1.15	
0.29	0.28	4.36	2.07	3.22	1.62	
0.29	0.33	5.32	2.73	4.03	1.83	
0.29	0.37	7.59	4.47	6.03	2.20	
0.29	0.42	6.74	4.05	5.40	1.91	
0.30	0.46	5.52	3.60	4.56	1.36	
0.33	0.29	5.06	1.81	3.43	2.30	
0.33	0.33	6.65	3.01	4.83	2.57	
0.34	0.37	14.26	5.91	10.08	5.90	
0.34	0.42	8.35	4.20	6.28	2.94	
0.34	0.46	7.54	4.28	5.91	2.30	
0.34	0.51	6.34	2.91	4.63	2.43	
0.37	0.29	3.61	2.82	3.22	0.56	
0.38	0.33	5.51	5.99	5.75	0.34	
0.38	0.38	12.78	7.55	10.16	3.70	
0.38	0.42	9.23	5.00	7.11	2.99	
0.39	0.46	8.69	5.40	7.05	2.33	
0.38	0.51	6.18	4.65	5.41	1.09	
0.42	0.34	5.42	3.31	4.36	1.49	
0.43	0.38	5.97	5.01	5.49	0.68	
0.42	0.42	7.78	5.77	6.77	1.42	
0.43	0.46	6.57	6.17	6.37	0.29	
0.46	0.33	4.21	2.93	3.57	0.90	
0.47	0.38	5.10	2.46	3.78	1.87	
0.47	0.42	4.79	4.68	4.73	0.08	
<b>Illuminant Condition BY</b>						
0.29	0.37	5.90	6.22	6.06	0.22	
0.29	0.42	7.08	7.64	7.36	0.40	
0.33	0.33	5.21	3.58	4.40	1.15	
0.33	0.37	7.78	5.58	6.68	1.56	
0.34	0.42	8.83	6.44	7.64	1.69	
0.34	0.46	7.84	5.67	6.75	1.54	
0.34	0.51	7.37	7.27	7.32	0.07	
0.37	0.33	3.46	7.65	5.56	2.96	
0.37	0.37	10.67	9.81	10.24	0.61	
0.38	0.42	12.98	7.25	10.11	4.05	
0.38	0.46	5.16	9.83	7.50	3.31	
0.38	0.51	7.42	4.81	6.12	1.85	
0.42	0.33	6.42	6.05	6.23	0.26	
0.42	0.38	12.29	10.10	11.19	1.55	
0.41	0.42	10.96	16.83	13.89	4.15	
0.42	0.46	11.31	9.36	10.34	1.38	
0.46	0.33	8.01	5.53	6.77	1.76	
0.46	0.38	12.58	—	11.13	2.04	
0.46	0.42	9.54	10.70	10.12	0.81	
0.46	0.47	13.90	12.75	13.33	0.82	
0.51	0.38	8.73	8.21	8.47	0.37	
0.51	0.42	8.83	7.70	8.26	0.80	

To examine the variability of our data, we computed the difference between individual session luminosity thresholds and the mean threshold computed across sessions. Histograms of these differences for each observer (accumulated across test chromaticities and illuminant conditions) are shown in Fig. 6. These histograms summarize the precision to which observers set thresholds across sessions. The error can be quite large, particularly for ob-

**Table 4. Luminosity Threshold Data for Subject KI**

Chromaticity		Threshold Luminance (cd/m <sup>2</sup> )			Standard Deviation T
x	y	T1	T2	Mean T	
<b>Illuminant Condition Y1</b>					
0.28	0.27	5.02	6.62	5.82	1.13
0.32	0.28	5.74	8.17	6.95	1.72
0.33	0.32	7.57	9.61	8.59	1.44
0.33	0.37	9.45	12.49	10.97	2.15
0.33	0.41	10.12	11.41	10.76	0.91
0.33	0.45	8.96	11.24	10.10	1.61
0.37	0.28	6.28	8.67	7.47	1.69
0.37	0.32	8.56	11.73	10.14	2.24
0.38	0.37	10.19	13.88	12.03	2.61
0.38	0.42	12.48	15.55	14.02	2.17
0.38	0.46	10.19	13.24	11.71	2.15
0.38	0.51	8.39	13.90	11.14	3.89
0.42	0.33	8.71	14.10	11.40	3.82
0.42	0.37	11.45	15.00	13.22	2.51
0.42	0.42	12.73	17.22	14.97	3.18
0.42	0.46	11.77	13.94	12.86	1.53
0.42	0.51	9.55	14.97	12.26	3.83
0.46	0.33	11.84	14.89	13.36	2.16
0.46	0.37	10.16	17.14	13.65	4.93
0.46	0.42	11.69	17.09	14.39	3.82
0.47	0.46	11.41	13.96	12.68	1.80
0.50	0.33	12.27	13.48	12.87	0.86
0.50	0.38	10.71	15.74	13.22	3.56
0.51	0.42	12.39	17.87	15.13	3.88
0.60	0.38	10.55	10.62	10.59	0.05
<b>Illuminant Condition B1</b>					
0.29	0.41	14.03	17.93	15.98	2.76
0.29	0.45	13.58	15.21	14.39	1.16
0.33	0.32	13.58	16.32	14.95	1.94
0.33	0.37	15.97	18.10	17.03	1.51
0.34	0.42	12.84	19.53	16.18	4.73
0.34	0.46	14.09	18.48	16.29	3.10
0.34	0.50	15.87	15.90	15.88	0.02
0.37	0.28	11.03	11.38	11.20	0.25
0.37	0.33	14.05	15.14	14.59	0.77
0.38	0.37	13.26	19.61	16.43	4.49
0.38	0.41	14.93	18.69	16.81	2.66
0.38	0.46	16.60	22.03	19.32	3.84
0.38	0.50	16.52	17.44	16.98	0.65
0.41	0.33	12.78	15.11	13.95	1.65
0.42	0.38	14.84	16.25	15.54	1.00
0.42	0.42	16.14	20.21	18.17	2.88
0.42	0.46	18.85	21.65	20.25	1.98
0.46	0.33	13.07	11.42	12.24	1.17
0.46	0.38	11.32	15.96	13.64	3.28
0.46	0.42	13.50	17.36	15.43	2.73
0.51	0.38	12.15	12.67	12.41	0.37
0.51	0.38	11.86	12.68	12.27	0.58
<b>Illuminant Condition Y1</b>					
0.28	0.23	2.40	2.66	2.53	0.19
0.28	0.27	4.57	3.71	4.14	0.61
0.29	0.32	5.67	4.66	5.17	0.71
0.29	0.37	5.30	4.98	5.14	0.22
0.33	0.28	4.17	5.38	4.77	0.86
0.33	0.32	5.23	5.65	5.44	0.30
0.33	0.37	7.50	7.25	7.37	0.18
0.33	0.42	6.57	6.01	6.29	0.40
0.33	0.45	5.72	5.45	5.59	0.19
0.37	0.28	5.42	5.80	5.61	0.26
0.37	0.33	6.06	7.51	6.78	1.03
0.38	0.37	6.73	8.16	7.44	1.01
0.38	0.42	8.13	7.62	7.88	0.36

**Table 4. (Continued)**

Chromaticity		Threshold Luminance (cd/m <sup>2</sup> )			Standard Deviation T
x	y	T1	T2	Mean T	
<b>Illuminant Condition Y1 (continued)</b>					
0.38	0.46	7.15	6.92	7.04	0.16
0.38	0.51	7.78	6.80	7.29	0.70
0.38	0.55	5.64	5.71	5.67	0.05
0.42	0.28	5.19	5.93	5.56	0.52
0.42	0.33	7.02	7.44	7.23	0.29
0.42	0.37	7.89	8.71	8.30	0.58
0.42	0.42	7.94	8.20	8.07	0.19
0.42	0.46	8.06	7.10	7.58	0.68
0.43	0.51	8.72	7.12	7.92	1.13
0.45	0.33	8.04	8.69	8.36	0.46
0.46	0.37	9.34	8.71	9.02	0.44
0.46	0.42	7.76	8.70	8.23	0.67
0.47	0.46	8.37	8.86	8.62	0.35
0.50	0.33	7.51	7.80	7.66	0.21
0.51	0.37	7.86	7.56	7.71	0.21
0.51	0.42	8.72	10.43	9.57	1.20
0.51	0.47	8.82	10.61	9.72	1.27
0.55	0.38	7.87	7.45	7.66	0.30
0.56	0.42	7.50	8.58	8.04	0.76
0.60	0.38	6.46	5.38	5.92	0.76
0.60	0.38	6.09	5.11	5.60	0.70
<b>Illuminant Condition B2</b>					
0.24	0.23	4.28	3.85	4.07	0.30
0.24	0.27	5.64	5.18	5.41	0.32
0.24	0.32	5.88	7.29	6.59	1.00
0.28	0.23	5.20	5.01	5.10	0.13
0.29	0.28	5.33	6.03	5.68	0.49
0.29	0.32	7.58	9.16	8.37	1.12
0.29	0.37	8.92	8.61	8.76	0.23
0.29	0.42	7.48	8.80	8.14	0.93
0.29	0.45	7.79	8.90	8.35	0.79
0.33	0.28	7.14	7.69	7.42	0.39
0.33	0.33	8.11	8.53	8.32	0.29
0.33	0.37	10.57	11.58	11.07	0.72
0.34	0.42	8.73	10.11	9.42	0.98
0.34	0.45	6.65	9.05	7.85	1.70
0.34	0.50	6.70	10.43	8.57	2.64
0.34	0.55	7.70	7.58	7.64	0.08
0.37	0.28	6.56	7.33	6.94	0.54
0.38	0.33	7.93	7.89	7.91	0.03
0.38	0.37	8.09	11.88	9.98	2.68
0.38	0.42	7.10	11.35	9.22	3.00
0.38	0.46	9.92	11.00	10.46	0.77
0.38	0.51	6.83	10.58	8.70	2.65
0.42	0.28	5.61	7.33	6.47	1.21
0.42	0.33	7.33	10.61	8.97	2.32
0.42	0.38	8.13	13.91	11.02	4.09
0.42	0.42	6.65	11.88	9.26	3.69
0.43	0.46	8.12	12.84	10.48	3.34
0.46	0.33	6.26	8.79	7.52	1.79
0.46	0.38	7.03	9.32	8.17	1.62
0.46	0.42	7.49	11.02	9.25	2.49
0.51	0.38	6.03	5.99	6.01	0.03
0.51	0.42	7.42	7.92	7.67	0.36
<b>Illuminant Condition BY</b>					
0.24	0.27	6.13	6.59	6.36	0.33
0.24	0.32	8.17	9.27	8.72	0.77
0.28	0.23	5.48	5.61	5.54	0.09
0.28	0.28	6.53	7.33	6.93	0.56
0.29	0.41	11.10	15.70	13.40	3.25
0.33	0.41	13.46	18.65	16.05	3.67
0.33	0.46	14.12	17.56	15.84	2.43
0.33	0.50	11.74	13.98	12.86	1.58

**Table 4. (Continued)**

Chromaticity		Threshold Luminance		Mean T	Standard Deviation T
		(cd/m <sup>2</sup> )			
x	y	T1	T2		
Illuminant Condition BY (continued)					
0.37	0.28	12.26	9.50	10.88	1.95
0.38	0.42	15.70	21.06	18.38	3.79
0.38	0.46	15.39	13.73	14.56	1.17
0.38	0.51	14.07	17.84	15.95	2.66
0.41	0.28	11.53	11.80	11.66	0.19
0.42	0.33	14.07	15.28	14.67	0.85
0.42	0.37	15.80	17.27	16.54	1.04
0.42	0.42	14.38	22.13	18.26	5.48
0.42	0.46	16.98	23.03	20.01	4.28
0.42	0.51	16.24	18.60	17.42	1.67
0.46	0.37	14.20	15.48	14.84	0.90
0.46	0.42	14.54	19.16	16.85	3.26
0.46	0.47	19.54	22.49	21.02	2.09

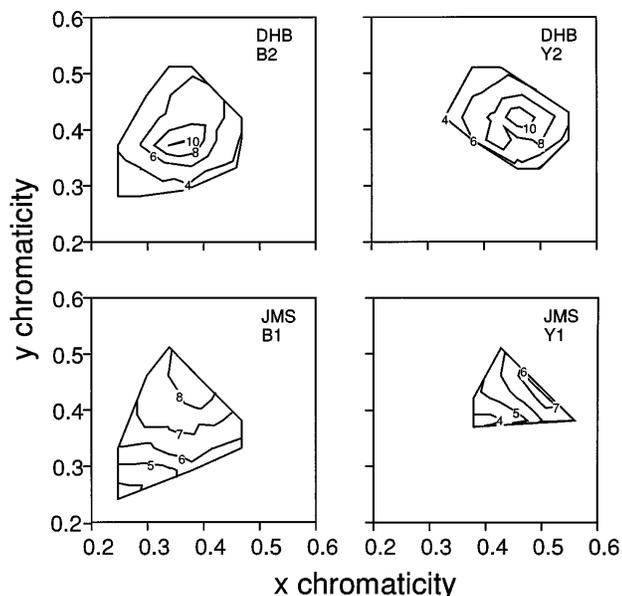


Fig. 3. Contour plots of the luminosity threshold surfaces for observers JMS and DHB as a function of  $x$  and  $y$  chromaticity, each for a single observer/illuminant condition. Each contour is labeled with its luminance in candelas per square meter. The pairs of plots for each observer illustrate that the shape of the luminosity threshold surface depends on the illumination.

server DHB. This fact is also apparent in the large error bars of Fig. 4.

In comparing data collected across sessions for single observers we noticed that the shape of the threshold surface was roughly the same in each session but that the overall level of the thresholds was different. To illustrate this point, Fig. 7 shows slices through the luminosity threshold surface from individual sessions for observer DHB. Notice that across sessions the slices have approximately the same shape but different overall luminances. We attribute these shifts to changes in criterion across sessions.

To evaluate the contribution of criterion shifts to the overall measurement error, we partitioned the differences shown in Fig. 6 into two components: a component at-

tributable to criterion shift alone and a residual component due to other sources of variability. To make the partition, we found multiplicative scale factors that shifted the mean (across session) data as close as possible to the individual session data. A single scale factor was used for all test patch chromaticities within a single session, so that there was one free parameter per session to account for criterion shifts. The scale factors are tabulated in Table 5. Histograms of the differences between the shifted mean and individual session thresholds are shown in Fig. 8. These provide a metric for how well our data establish the relative shape of the luminosity thresholds, without regard to their overall level. The corrected histograms are considerably narrower than their

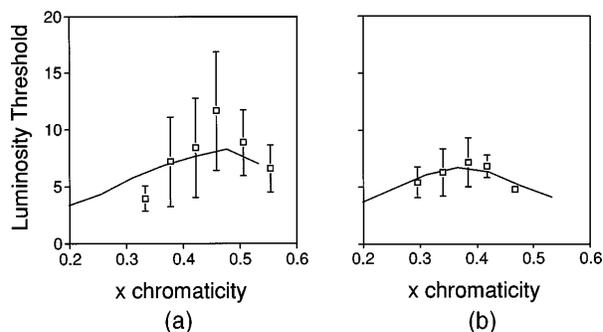


Fig. 4. Slices through the luminosity threshold surfaces for observer DHB: (a) illuminant Y2 and  $y$  chromaticity 0.42; (b) illuminant B2 and  $y$  chromaticity 0.42. The solid curves show a model fit, which is discussed below. The thresholds are given in units of  $\text{cd}/\text{m}^2$ . The error bars indicate the standard error of the mean computed across sessions.

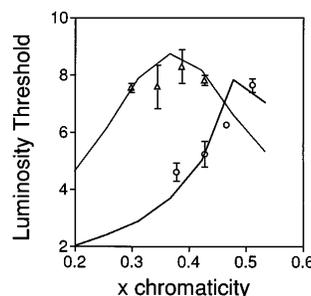


Fig. 5. Slices through the luminosity threshold surfaces for observer JMS: illuminants Y1 (open circles) and B1 (open triangles) at a  $y$  chromaticity of 0.42. The solid curves show a model fit, which is discussed below. The thresholds are given in candelas per square meter. The error bars indicate the standard error of the mean computed across sessions.

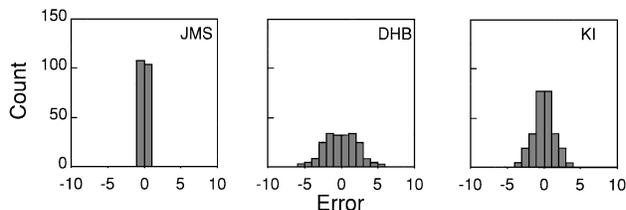


Fig. 6. Each panel shows a histogram of the luminance error in candelas per square meter between the luminosity thresholds for individual sessions and the corresponding mean threshold. The errors for each observer are accumulated across both sessions and conditions. The RMSE's for observers JMS, DHB, and KI were 0.26, 2.20, and 1.42  $\text{cd}/\text{m}^2$ , respectively.

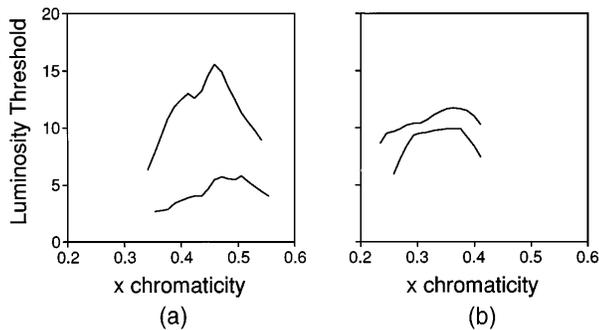


Fig. 7. Individual session variability for observer DHB. The two panels show slices through the threshold surfaces for the same conditions but collected in different sessions. We obtained the slice data by applying a two-dimensional linear-interpolation algorithm to the raw data and then slicing the interpolation, allowing us to compute slices for arbitrary values of  $y$  chromaticity. The thresholds are in candelas per square meter. (a) Illuminant Y2,  $y$  chromaticity 0.41; (b) illuminant condition B1,  $y$  chromaticity 0.31. The variability across sessions is primarily one of overall luminance; across sessions the relative shape of the slices is approximately constant.

**Table 5. Scale Factors Providing the Best Fit between Mean Thresholds and Thresholds Measured in Each Session<sup>a</sup>**

Subject	Condition	Scale Factor		
		1	2	3
JMS	Y1	1.01	0.93	1.05
	B1	1.01	0.99	—
	Y2	0.99	1.01	—
	B2	0.93	1.07	—
	BY	1.00	1.00	—
DHB	Y1	1.23	0.77	—
	B1	1.16	0.84	—
	Y2	1.42	0.58	—
	B2	1.24	0.76	—
	BY	1.03	0.97	—
KI	Y1	0.85	1.15	—
	B1	0.91	1.09	—
	Y2	0.99	1.01	—
	B2	0.88	1.12	—
	BY	0.91	1.09	—

<sup>a</sup>For each illuminant condition the scale factors shift the mean threshold data to the individual session data.

raw counterparts of Fig. 6, indicating that criterion shifts across sessions account for a substantial portion of our experimental variability. Evans<sup>4</sup> also reported difficulties with criterion shifts, which he dealt with by providing extensive training to a single observer. We take a different tack and restrict further analysis to understanding the relative shapes of our measured luminosity threshold surfaces. We do not know whether the observed criterion shifts were perceptual or decisional in nature.

#### 4. MODEL

A complete understanding of our data requires two parts. First, we need to understand the parametric form of the luminosity threshold surface for any illuminant. Second, we need to understand how the parameters that describe

the luminosity thresholds depend on the illumination. In this section we address these two questions in turn. First we develop and evaluate a physics-based model that predicts the relative shape of the luminosity threshold surfaces. Second, we ask whether a simple illuminant linearity principle can explain how the parameters of our model depend on the illumination.

The key idea underlying our model is the physical fact that surfaces cannot reflect more light than strikes them. The roots for using this fact to understand the appearance of luminosity extend back to Hering,<sup>22</sup> who suggested that stimuli should appear self-luminous when their luminance exceeds that of the brightest white in the scene. Buhler's refinement of Hering's notion<sup>2</sup> was to link the appearance of luminosity explicitly to the perceived intensity of the illuminant. Evans used MacAdam's computations of the physical gamut of reflective surfaces<sup>23</sup> to generate a quantitative model for his fluorence judgments.<sup>4</sup> Although Evans did not find good agreement with his data and later proposed other models, we believe the idea deserves further consideration. We have extended Evans's model in two ways. As suggested by Buhler,<sup>2</sup> we allow for the possibility that the observer misestimates the illuminant. Second, we have incorporated linear-model constraints into our calculation of the physical gamut of reflective surfaces.

To help understand our model, consider Fig. 9. The figure illustrates a simplified version of the physics of reflective image formation. A set of surfaces with reflectance functions  $S_j(\lambda)$  are illuminated by a common illuminant with spectral power distribution  $E(\lambda)$ . The color signal reaching the observer's eye at each location,  $C_j(\lambda)$ , is given as the wavelength-by-wavelength product of  $E(\lambda)$  and  $S_j(\lambda)$ . We assume that the visual system attempts to parse the color signals from the image into an estimate of the common illuminant spectral power distribution and the surface-reflectance functions at each location.<sup>1,24,25</sup> Research in computational color constancy has shown that it is possible to make approximately correct estimates of the common illuminant when the image contains multiple surfaces.<sup>20,21,26,27</sup>

Physically realizable surfaces cannot reflect more light than is incident upon them. This constraint is expressed

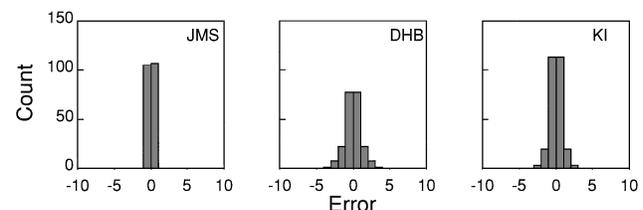


Fig. 8. Each panel shows a histogram of the luminance error in candelas per square meter between the luminosity thresholds for individual sessions and the mean thresholds after criterion shifts have been taken into account. To model criterion shifts, we found for each observer/illuminant condition the single scale factor that shifted the mean (across session) data for all test patch chromaticities as close as possible to the individual session data. The errors for each observer are accumulated across both sessions and conditions. The RMSE's for observers JMS, DHB, and KI were 0.21, 1.03, and 0.78  $\text{cd}/\text{m}^2$ , respectively. For observers DHB and KI the errors after scaling were substantially reduced from the raw errors: 2.20 and 1.42  $\text{cd}/\text{m}^2$ , respectively. Observer JMS showed less criterion shift across sessions.

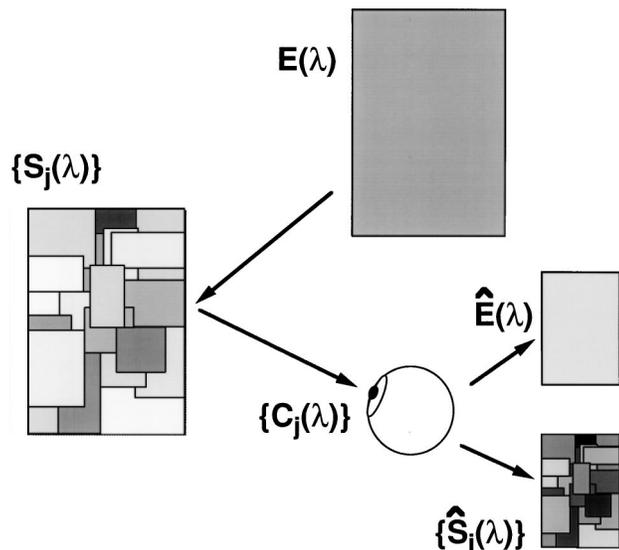


Fig. 9. Physics of reflective image formation. The light reaching the observer's eye at each location,  $C_j(\lambda)$ , is the wavelength-by-wavelength product of the illuminant spectral power distribution,  $E(\lambda)$ , and the surface reflectance function at that location,  $S_j(\lambda)$ . We assume that the visual system attempts to parse the color signal into an estimate of illuminant spectral power distribution and surface-reflectance functions, as shown at the lower right.

quantitatively by the statement that a surface-reflectance function cannot exceed 1. In addition, the reflectance functions for physically realizable surfaces must be positive. Given a particular estimate of the common illuminant, there may be image locations where the surface-reflectance function that is consistent with both the illuminant estimate and the color signal is not physically realizable. We link the appearance of luminosity to estimates of illuminant and surface properties by assuming that an image location appears luminous when the color signal cannot be explained by the observer's estimate of the illuminant together with a physically realizable surface. Physical realizability constraints have also been employed in computational color constancy.<sup>28-30</sup>

We do not have direct experimental access to the model constructs described above. Nonetheless, we can use our ideas to develop a parametric model of the measured luminosity threshold surfaces. The parameters of the model are exactly those required to describe the observer's estimate of the illuminant. Given values for these parameters, we can predict the luminosity threshold for any test patch chromaticity. To make the prediction, we start with the chromaticity of a test patch. We then use a search procedure over the space of surface-reflectance functions to find the highest luminance at this chromaticity that is consistent with a physically realizable surface and the illuminant estimate. This luminance is then taken as the predicted luminosity threshold.

We can fit our model to the measured luminosity threshold surfaces for each observer/illuminant condition. We do this by using a second search procedure, this one over the space of illuminants. For each illuminant we compute the predicted luminosity thresholds for all of the test patch chromaticities (using the search procedure described in the preceding paragraph). The second search procedure finds the illuminant that minimizes the sum of

squared errors between the predicted and the measured luminosity thresholds. We call the illuminant found by the second search procedure the equivalent illuminant. To the extent that the predicted thresholds match the data, the equivalent illuminant together with our model provide a compact description of how luminosity thresholds vary as a function of chromaticity.

To constrain the dimensionality of both searches, we use finite-dimensional linear models to represent surface and illuminant spectra.<sup>18</sup> The spectra of naturally occurring surfaces and illuminants are not arbitrary. Rather, they show a great deal of regularity, which may be described by small-dimensional linear models.<sup>31-33</sup> If our visual system has evolved to incorporate regularities exhibited by naturally occurring spectra, then it makes sense to incorporate such regularities in our models.<sup>1,16</sup>

Our model computations are described in detail in Appendix A.

### A. Model Fit

Using the procedure outlined above, we determined the equivalent illuminants for the three observers in each illuminant condition. It is not computationally feasible to search over all possible choices of linear models nor to use linear models with high dimensionality. We have fitted our model with a few different choices of linear models for both surface and illuminant spectra. Of the models tried,<sup>34</sup> the best fit was obtained by use of a four-dimensional linear model for surfaces obtained by principal-components analysis of a set of measured Munsell papers.<sup>18,35,36</sup> and the CIE three-dimensional linear model for daylights.<sup>17</sup>

The solid curves shown in Figs. 4 and 5 are the predicted luminosity thresholds obtained with the selected linear models. The chromaticities of the equivalent illuminants are shown in Fig. 10 along with the chromaticities of the physical illuminants. For observer JMS the recovered illuminants were fairly close to the physical illuminants. For observers DHB and KI there was less agreement, although across conditions the equivalent-illuminant chromaticities shift in rough accordance with the physical-illuminant chromaticities. Recall that be-

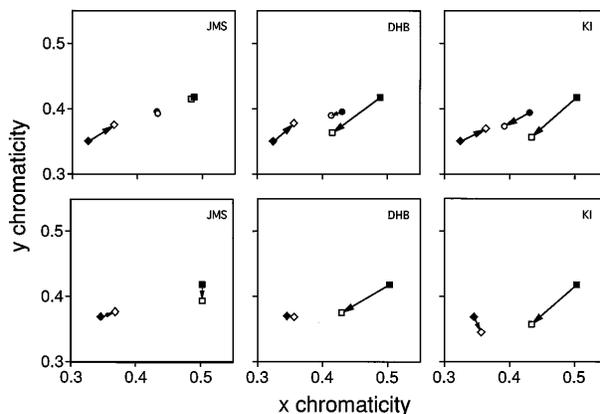


Fig. 10. Equivalent-illuminant chromaticities. Each panel compares the chromaticities of the equivalent illuminants to the physical illuminant for one observer. Open symbols, equivalent-illuminant chromaticities; filled symbols, corresponding physical illuminants. Top panels, chromaticities for conditions Y1, B1, and BY; bottom panels, chromaticities for conditions Y2 and B2.

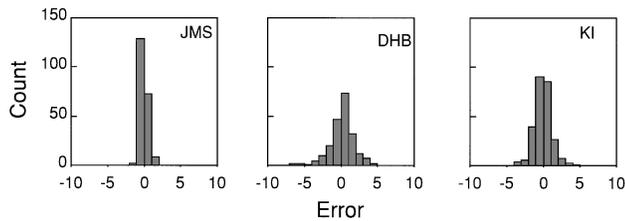


Fig. 11. Each panel shows a histogram of the luminance error between the luminosity thresholds for individual sessions and the equivalent-illuminant model predictions after criterion shifts have been taken into account. The errors for each observer are accumulated across both sessions and conditions. The RMSE's for observers JMS, DHB, and KI were 0.49, 1.70, and 1.17  $\text{cd}/\text{m}^2$ , respectively. For observer KI there were some test patch chromaticities for which he was able to set a luminosity threshold but for which the model predicted that the threshold should be zero. This indicates that the test patch chromaticity was outside the gamut of surfaces realizable within our linear model. We did not include these points in the histogram. Table 6 gives the number of points excluded for each modeling condition. Appendix A contains further discussion of this issue.

cause of criterion shifts we fitted only the relative luminosity thresholds. For this reason, the equivalent illuminants do not have a well-defined luminance.

Figure 11 shows, for each observer, histograms of the error between the model predictions and the measured luminosity thresholds. As described above, we show only the component of the error that cannot be explained by a criterion shift across sessions. We computed this error for each condition by scaling the model predictions to best fit the individual session data. Thus the histograms in Fig. 11 should be compared with those in Fig. 8.

To make the comparison easier, we computed the root-mean-squared error (RMSE) corresponding to each histogram. The first two bars of Fig. 12 show, for each observer, the RMSE corresponding to Fig. 8 (labeled Data precision) and to Fig. 11 (labeled Equivalent illuminant). To evaluate the quality of our model's fit, it is useful to compare our model with several alternatives. Thus the figure also shows the RMSE corresponding to that of four other models. The first two comparison models attempt to explain the appearance of luminosity as the result of simple mechanistic computations that might occur early in the visual pathways. We formulated both of these models so that they had the same number of free parameters as our equivalent illuminant model.

The first comparison model (labeled Linear mechanism) predicts the luminosity thresholds on the assumption that stimuli appear luminous when they excite a linear post-receptor mechanism above a criterion level. To fit this model, we found (for each observer/illuminant condition) the three weights  $w_L$ ,  $w_M$ , and  $w_S$  that best predicted the luminosity thresholds under the assumption that a stimulus appeared luminous when  $w_L L + w_M M + w_S S \geq 1$ , where  $L$ ,  $M$ , and  $S$  represent the cone coordinates of the stimulus ( $L$ , long-;  $M$ , medium-;  $S$ , short-wavelength-sensitive). This model includes as a special case the hypothesis that a stimulus appears self-luminous when its photopic luminance exceeds a criterion threshold. We used the Smith-Pokorny estimates of the cone sensitivities to compute the cone coordinates of the luminosity thresholds.<sup>37,38</sup>

The second comparison model (labeled Evans receptor) is based on Evans's proposal that a region appears

self-luminous when the cone response for at least one cone class exceeds the corresponding cone response to the background.<sup>4</sup> Evans formulated this model to handle only viewing situations in which the patch was viewed on a uniform background. His idea was that the visual system made simple comparisons between the test and the immediately adjacent background. In our experiments there was no simple background. The immediate surround of the test was a thin black border, and the broader surround contained a number of different colored surfaces (see Section 2). We therefore generalized Evans's proposal by assuming that our complex surround had the same effect as some unknown *equivalent background*. The equivalent background plays the same role in our version of Evans's model as the physical background did in the original version. Because the cone responses of the equivalent background are not known, we fitted this model by searching for the equivalent background cone coordinates  $c_L$ ,  $c_M$ , and  $c_S$  that resulted in the best fit to our data set. To calculate the predictions, we simply assumed that a stimulus appeared self-luminous when  $L \geq c_L$ ,  $M \geq c_M$ , or  $S \geq c_S$ , where  $L$ ,  $M$ , and  $S$  represent the cone coordinates of the stimulus.

The third model (labeled Physical illuminant) is identical to our equivalent-illuminant model except that it contains no free parameters for the equivalent illuminant. Rather, we simply assume that the equivalent illuminant is the same as the physical illuminant. The final model (labeled No effect) is based on the assumption that the luminosity thresholds are independent of the test patch chromaticity for each observer/illuminant condition. The RMSE corresponding to this model is a measure of the variance of our data set.

Table 6 provides the RMSE's along with other information useful for model comparison.

A comparison of the RMSE values shows that the equivalent-illuminant model performs better than the other models. For observer JMS the linear-mechanism

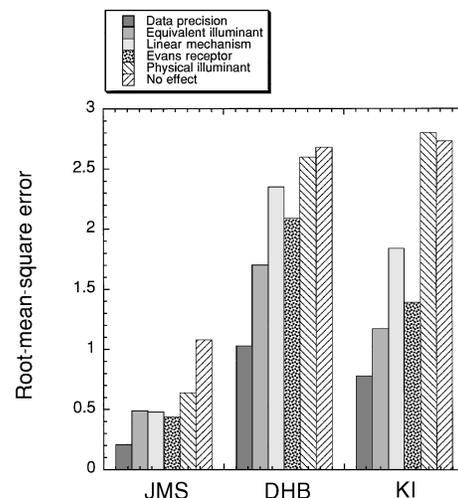


Fig. 12. RMSE's in candelas per square meter for various models. The data-precision model allowed a parameter for each test patch chromaticity. The corresponding RMSE represents the precision of the data after criterion shifts have been taken into account. All other models were fitted to the mean threshold data for each observer/condition as described in the text. Criterion shifts were taken into account by allowing a single scale factor for each individual session.

**Table 6. Summary of Model Fits<sup>a</sup>**

Subject	Model	Points		Parameter	SSE	RMSE
		Measured	Fitted			
JMS	Data precision	212	212	106	9	0.21
	Equivalent illuminant	212	212	21	52	0.49
	Linear mechanism	212	212	21	48	0.48
	Evans receptor	212	212	21	40	0.44
	Physical illuminant	212	212	6	86	0.64
	No model	212	212	6	247	1.08
	Constrained linear	212	212	16	60	0.53
	DHB	Data precision	214	214	114	229
Equivalent illuminant	214	214	20	618	1.70	
Linear mechanism	214	214	20	1180	2.35	
Evans receptor	214	214	20	931	2.09	
Physical illuminant	214	214	5	1452	2.60	
No model	214	214	5	1536	2.68	
Constrained linear	214	214	15	668	1.77	
KI	Data precision	270	270	140	166	0.78
	Equivalent illuminant	270	264	20	364	1.17
	Linear mechanism	270	252	20	856	1.84
	Evans receptor	270	270	20	518	1.39
	Physical illuminant	270	268	5	2108	2.80
	No model	270	270	5	2008	2.73
	Constrained linear	270	265	15	405	1.24

<sup>a</sup>The table provides the number of test patch chromaticities at which luminosity thresholds were obtained. The number of parameters for each model takes into account the criterion shift. For example, the equivalent-illuminant model has three free parameters for each illuminant condition. One of these parameters, however, is redundant to the criterion-shift scale factors. Thus for five illuminant conditions with two sessions per condition, the equivalent illuminant model has ten parameters to describe the five equivalent illuminants and ten parameters to describe the session-by-session criterion shifts. This corresponds to the 20 parameters listed for this model for observers DHB and KI. (Observer JMS ran one additional session, so for him the model has 21 parameters.) Similar calculations lead to the parameter count for the other models.

and Evans-receptor models have slightly lower RMSE's than the equivalent-illuminant model, but for the other two observers the equivalent-illuminant model is substantially better. Of the models considered, the equivalent-illuminant, linear-mechanism, and Evans-receptor models all have the same number of free parameters. Thus the RMSE values for these models are directly comparable. The generally lower RMSE corresponding to the equivalent-illuminant model thus favors it over the other two.

The physical-illuminant and no-effect models have higher RMSE's than the equivalent-illuminant model but have fewer parameters. Because these models are not nested within the equivalent-illuminant model, there is no easily applied goodness-of-fit test to decide whether the differences in RMSE are of statistical significance. (The physical illuminant generated by our apparatus is not within the CIE linear model for daylight that we used to find the equivalent illuminant.) If we ignore the fact that the models are not nested, however, we can apply an  $F$ -ratio goodness-of-fit test. We did so and found that we can reject the null hypothesis that the difference in RMSE is due to chance alone at the  $p < 0.001$  level for all three observers. The  $F$ -ratios we obtained were for JMS,  $F(15,191) = 8.50$ ; for DHB,  $F(15,194) = 17.45$ ; and for KI,  $F(15,250) = 78.35$ . We applied the test as described by Ashby and Lee.<sup>39</sup> For observer KI the application of the test is complicated by the fact that not all of the measured points are within the model gamut. We corrected for this by assigning a squared error for the out-of-model gamut points equal to the mean of the

squared error for the points that were in gamut. This issue is discussed at the end of Appendix A.

The equivalent-illuminant model fits better than the other models evaluated, but it does not describe the data to the measurement precision. Our data-precision measure may be thought of as the RMSE for a model containing a free parameter for every test patch chromaticity. To ask whether the difference between the equivalent-illuminant and the data-precision RMSE's was due to chance alone, we applied an  $F$ -ratio goodness-of-fit test. Here the models are nested, as any prediction made by the equivalent-illuminant model can be replicated by the data-precision model. Again we were able to reject the hypothesis that the differences in RMSE were due to chance alone at the  $p < 0.001$  level for JMS, at the  $p < 0.01$  level for DHB, and at the  $p < 0.05$  level for KI. The  $F$ -ratios obtained were for JMS,  $F(85,106) = 5.98$ ; for DHB,  $F(94,100) = 1.81$ ; and for KI,  $F(114,130) = 1.35$ . We again corrected for KI's out-of-gamut points during performance of this test. The fact that there is regularity in the data that is not described by the equivalent-illuminant model is also suggested by the plots of the model fits shown in Figs. 4 and 5. Although the model captures the qualitative features of the data, it misses in detail. One possible reason is that to fit the model to the data, we had to choose specific linear models for surface reflectances and illuminant spectral power distributions. This *a priori* choice affects the model fits. As noted above, we did try some other combinations of surface and illuminant linear models. It seems likely, however, that some other choice of linear models would

provide a better fit, although how much better is difficult to evaluate. Computational limitations prohibit an extensive search over the spaces of linear models.

### B. Illuminant Linearity

Although the equivalent-illuminant model does not provide a complete description of the data, we believe that it captures enough of the variance to warrant its use for addressing questions about how luminosity thresholds vary with the illumination. To demonstrate this, we investigate whether the equivalent illuminant depends linearly on the physical illuminant. This is a hypothesis that has been successfully used to understand how surface color appearance varies with the illumination.<sup>40</sup> The basic idea is as follows. Suppose that we measure equivalent illuminants  $\hat{E}_1(\lambda)$  and  $\hat{E}_2(\lambda)$  under two physical illuminants  $E_1(\lambda)$  and  $E_2(\lambda)$ . If linearity holds, then when the physical illuminant is any linear combination

$$E(\lambda) = aE_1(\lambda) + bE_2(\lambda) \quad (1)$$

(with  $a$  and  $b$  any scalars), the corresponding equivalent illuminant will be given by

$$\hat{E}(\lambda) = a\hat{E}_1(\lambda) + b\hat{E}_2(\lambda) \quad (2)$$

for the same scalars  $a$  and  $b$ . The simplifying advantage of this illuminant linearity is that the equivalent illuminants corresponding to a wide range of physical illuminants may be predicted from measurements of the equivalent illuminant for a small number of physical illuminants. Since the equivalent illuminants in turn predict the luminosity thresholds, illuminant linearity provides the promise of a very compact description of a large number of measurements. This sort of simplification is especially attractive because the range of naturally occurring illuminants may in fact be described as linear combinations of a small number of basis illuminants.<sup>17,18,32</sup>

Our five physical illuminants may be accurately described as a linear combination of two basis illuminants. That is, we can describe all of our experimental illuminants with Eq. (1), where  $E_1(\lambda)$ ,  $E_2(\lambda)$ , and the five scalar pairs  $(a, b)$  are known. If linearity holds, we should be able to find  $\hat{E}_1(\lambda)$  and  $\hat{E}_2(\lambda)$  such that when they are used in Eq. (2) with the five scalar pairs  $(a, b)$ , we get equivalent illuminants that provide a good description of our threshold data. For each observer we repeated our search procedure with the added constraint that the equivalent illuminants satisfy illuminant linearity. Adding the linearity constraint does not substantially affect the RMSE error. Figure 13 compares the RMSE's of the fits to the luminosity threshold data with and without the linearity constraint. An  $F$ -ratio goodness-of-fit test does, however, indicate that the difference between the constrained and the unconstrained fits is statistically significant ( $p < 0.001$  for JMS and KI,  $p < 0.01$  for DHB). The  $F$ -ratios we obtained were for JMS,  $F(5,191) = 6.10$ ; for DHB,  $F(5,194) = 3.12$ ; and for KI,  $F(5,250) = 5.34$ . Although this test indicates that illuminant linearity does not hold perfectly, we conclude from the comparison of RMSE's that the violation is very small. Indeed, to the precision at which the unconstrained equivalent-illuminant model describes the data, illuminant linearity holds quite well.

## 5. SUMMARY AND DISCUSSION

In this paper we extended previous measurements of the perception of luminosity by having observers set luminosity thresholds under quite natural viewing conditions. In addition, we developed a physics-based model for luminosity thresholds and showed that this model provides a reasonable description of the data.

Our experiments show that observers can set luminosity thresholds. A disappointing feature of our data is that there was considerable criterion shift across sessions (for some observers) and that the agreement among observers was poor. Perhaps a sharper specification of the observer's task or more practice would allow collection of more consistent data. In his classic studies, Evans acted as his own primary observer and made fine distinctions among different forms of non-surface-mode appearance.<sup>6-8</sup> We were initially hesitant to impose these distinctions on our observers, because they did not strike us as natural at the time. A direction for future research clearly is to find methods for obtaining more-precise measurements. It would also be useful to modify our apparatus to allow measurements for a wider gamut of chromaticities.

Our model generalizes previous methods for calculating the gamut of physically realizable surfaces (often called optimal surface stimuli) by including linear-model constraints on the surface-reflectance functions.<sup>23,41-43</sup> For this reason alone it may be of interest to color theorists, independent of its value for describing human performance. For example, the Deutsches Institut für Normung (DIN) color-order system uses optimal surface stimuli as a reference standard for determining the darkness of colored samples.<sup>44</sup> Our calculation provides a method to generate similar standards that take the regularities of natural surface-reflectance spectra into account. Previously such regularities could be incorpo-

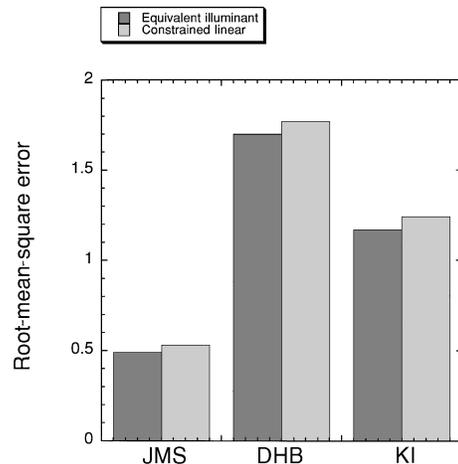


Fig. 13. RMSE's for observers JMS, DHB, and KI with and without the linearity constraint. Both the general and the constrained linear models were fitted to the mean threshold surfaces of each condition. For the constrained linear model the fitting procedure required the set of equivalent illuminants to satisfy the illuminant-linearity constraint. The RMSE's were 0.49, 1.70, and 1.17  $\text{cd}/\text{m}^2$  without the linearity constraint and 0.53, 1.77, and 1.24  $\text{cd}/\text{m}^2$  with the constraint for observers JMS, DHB, and KI, respectively.

rated only with brute-force methods, such as examining the physical gamut of a large set of measured samples.<sup>45</sup>

Our approach to understanding luminosity thresholds has much in common with the thinking of other investigators. In his early work Evans compared fluorence thresholds with the gamut of physically realizable surfaces under the illuminant specified by his stimulus background.<sup>4</sup> To make this comparison he used MacAdam's method for computing the surface gamut.<sup>23,41-43</sup> This approach is essentially the one implemented in the physical-illuminant model above. The main difference is that we computed the surface gamut under a linear-model constraint. Recall that the physical-illuminant model did not provide a good fit to our data. Our equivalent-illuminant model extends Evans's idea by allowing free parameters for the observer's illuminant estimate. Ikeda and co-workers also used this general approach to model judgments of luminosity.<sup>11</sup> They compared their data to the gamut of Munsell papers<sup>45</sup> but found only approximate agreement. Like Evans, they did not allow for the possibility that the observer misestimated the illuminant.

A notion similar to that of our equivalent illuminant was used recently by Bonato and Gilchrist<sup>10</sup> to facilitate understanding of their measurements of achromatic luminosity thresholds. They found that luminosity thresholds varied with the physical illumination and that the thresholds were well predicted with a model that incorporated a parameter to account for which stimulus the observer judged to be white. This parameter plays the same role in their thinking as the equivalent illuminant does in ours. A similar approach has also been taken by Ikeda and co-workers.<sup>12</sup> Ullman's<sup>9</sup> approach to understanding luminosity thresholds was also in terms of a physics-based model. Rather than emphasizing explicit estimates of surfaces and illuminants, however, he attempted to predict the thresholds on the basis of quantities that should be invariant in the absence of self-luminous objects.

One final approach has been used that aids in understanding judgments such as ours. Evans and Swenolt<sup>6-8</sup> related their fluorence threshold data to measures of the chromatic strength (e.g., saturation) of their test stimuli. They argued that these two types of appearance judgment were mediated by the same visual mechanisms. We have not explored this intriguing idea, but to us it suggests the possibility that saturation might be understood by consideration of a stimulus's relation to the gamut of physically realizable surfaces.

One feature of our model may puzzle those familiar with computational color constancy. In principle, both the relative spectral power distribution and the absolute intensity of the equivalent illuminant play an important role in our model calculations. Strictly speaking, however, the absolute intensity of the illuminant can be traded off against the absolute reflectance of the surfaces in the image. Halving the reflectance of every surface and doubling the illuminant intensity leaves the array of receptor responses unchanged. In light of this observation it might seem unreasonable to suppose that the visual system has any representation of the absolute illuminant intensity. Recent computational approaches to color constancy, however, have employed Bayesian meth-

ods that incorporate probabilistic descriptions of which surfaces and illuminants are most likely to occur.<sup>30,46,47</sup> In the Bayesian context it is possible to make a reasonable estimate both of the relative spectral power distribution and of the overall illuminant intensity.<sup>46,47</sup> In practice, we modeled only the relative shape of the luminosity threshold surfaces. For this reason the absolute intensity of the equivalent illuminant was undefined in our model fits. In the context of the above discussion, the criterion shifts that we observed could be interpreted as a consequence of uncertainty in the visual system's estimate of the absolute intensity of the illuminant.

Finally, it is worth noting that our equivalent-illuminant model has a character different from that of the more mechanistic models (i.e., the linear-mechanism and Evans-receptor models) with which we compared it. Our model is based on the assumption that the visual system has internalized the physics of reflectance. It models shifts in the luminosity thresholds with a calculation that explicitly represents both linear-model and physical-realizability constraints on naturally occurring spectra. The comparison models, on the other hand, make no reference to the nature of the external world. Rather, they are based on the action of early visual mechanisms. We believe that a complete theory of luminosity thresholds must include both a mechanistic and a computational explanation. The superior fit of our equivalent-illuminant model suggests that to develop a mechanistic explanation we may do well to consider how the action of visual mechanisms might approximate the calculations that our model embodies.<sup>48</sup>

## APPENDIX A: MODEL IMPLEMENTATION

To implement our model, we require a method to calculate the luminosity threshold given an equivalent illuminant. We will use the general linear-model-based colorimetric techniques described by Brainard.<sup>18</sup> Let  $s$  be an  $N_\lambda$ -dimensional column vector whose entries specify a surface reflectance function sampled at  $N_\lambda$  evenly spaced wavelengths throughout the visible spectrum. The vector  $\mathbf{s}$  lies within an  $N_s$ -dimensional linear model if we can write

$$\mathbf{s} = \mathbf{B}_s \mathbf{w}_s, \quad (\text{A1})$$

where  $\mathbf{B}_s$  is a fixed  $N_\lambda \times N_s$ -dimensional matrix. We call  $\mathbf{B}_s$  the basis matrix of the linear model; we call the columns of  $\mathbf{B}_s$  the basis vectors of the linear model. A linear model is specified completely by its basis matrix. Once a basis matrix for surface spectra has been chosen, we can specify any vector within the linear model by its vector of weights  $\mathbf{w}_s$ . In parallel fashion, let  $\mathbf{e}$  be an  $N_\lambda$ -dimensional column vector whose entries specify illuminant power sampled at  $N_\lambda$  evenly spaced wavelengths throughout the visible spectrum. The vector  $\mathbf{e}$  lies within an  $N_e$ -dimensional linear model if we can write

$$\mathbf{e} = \mathbf{B}_e \mathbf{w}_e, \quad (\text{A2})$$

where  $\mathbf{B}_e$  is a fixed  $N_\lambda \times N_e$ -dimensional matrix. In our calculations we assume that both surface and illuminant spectra are constrained to lie within linear mod-

els and that the basis matrices for the two linear models are known.

Given a particular equivalent illuminant specified by  $\mathbf{w}_e$  (within the linear model defined by  $\mathbf{B}_e$ ) and a test patch chromaticity  $(x, y)$ , we must calculate the highest color-signal luminance consistent with a physically realizable surface within the linear model defined by  $\mathbf{B}_s$ . This can be cast as a linear programming problem. We seek to maximize a linear function of the weights  $\mathbf{w}_s$  subject to linear equality and inequality constraints.

Let the  $3 \times N_\lambda$ -dimensional matrix  $\mathbf{T}$  be the matrix whose rows contain the CIE 1931 color-matching functions sampled at evenly spaced wavelengths throughout the visible spectrum, and let  $\mathbf{T}_y$  be the  $N_\lambda$ -dimensional row vector specifying the photopic luminosity function. The tristimulus coordinates  $r$  of the color signal reflected from a surface  $\mathbf{w}_s$  under an illuminant  $\mathbf{w}_e$  are given by

$$\mathbf{r} = \mathbf{T} \text{diag}(\mathbf{B}_e \mathbf{w}_e) \mathbf{B}_s \mathbf{w}_s, \quad (\text{A3})$$

where the expression  $\text{diag}(\mathbf{B}_e \mathbf{w}_e)$  indicates the diagonal matrix with the entries of  $\mathbf{B}_e \mathbf{w}_e$  along its main diagonal. Similarly, the luminance  $Y$  of the color signal is given by

$$Y = \mathbf{T}_y \text{diag}(\mathbf{B}_e \mathbf{w}_e) \mathbf{B}_s \mathbf{w}_s. \quad (\text{A4})$$

Equation (A4) implies that the stimulus luminance  $Y$  is a linear function of  $\mathbf{w}_s$ . We want to choose  $\mathbf{w}_s$  to maximize the luminance subject to two constraints. First, the chromaticity of the color signal must be at a specified value. Second, the surface reflectance function must be physically realizable. Both of these constraints may be expressed as linear equality or inequality constraints on  $\mathbf{w}_s$ .

Let  $\mathbf{r}_0$  be the unique tristimulus vector with unit length and chromaticity coordinates  $(x, y)$ . Let  $\text{null}(\mathbf{r}_0)$  be a  $2 \times 3$  matrix whose rows form a basis set for the subspace orthogonal to  $\mathbf{r}_0$ . The projection of  $\mathbf{r}_0$  onto its orthogonal subspace is zero:  $\text{null}(\mathbf{r}_0) \mathbf{r}_0 = \mathbf{0}$ . Any tristimulus vector  $\mathbf{r}$  that satisfies  $\text{null}(\mathbf{r}_0) \mathbf{r} = \mathbf{0}$  will have the same chromaticity  $(x, y)$  as  $\mathbf{r}_0$ , and all tristimulus vectors with chromaticity  $(x, y)$  satisfy  $\text{null}(\mathbf{r}_0) \mathbf{r} = \mathbf{0}$ . Thus we can express the chromaticity constraint as a linear constraint on  $\mathbf{w}_s$ :

$$\text{null}(\mathbf{r}_0) \mathbf{T} \text{diag}(\mathbf{B}_e \mathbf{w}_e) \mathbf{B}_s \mathbf{w}_s = \mathbf{0}. \quad (\text{A5})$$

The constraints that the surface reflectance be physically realizable are expressed as the linear inequalities

$$\mathbf{B}_s \mathbf{w}_s \geq \mathbf{0}, \quad \mathbf{B}_s \mathbf{w}_s \leq \mathbf{1}. \quad (\text{A6})$$

Thus finding the maximum luminance consistent with a given equivalent illuminant  $\mathbf{w}_e$  and color signal chromaticity  $(x, y)$  is equivalent to the linear programming problem of finding the surface-weight vector  $\mathbf{w}_s$  that maximizes the linear function of Eq. (A4) subject to the linear constraints of Eqs. (A5) and (A6). We used the linear programming routine `lp` provided as part of the `MATLAB` Optimization Toolbox to implement the maximization.<sup>49,50</sup>

To find the equivalent illuminant that provided the best fit to the data, we implemented a numerical search procedure

over the equivalent illuminant  $\mathbf{w}_e$ . For each choice of  $\mathbf{w}_e$  we computed the predicted luminosity thresholds for all test patch chromaticities, using the linear programming technique described above. This allowed us to compute the sum of squared errors between the data and the prediction corresponding to any choice of  $\mathbf{w}_e$ . We searched for the  $\mathbf{w}_e$  that minimized this error.

One subtlety is worth noting. For some choices of  $\mathbf{w}_e$  and test patch chromaticity there will be no feasible solution to the linear programming problem. In these cases the model predicts that the test patch should appear luminous at any luminance. Physically this corresponds to the test patch chromaticity being outside the gamut of physically realizable surfaces that are within the linear model defined by  $\mathbf{B}_s$ . In computing the fit penalty in this case, we took the predicted luminosity threshold to be zero. This choice made the error for such points large and tended to drive the search toward choosing equivalent illuminants for which all measured points were within gamut. In some cases, however, the final equivalent illuminant resulted in out-of-gamut predictions for some test chromaticities. We excluded such points from the error histograms shown in Fig. 11. When we computed SSE's and RMSE's, we took the error for such points to be the average of the error for within-gamut points. This was necessary only for a small number of points for observer KI. We believe that it gives a more accurate picture of model performance, since presumably small modifications in the linear model for surfaces would have put all points within gamut without drastically changing the overall fit.

## ACKNOWLEDGMENTS

We thank L. Arend, M. D'Zmura, A. Gilchrist, K. Ishigami, A. T. Young, and an anonymous reviewer for useful discussions and suggestions. We also thank J. Tietz for his role in apparatus and software development. K. Ishigami observed in the experiments reported here. P. Wise and L. Markan observed in pilot experiments. This work was supported by National Eye Institute grant EY10016.

## REFERENCES AND NOTES

1. D. H. Brainard, B. A. Wandell, and E.-J. Chichilnisky, "Color constancy: from physics to appearance," *Curr. Dir. Psychol. Sci.* **2**, 165–170 (1993).
2. K. Buhler, "Gegenbemerungen," *Psychol. Forsch.* **5**, 182–188 (1924), as discussed in Ref. 3.
3. G. Kreezer, "Luminous appearances," *J. Gen. Psychol.* **4**, 247–281 (1930).
4. R. M. Evans, "Fluorescence and gray content of surface colors," *J. Opt. Soc. Am.* **49**, 1049–1059 (1959).
5. R. M. Evans, "Variables of perceived color," *J. Opt. Soc. Am.* **54**, 1467–1474 (1964).
6. R. M. Evans and B. K. Swenholt, "Chromatic strength of colors: dominant wavelength and purity," *J. Opt. Soc. Am.* **57**, 1319–1324 (1967).
7. R. M. Evans and B. K. Swenholt, "Chromatic strength of colors. Part II. The Munsell system," *J. Opt. Soc. Am.* **58**, 580–584 (1968).
8. R. M. Evans and B. K. Swenholt, "Chromatic strength of colors. III. Chromatic surrounds and discussion," *J. Opt. Soc. Am.* **59**, 628–634 (1969).
9. S. Ullman, "On visual detection of light sources," *Biol. Cybern.* **21**, 205–212 (1976).

10. F. Bonato and A. L. Gilchrist, "The perception of luminosity on different backgrounds and in different illuminations," *Perception* **23**, 991–1006 (1994).
11. M. Ikeda, K. Motonaga, N. Matsuzawa, and T. Ishida, "Threshold determination for unnatural color appearance with local illumination," *Kogaku* **22**, 289–298 (1993).
12. M. Ikeda, S. Fukumura, N. Matsuzawa, and T. Ishida, "Influence of surrounding visual information on the recognition threshold of local illumination," *Kogaku* **23**, 42–49 (1994).
13. Our usage of the term luminosity threshold follows that of Kreezer<sup>3</sup> and of Bonato and Gilchrist.<sup>10</sup> The term luminosity is used in the literature both to describe physical properties of light<sup>14</sup> and to describe perceptual experience.<sup>15</sup> Ideal terminology would distinguish these two uses, but at present we feel that less confusion will be generated by our use of luminosity threshold than if we attempt to define new terminology.
14. D. L. MacAdam, *Color Measurement: Theme and Variations* (Springer-Verlag, New York, 1981).
15. D. Judd and G. Wyszecki, *Color in Business, Science, and Industry* (Wiley, New York, 1975).
16. R. N. Shepard, "The perceptual organization of colors: an adaptation to regularities of the terrestrial world?" in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. H. Barkow, L. Cosmides, and J. Tooby, eds. (Oxford U. Press, New York, 1992).
17. CIE, *Colorimetry*, 2nd ed. (Bureau Central de la CIE, Paris, 1986).
18. D. H. Brainard, "Colorimetry," in *Handbook of Optics: Vol. 1. Fundamentals, Techniques, and Design*, M. Bass, ed. (McGraw-Hill, New York, 1995).
19. J. M. Speigle and D. H. Brainard, "Fluorescence thresholds depend on the illumination," *Invest. Ophthalmol. Vis. Sci. Suppl.* **35**, 1656 (1994).
20. L. T. Maloney and B. Wandell, "A computational model of color constancy," *Invest. Ophthalmol. Vis. Suppl.* **26**, 206 (1985).
21. M. D'Zmura and G. Iverson, "Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces," *J. Opt. Soc. Am. A* **10**, 2148–2165 (1993).
22. E. Hering, "Der Raumsinn und die Bewegungen des Auges," in *Handbuch der Physiologie*, B. L. Hermann, ed. (Vogel, Leipzig, 1879), Vol. 3, Part 1, as discussed in Ref. 3 above.
23. D. L. MacAdam, "The theory of the maximum visual efficiency of colored materials," *J. Opt. Soc. Am.* **25**, 249–252 (1935).
24. A. L. Gilchrist, S. Delman, and A. Jacobsen, "The classification and integration of edges as critical to the perception of reflectance and illumination," *Percept. Psychophys.* **33**, 425–436 (1983).
25. A. L. Gilchrist, "Lightness contrast and failures of constancy: a common explanation," *Percept. Psychophys.* **43**, 415–424 (1988).
26. G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.* **310**, 1–26 (1980).
27. D. H. Brainard and W. T. Freeman, "Bayesian method for recovering surface and illuminant properties from photoreceptor responses," in *Human Vision, Visual Processing, and Digital Display V*, B. E. Rogowitz and J. P. Allebach, eds., *Proc. Soc. Photo-Opt. Instrum. Eng.* **2179**, 364–376 (1994).
28. D. Forsyth, "A novel algorithm for color constancy," *Int. J. Comput. Vision* **5**, 5–36 (1990).
29. G. D. Finlayson, "Color constancy in diagonal chromaticity space," in *Proceedings of the 5th International Conference on Computer Vision* (IEEE, Cambridge, Mass., 1995), pp. 218–223.
30. M. D'Zmura, G. Iverson, and B. Singer, "Probabilistic color constancy," in *Geometric Representations of Perceptual Phenomena: Papers in Honor of Tarow Indow's 70th Birthday*, R. D. Luce, M. D'Zmura, D. Hoffman, G. Iverson, and A. K. Romney, eds., (Erlbaum, Hillsdale, N.J., 1995).
31. J. Cohen, "Dependency of the spectral reflectance curves of the Munsell color chips," *Psychon. Sci.* **1**, 369–370 (1964).
32. D. B. Judd, D. L. MacAdam, and G. W. Wyszecki, "Spectral distribution of typical daylight as a function of correlated color temperature," *J. Opt. Soc. Am.* **54**, 1031–1040 (1964).
33. L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *J. Opt. Soc. Am. A* **3**, 1673–1683 (1986).
34. The other combinations of linear models we tried were (1) the CIE three-dimensional daylight model together with a three-dimensional linear model for the Munsell papers, (2) the CIE three-dimensional daylight model together with a six-dimensional linear model for the Munsell papers, and (3) a two-dimensional linear model for illuminants obtained from a principal-components analysis of our experimental illuminants together with the four-dimensional linear model for Munsell papers. These choices did not provide a better fit than the combination we used. We evaluated only choice (2) on a subset of our data, because it was computationally expensive to compute the model fits for a six-dimensional-surface linear model.
35. K. L. Kelly, K. S. Gibson, and D. Nickerson, "Tristimulus specification of the Munsell Book of Color from spectrophotometric measurements," *J. Opt. Soc. Am.* **33**, 355–376 (1943).
36. D. Nickerson, "Spectrophotometric data for a collection of Munsell samples" (U.S. Department of Agriculture, Washington, D.C., 1957; available from Munsell Color Company, Baltimore, Md.).
37. V. Smith and J. Pokorny, "Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm," *Vision Res.* **15**, 161–171 (1975).
38. P. DeMarco, J. Pokorny, and V. C. Smith, "Full-spectrum cone sensitivity functions for X-chromosome-linked anomalous trichromats," *J. Opt. Soc. Am. A* **9**, 1465–1476 (1992).
39. F. G. Ashby and W. W. Lee, "Predicting similarity and categorization from identification," *J. Exp. Psychol. Gen.* **120**, 150–172 (1991).
40. D. H. Brainard and B. A. Wandell, "Asymmetric color-matching: how color appearance depends on the illuminant," *J. Opt. Soc. Am. A* **9**, 1433–1448 (1992).
41. E. Schrodinger, "Theorie der pigmente von grosster Leuchtkraft," *Ann. Phys.* **62**, 603–622 (1920), as discussed in Ref. 50 below.
42. S. Rosch, "Die Kennzeichnung der Farben," *Phys. Z.* **29**, 83–91 (1928), as discussed in Ref. 50 below.
43. G. Wyszecki and W. S. Stiles, *Color Science—Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. (Wiley, New York, 1982).
44. M. Richter and K. Witt, "The story of the DIN color system," *Color Res. Appl.* **11**, 138–145 (1986).
45. M. R. Pointer, "The gamut of real surface colours," *Color Res. Appl.* **5**, 145–155 (1980).
46. W. T. Freeman and D. H. Brainard, "Bayesian decision theory, the local mass estimate, and color constancy," in *Proceedings of the 5th International Conference on Computer Vision* (IEEE, Cambridge, Mass., 1995), pp. 210–217.
47. D. Marr, *Vision*, (Freeman, San Francisco, 1982).
48. J. Little and C. Moler, *MATLAB User's Guide* (The MathWorks, Natick, Mass., 1991).
49. A. Grace, *Optimization Toolbox for Use with MATLAB User's Guide* (The MathWorks, Natick, Mass., 1990).
50. G. Wyszecki, "Color Appearance," in *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, eds. (Wiley, New York, 1986).