PNAS

Supporting Information for
Perceptual and neural constraints on photometric measures of heterochromatic brightness

Shuchen Guan, Jing Chen, Robert Ennis, Matteo Toscani, Matteo Valsecchi, Andrea van Doorn, Jan Koenderink, Karl R. Gegenfurtner

Karl R. Gegenfurtner
Email: gegenfurtner@uni-giessen.de

This PDF file includes:

Supplementary Materials and Methods
Figures S1 to S10
SI References

Supplementary Materials and Methods

Behavioral Ranking - Laboratory Experiments

Observers. Data were collected from 45 naïve observers (28 females, 25.3 ± 4.1 years old). One of them was excluded for further analysis due to low accuracy in the catch trial (<75%). 21 observers returned for a retest experiment after six months or more than six months. Observers were students or employees of Giessen University, Germany. They had normal or corrected-to-normal vision and passed the Ishihara color blindness test (1). Informed consent was obtained from each observer before data collection. All the tasks complied with the Helsinki Declaration and were approved by the local ethics committee (Giessen LEK 2020-0015).

Setup. Stimuli were presented on a calibrated Sony OLED display (model number: PVM-2541A, 54.4 × 30.6 cm, 1920 × 1080, 60 Hz), calibrated with a Konica Minolta Spectroradiometer CS-2000A (Konica Minolta Holdings Inc., Marunouchi, Tokyo, Japan) and set to sRGB simulation mode. Its gamut at CIE 1931 xyY coordinate (2) was: R = [0.63, 0.33, 28.89], G = [0.30, 0.60, 98.47], B = [0.15, 0.06, 11.64], with gamma values of 2.01, 2.24, and 1.82. Stimuli were gamma-corrected to a linear scale. Observers sat in a dimly lit room, ~65 cm from the display.

Visual stimuli. Color patches were randomly selected from 144 colors, varying in hue and intensity (see Fig. 1A). The 12 hues included three primary (red [1,0,0], green [0,1,0], blue [0,0,1]), three secondary (yellow [1,1,0], cyan [0,1,1], and magenta [1,0,1]), and six intermediate ([1,0.5,0], [0.5,1,0], [0,1,0.5], [0,0.5,1], [1,0,0.5] and [0.5,0,1]) hues. Intensities linearly ranged from 10% to 80% in 12 steps. Each trial presented 12 square patches (4.6×4.6°) on a black background in a 4×3 array, covering all hues and intensities. Observers ranked them from brightest to darkest by clicking sequentially; selected patches disappeared. They could move their eyes freely and had no time limits. The experiment began with six practice trials, followed by 66 trials, including six catch trials with 12 intensity levels that were monochromatic to avoid conflicting brightness-metric predictions. Stimuli were identical across observers, but patch locations and trial order were randomized. The experiment lasted 30–45 minutes. Some observers returned for an identical retest session to assess experimental reliability. The stimuli and task were programmed in TypeScript using the p5.js library and were presented in a fullscreen WebGL context. We validated that all major browsers, across MacOS, Linux, and Windows and across Nvidia, Apple Silicon, and AMD GPUs, presented colors equivalently and correctly in this context.

Model comparison. Correlation analyses were conducted to evaluate the performance of models in predicting brightness rankings. Model-predicted values were calculated for each color patch by transforming the stimulus RGB values to XYZ, from which luminance and the outputs of other models such as CAM16 were computed. Spearman correlations were then computed between these model predictions and the mean brightness rankings of the patches, averaged across trials and observers. The unexplained variance was defined as $1-R^2$, where R^2 represents the squared correlation coefficient. These models include **Luminance** ($cd \cdot m^{-2}$)(3), **L*** in CIELAB (4), which applies a nonlinear transformation to better approximate human brightness perception, and the **CIE Brightness Matching Model** (5), which is based on observer's matching brightness across colors and extends beyond luminance. Several more perceptually grounded models were also evaluated, including **Guth's Vector Model** (6), the **Nayatani Model** (7), and the **Hunt Model** (8), incorporating opponent color processing, adaptation mechanisms, and contextual influences. In addition, **Radiance** ($W \cdot sr^{-1} \cdot m^{-2}$), a measure of emitted or reflected energy based on spectral composition without weighting, was included.

More recent color appearance models were also tested, including ***Q*** in **CIECAM16** (9) and **CIECAM02** (10). Several modified models, primarily based on CIECAM02 or CIECAM16 and designed to incorporate the Helmholtz-Kohlrausch effect, like the brightness in **sCAM** (11), the lightness in **sUCS** (11), the **High Model** (12), the **Hellwig Model** (13), **J** in **Kim's Model** (14), **Q** in **Park's Model** (15), **L*** modified by Liao et al (16), and **L**** in **Fairchild & Pirrotta's Model** (17). Beyond these models, a more generalized **Minkowski metric** was considered:

I* = $((\boldsymbol{w_R} \cdot \boldsymbol{R})^{\boldsymbol{n}} + (\boldsymbol{w_G} \cdot \boldsymbol{G})^{\boldsymbol{n}} + (\boldsymbol{w_B} \cdot \boldsymbol{B})^{\boldsymbol{n}})^{\frac{1}{\boldsymbol{n}}}$

The weights w_R, w_G and w_B determine the relative weights on the three display primaries *R*, *G* and *B.* This analysis focused on two key variations: a linear variation and a nonlinear variation, determined by the exponent of ***n***. When ***n*** equals 1, the weighted sumRGB: $\mathrm{Sum}\ [w_R \cdot R, w_G \cdot G, w_B \cdot B]$ is derived. When ***n*** approaches infinity, the weighted maxRGB (wMaxRGB): $\mathrm{Max}\ [w_R \cdot R, w_G \cdot G,\ w_B \cdot B]$ is derived. The weighted sumRGB and wMaxRGB were then fitted to both the mean and individual brightness rankings to obtain the optimal w_R, w_G and w_B by minimizing the unexplained variance. The optimal w_R, w_G and w_B were also used to examine the transition from linear to nonlinear rule by calculating the unexplained variance for increasing exponent ***n***. Similar weight fitting procedures were applied in the LMS and DKL color spaces(18, 19). For each color patch, the three component values in each color space were calculated and normalized separately to the 0–1 range for each of the three dimensions before performing the weight fitting.

All data analysis was conducted using MATLAB 2022b (The Mathworks, Natick, MA, USA).

<u>Behavioral Ranking - Online Experiments</u>

To extend the experimental sample, we conducted online experiments. To assess the reliability of the online experiment, the same group of observers from the laboratory experiment completed the same task at home using their own computers and displays. Data collection was then expanded to a larger group of online observers on Prolific (www.prolific.com), with different background colors tested across separate sub-groups.

Observers. Four groups of observers participated in the online data collection. **I**) For the at-home experiment, data were collected from the same 45 observers as in the laboratory experiment, with one excluded from further analysis due to low accuracy in the catch trials. **II**) In the Prolific experiment, data were collected from three groups of observers with patches presented on different background colors: 186 (52 females, 124 males, 10 unspecified; mean age 32.3 ± 9.6 years) for the black background, 146 (44 females, 100 males, 2 unspecified; 38.8 ± 11.5 years) for the gray background, and 154 (60 females, 90 males, 4 unspecified; 39.8 ± 13.3 years) for the white background. Observers that did not finish the experiment or with low catch-trial accuracy were excluded: 4 from the black, 25 from the gray, and 16 from the white background conditions. Beforehand, they were asked whether they were color-deficient (those who were color-deficient should not continue the experiment), and informed consents were obtained from them before data collection. All tasks complied with the Helsinki Declaration and were approved by the local ethics committee (Giessen LEK 2020-0015).

Setup. Stimuli were presented on the observers' own computers and displays at home. Observers were instructed to turn off the automatic brightness adjustment and the night shift feature on their displays. **I**) Observers in the home experiment brought their own displays or laptops to the laboratory afterwards, where we measured these displays using the Konica Minolta Spectroradiometer CS-2000A. The gamuts (fig. S2A) and gammas, the brands and models of 43 displays were recorded (one observer's display was unavailable). **II**) For the online experiment on Prolific, we were unable to measure the displays but assumed an OLED display that was used for the laboratory experiment. Observers completed the experiment without any lighting limitations.

Visual stimuli. **I**) For the at-home experiment, stimuli were identical to those used in the laboratory experiment, with gamma-uncorrected (nonlinear) RGB values of these patches matching those used in the laboratory experiment. The patches were square, with a width and height approximately 10% of the display width. **II**) For the online experiment on Prolific, an attentional check question was added to ensure the observers were paying attention and understood the instructions. The color patches remained the same, but were presented with different background colors: black, mid-gray (intensity: 0.465) and white. Each background color was tested to a separate group of observers. The same TypeScript/p5.js code was used as before, again running in a full screen WebGL context.

Data analysis. **Display Comparison.** For observers who participated in both the laboratory and home experiments, their displays (N = 43, one observer's display was unavailable) were measured. The resulting display gamuts were then compared with that of the OLED display used in the laboratory experiment. To characterize differences between the observers' displays and the OLED, we calculated the Luminance, Radiance, wMaxRGB, and weighted sumRGB of 144 color patches based on each display's gamut and gammas. The patches were then paired, generating 10,296 pairs. For each pair, predictor values were used to rank the two patches. To compare the home displays with the OLED, the percentage of pairs that maintained a consistent ranking was calculated for each predictor. This resulted in a distribution of consistency values across all displays for each predictor (fig. S2E). **Ranking Comparison.** The basic data analysis for both the home experiment and the online experiment on Prolific followed the same procedure as the laboratory experiment. The RGB values of the patches used in all experiments were based on the assumption of an OLED display. To assess ranking consistency, Spearman correlations between the laboratory and home experiments were calculated both at the individual and group (averaged) levels. A similar analysis was conducted for the online experiment on Prolific. **Model Prediction Comparison.** All models used in the laboratory experiment were also evaluated on the home and online datasets. The RGB weights for the wMaxRGB and weighted sumRGB models were fitted separately for each experiment.

<u>Behavioral Ranking - Chroma Variation</u>

In chroma-variation experiment, we expanded our stimulus set to the full RGB cube. This experiment was conducted in two stages: In the first stage, a group of observers completed the same experiment both in the laboratory environment and at home. In the second stage, another group of observers on Prolific completed the experiment using their own displays.

Observers. **I**) For the laboratory and at-home experiments, data were collected from 48 university observers (35 females), with three excluded from further analysis due to low accuracy in the catch trials. All observers passed the Ishihara color blindness test, and the order of the two experiments was counterbalanced. **II**) For the Prolific experiment, data were collected from 142 observers (54 females, 87 males, 1 unspecified; 37.9 ± 12.1 years), with three excluded due to low accuracy in the catch trials (< 75%). Observers confirmed they were not color-deficient. For both groups, informed consents were obtained beforehand, and all tasks complied with the Helsinki Declaration and were approved by the local ethics committee (Giessen LEK 2020-0015).

Setup. Stimuli were presented on the calibrated OLED display in the laboratory experiment, and on the observers' own displays for the at-home and Prolific experiments (no specific lighting limitations). The automatic brightness adjustment and night shift features were turned off. Additionally, the displays (N = 45) of observers from the home experiment were measured using the Konica Minolta Spectroradiometer CS-2000A (fig. S2B). The same stimulus presentation software (TypeScript/p5.js) was used as in the earlier experiments.

Visual stimuli. In this experiment, chroma levels were expanded from a single highest chroma to six. The 144 color patches remained unchanged, with 12 hues and 12 intensities (ranging from 20% to 100% in 12 steps). For each of these patches, six chroma levels were interpolated between the target patch and a gray patch (with no chroma) of the same intensity. This produced 732 color patches (fig. S4A, comprising 12 hues × 12 intensities × 5 chroma levels, plus 12 gray patches with no hue variation). Each trial presented 12 square patches on a black background arranged in a 4×3 array, covering all hues, intensities, and chroma levels, with each chroma level repeated twice. The experiment began with five practice trials, followed by 79 trials, including seven catch trials with monotonic intensity changes. Different observers were tested with different color patches; however, for the observers who completed both laboratory and at-home experiments, each one was tested with matched color patches across the two sessions, with locations and trial order randomized. All RGB values of those patches were gamma-corrected based on the OLED settings. All other settings and procedures were consistent with the previous experiment.

Data analysis. The model predictions were compared by correlating brightness rankings with model predictions (fig. S5). Most of the analysis remained unchanged. We fitted the wMaxRGB and weighted sumRGB models using both the full dataset across all six chroma levels (732 patch rankings) and subsets corresponding to each individual chroma level (144 patches per level, excluding the gray patches), to examine how chroma variations affect heterochromatic brightness (fig. S4D). **Experimental Consistency.** To assess consistency with the previous fixed-chroma experiment, we compared the rankings of the highest chroma patches from the current experiment with those from the earlier fixed-chroma experiment using Spearman correlation. **Chroma Effect.** Several ways were used to test the chroma effect on brightness: 1) Model performance with chroma change (fig. S4C). 2) The weight shifts by chroma (fig. S4B). **Ranking Comparison Across Experimental Groups.** Rankings of the 732 patches were correlated across the laboratory, home, and Prolific experiment groups (fig. S7).

Steady-State Visual Evoked Potentials (SSVEPs)

SSVEPs were recorded to examine neural responses to luminance and heterochromatic brightness using color discs flickering at 3 Hz and 15 Hz, with EEG signals collected from occipital electrodes.

Observers. Ten observers (seven females, mean age: 24 years old, range: 23-26 years old) with normal or corrected-to-normal visual acuity and normal color vision (tested by Ishihara plates) participated. This study was in accordance with the Declaration of Helsinki, and was approved by the local ethics committee (No. 102772019RT009). All observers gave informed consent before participation.

Setup. Stimuli were presented on a LCD display (1440 × 900, 35° × 22° at a viewing distance of 65 cm, 60 Hz), calibrated with a X-Rite i1Pro2 spectrophotometer (X-Rite, Michigan, USA). Its gamut (CIE 1931 xyY) was: R = [0.64, 0.34, 16.5], G = [0.30, 0.61, 59.0], B = [0.15, 0.073, 8.7]. Stimuli were gamma-corrected to a linear scale. Observers sat in a dimly lit room, ~65 cm from the display. The stimuli and task were programmed in MATLAB using the Psychophysics Toolbox (20). In the SSVEP experiment, EEGs were recorded from 3 scalp sites (O1, Oz, O2) according to the international 10-20 system, at 1000 Hz sampling rate through a BrainAmp amplifier (Brain Products, Munich, Germany). Only O1, Oz, and O2 electrodes were recorded because SSVEP responses were confined to these 3 occipital electrodes in all of our previous results with similar setups (21-23). The online reference was placed at FCz electrode, and the ground was placed at AFz. The electrode impedances were kept below 5 kΩ.

Visual stimuli. Seven primary and secondary colors were used in this study, with intensities capped at 50% to avoid saturation, all stimuli were displayed against a black background (0.2 cd/m^2). The seven colors were red (xyY: [0.64, 0.34, 8.3]), green ([0.30, 0.61, 29.6]), blue ([0.15, 0.073, 4.4]), yellow ([0.41, 0.52, 37.9]), magenta ([0.29, 0.15, 12.6]), cyan ([0.22, 0.31, 34.0]), and white ([0.29, 0.32, 42.3]).

I) In the SSVEP experiment, a flickering color disc (radius = 1°) was presented at the center of the monitor in each trial (fig. S8A). The disc had one of seven colors and flickered at either 3 Hz or 15 Hz. Its intensity was sinusoidally modulated between 0% and one of four levels (12.5%, 25%, 37.5%, 50% of the maximum, fig. S8D), each lasting 1 second, making each trial 4 seconds long. Observers self-paced the trials by pressing the spacebar to start the trial. They were required to maintain fixation on the disc without blinking while EEGs were monitored. Each color (seven colors) at each frequency (two frequencies) was presented 10 times, totaling 140 trials per observer. **II**) In the brightness ranking experiment, the same observers ranked seven color discs by perceived brightness. Each trial displayed the seven colors at randomly assigned positions on a circle (fig. S8E), with contrast levels drawn from [7%, 14%, 21%, 29%, 36%, 43%, 50%]. In each block of 7 trials, the disc with a specific color (e.g., red disc) went through all 7 levels of contrast. The discs were flickered at either 3 Hz or 15 Hz, across different trials. Observers sequentially clicked the brightest disc, which disappeared after clicking, until all were ranked. Each observer completed 140 trials.

Data Analysis. **SSVEP Analyses.** Functions from EEGlab toolbox (24) and customized scripts in MATLAB were used for analyzing EEG data. The 4-s signal from each trial was decomposed into 4 successive 1-s epochs. Each epoch corresponded to one intensity level of the target (fig. S8B). Epochs from the same condition at three electrodes (O1, Oz, O2) of each observer were averaged. The averaged epoch was converted to the frequency domain via fast-Fourier transformation (fft.m in MATLAB) with a resolution of 1 Hz (fig. S8C). At a certain frequency of interest (e.g., 15 Hz), we subtracted from the peak amplitude the average amplitude at two adjacent frequencies (e.g., 14 Hz and 16 Hz), in order to discount background noise from the computed SSVEP amplitude, e.g. as in (25). In order to calculate the overall SSVEP response to the 3Hz stimuli or the 15Hz stimuli, we took the average of all harmonics below 30Hz (26). That is, for 3Hz stimuli, the overall SSVEPs were calculated by averaging ten harmonics (i.e., 3, 6, 9, ... 30 Hz); whereas, for 15Hz stimuli, the overall SSVEPs were calculated by averaging two harmonics (15 and 30Hz). Data for the two flickering conditions were analyzed separately. SSVEPs to stimuli (intensity × color) were correlated with the stimuli's Luminance, MaxRGB.

<u>Real-world lighting experiments</u>

Comparing to the surface colors that were presented on the displays, we further explored how the brightness of real-world lighting was perceived. Tunable LED lamps were used to simulate real-world lighting.

Observers. Data were collected from 31 observers (19 females, 26.6 ± 4.4 years old), with one excluded from further analysis due to low accuracy in the catch trials. All observers passed the Ishihara color blindness test, informed consents were obtained before the experiment, and all tasks complied with the Helsinki Declaration and were approved by the local ethics committee (Giessen LEK 2020-0015).

Setup. An illuminated room (220 × 240 × 176.5 cm) was built for this experiment, painted with neutral RAL 7047 matt paint. Two LED-based, spectrally tunable illuminators (JUST Normlicht DLS moduLight 2-800, JUST Normlicht GmbH, Weilheim, Germany) were positioned at the back, 145 cm from the front wall. Observers sat in the room, 135 cm from the front wall (fig. S9B). Each illuminator (94 × 39 × 11.9 cm) had a 70 × 80 cm illumination area. Five of the seven LED channels within the visible spectrum were used (fig. S9A): four narrow-band channels (peak wavelengths/FWHM: 415 nm/16 nm, 463 nm/20 nm, 503 nm/43 nm, 631 nm/15 nm) and one broad-band channel (peaks at 449 nm/18 nm and 618 nm/188 nm). The illuminators produced light from the left and right sides of the room. Two JETI Spectraval1501 spectroradiometers (JETI Technische Instrumente GmbH, Jena, Germany), mounted on each illuminator, measured reflected light from 380 to 780 nm in 1 nm increments. Experiments were controlled and analyzed in MATLAB 2020b (The Mathworks, Natick, MA, USA).

Visual stimuli. Seven predictors were used in this experiment: Luminance, Radiance, Melanopsin Excitation (a photopigment expressed by ipRGCs)(27), SumRGB, MaxRGB, weighted sumRGB, and wMaxRGB with optimal weights fitted to the results. We adopted an RGB representation based on a tripartition of the visible spectrum, converting spectral data into blue (380–482.65 nm), green (482.65–565.43 nm), and red (565.43–780 nm) components, analogous to the canonical representation in colorimetry (28). Predictors were combined into 21 pairs, creating scenarios where one illuminant had a higher value for one predictor and the other had a higher value for a different predictor (e.g., higher Luminance but lower Radiance; for examples see Fig. 4, D and E; fig. S9, C to F). Observers were required to choose the brighter side. Additionally, catch trials were introduced. In catch trials, all predictors were either higher or lower for the left illuminant. Illuminant pairs in the experiment were identical for all observers, but presented in a random order. In each trial, observers first adapted to an approximated D65 illuminant (61.80 cd/m², x = 0.311, y = 0.34), followed by a 200 ms beep (0.6 kHz) indicating the switch to the test illuminants for 2s, after which the illuminants returned to D65, and observers made their choice. The experiment began with 30 practice trials with feedback to familiarize observers. The main session consisted of 630 trials without feedback. The total duration was 1-1.5 hours per observer.

Data analysis. Pairwise comparisons were made for the 21 conflicting scenarios. For each scenario, all conflicting illuminants that met the conflicting criterion were identified. The choosing percentage of each predictor was then calculated, providing a direct comparison between each pair of predictors.

Figures

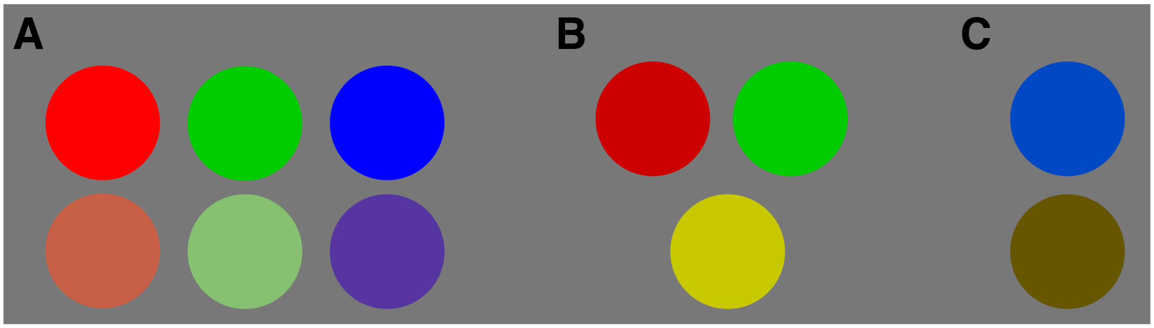

Fig. S1. Three known failures of luminance to account for heterochromatic brightness. (**A**) Helmholtz-Kohlrausch effect: lights of higher saturation (top row) are perceived brighter. (**B**) Sub-additivity: The additive mixture of red and green (yellow) appears less bright than the sum of the luminances would predict. (**C**) Underestimation of short wavelength light: The blue patch on top has approximately the same luminance as the brownish patch below, which appears drab and of much lower brightness.

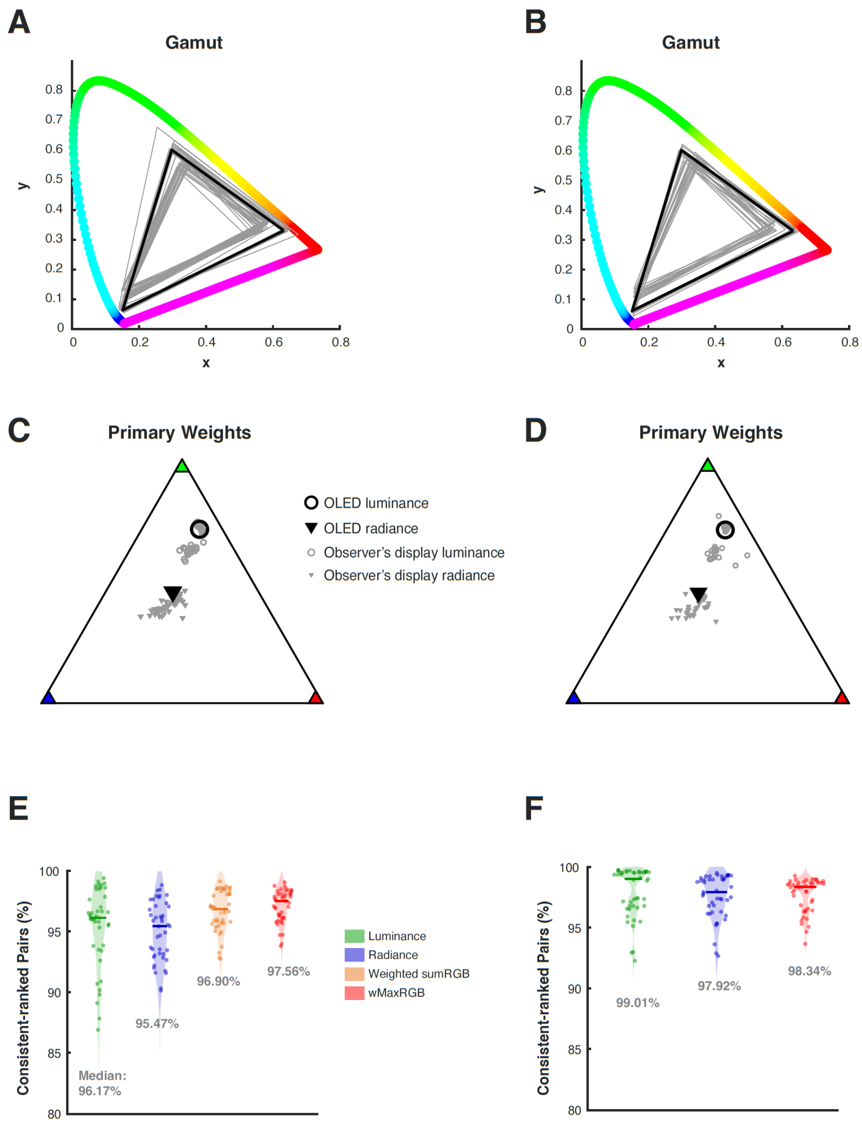

Fig. S2. Measured color appearance of OLED and observer displays. (**A** to **B**) Gamuts of the displays shown in CIE 1931 chromaticity coordinates. In total, 43 displays were measured for the fixed-chroma experiment (A**,** one observer's display was unavailable), and 45 for the chroma-variation experiment (B). (**C** to **D**) Red, green, and blue contributions to Luminance and Radiance across displays in the fixed-chroma experiment (C) and chroma-variation experiment (D). The distance from each circle/triangle to the side opposite the corresponding primary (e.g., red) indicates that primary's contribution to Luminance or Radiance. (**E** to **F**) Percentage of stimulus pairs that were consistently ranked (compared to OLED) by Luminance, Radiance, weighted sumRGB, and wMaxRGB models. For the fixed-chroma experiment (E), 144 color patches generated 10,296 pairwise comparisons. For each pair, the higher-Luminance patch on a given display was identified and compared to the OLED reference to compute consistency; each dot represents one display. The same was done for Radiance, weighted sumRGB, and wMaxRGB models. In the chroma-variation experiment (F), 732 patches generated 267,546 pairs, with consistency computed similarly.

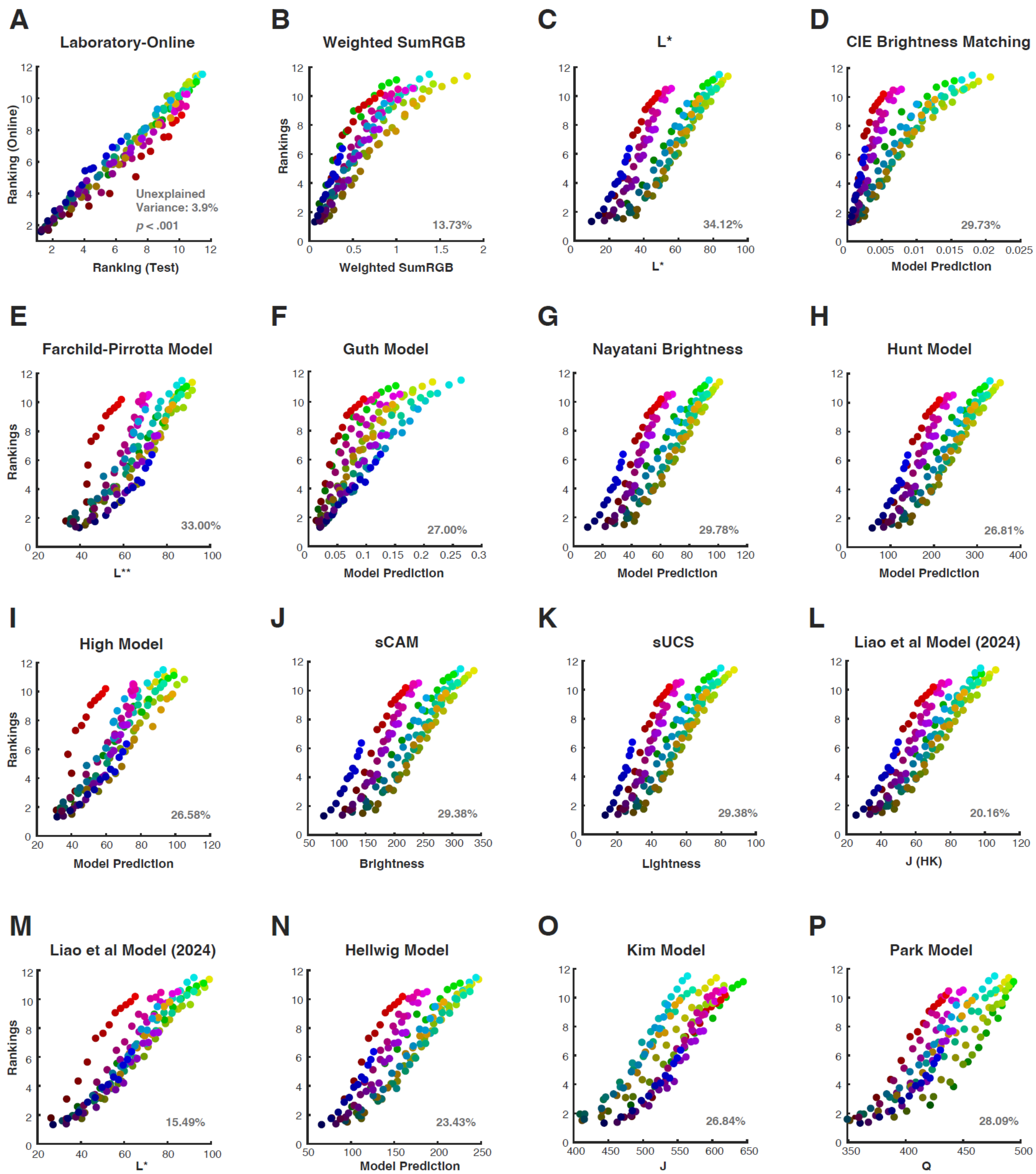

Fig. S3. Model predictions of brightness ranking in the laboratory experiment (test session). (**A**) Correlation of mean brightness rankings for 144 color patches between laboratory and online testing. (**B** to **P**) Correlations between model-predicted values from color appearance models and mean brightness rankings for 144 color patches. Model details are described in the Materials and Methods. Unexplained variance is shown for each plot.

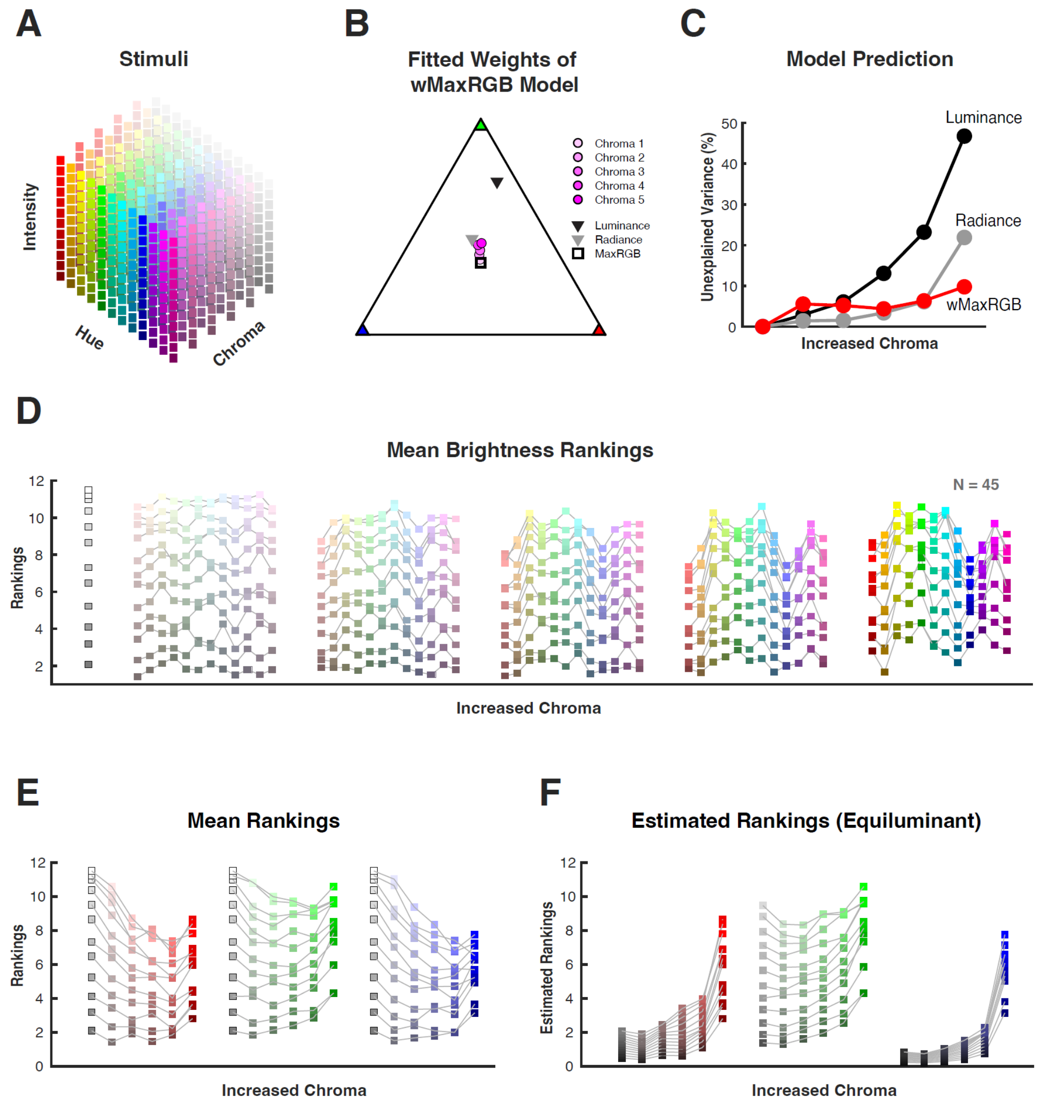

Fig. S4. Changes in brightness rankings and model predictions with chroma variation. (**A**) Stimuli used in the chroma-variation experiment. The original 144 color patches (12 hues × 12 intensity levels, from 20% to 100% in 12 steps) were retained. For each patch, six chroma levels were interpolated between the target patch and a gray patch (with no chroma) of the same intensity, resulting in a total of 732 patches. (**B**) Optimal weights of the wMaxRGB model fitted to the mean rankings at five chroma levels. Rankings for gray patches (no chroma) were excluded. The distance from each circle to the side opposite the corresponding primary (e.g., red) indicates the weight assigned to that primary. (**C**) Model comparison with chroma variation. For each chroma level, the wMaxRGB model was independently fitted to the corresponding mean rankings. (**D**) Brightness rankings at different chroma levels were averaged across observers and repetitions. (**E**) Mean brightness rankings vary with increasing chroma for three primary hues across different chroma levels. (**F**) Helmholtz–Kohlrausch effect based on ranking data. For each line in the plot, patches at lower chroma levels were adjusted to be equiluminant with the patch at the maximum chroma level on the same line. First, the luminance intensities of the lower-chroma patches were calculated to ensure equiluminance with the maximum-chroma patch. Then, their estimated brightness rankings were obtained using interpolation.

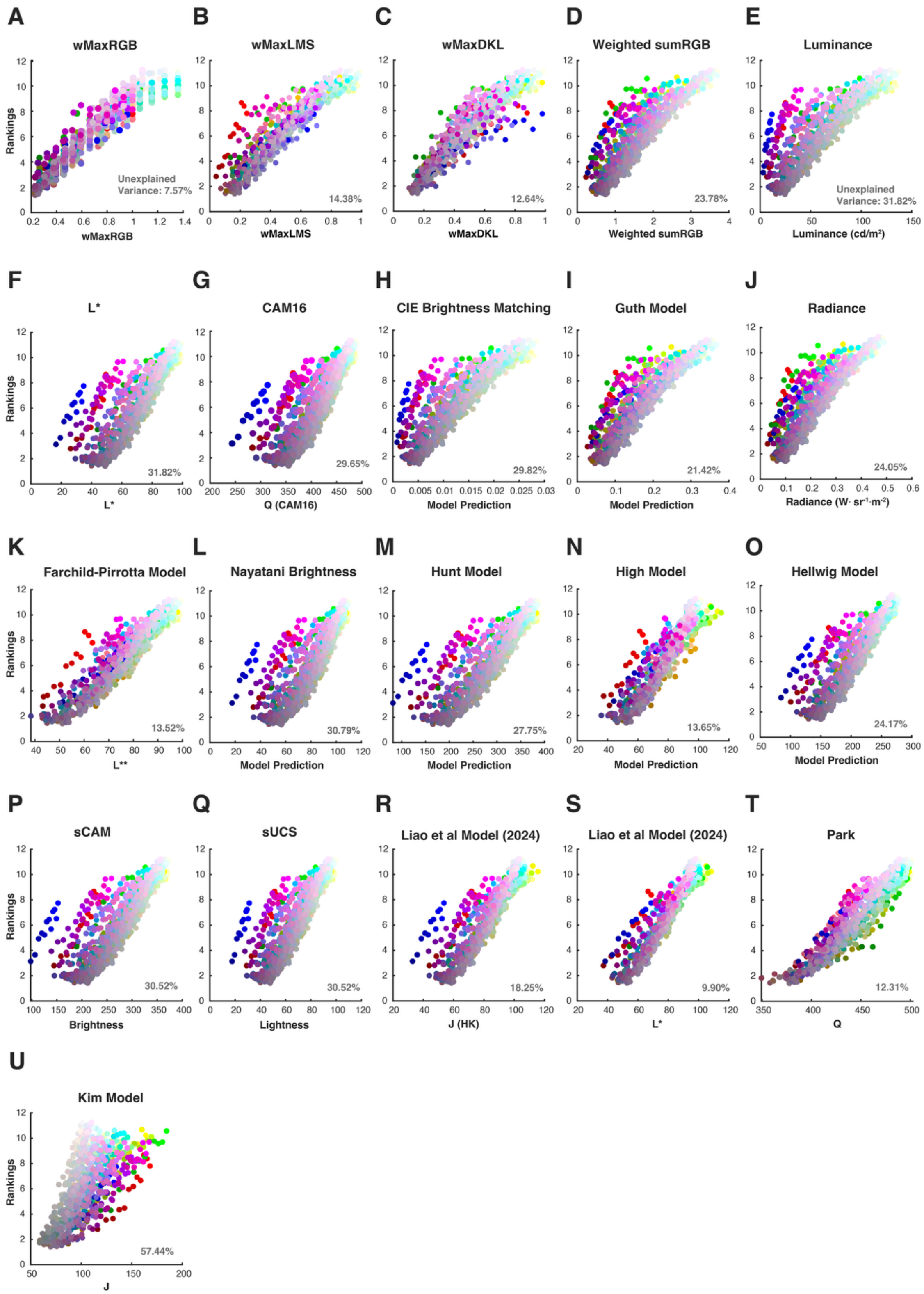

Fig. S5. Model predictions of brightness rankings in the chroma-variation experiment (test session). Correlations between model-predicted values from various color appearance models and the mean brightness rankings for 732 color patches. Model computations are the same as in fig. S3. Unexplained variance is indicated for each plot.

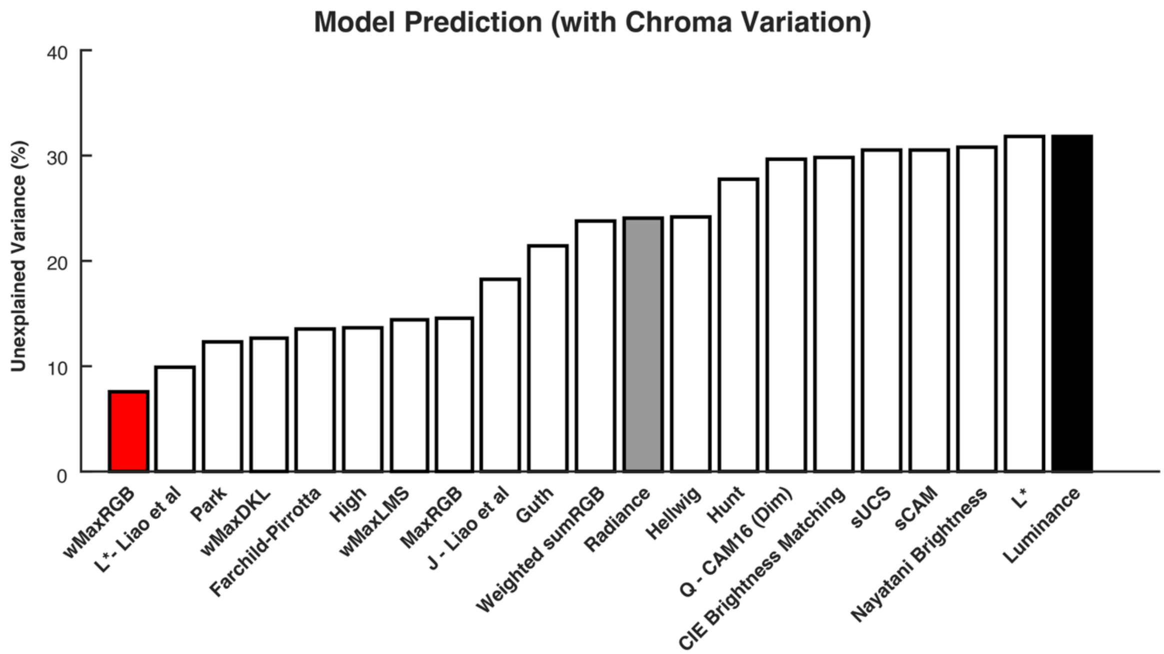

Fig. S6. Prediction of representative model classes for the full chroma-ranking dataset. Bars show unexplained variance for the nonlinear pooling model fit to the expanded chroma dataset comprising 732 stimuli. Although optimal weights differ slightly across chroma levels, nonlinear pooling consistently outperforms additive photometric models. Unexplained variance is defined as one minus the squared Spearman correlation between predicted and observed rankings. For models with free weights, parameters were estimated independently within each dataset by minimizing unexplained variance.

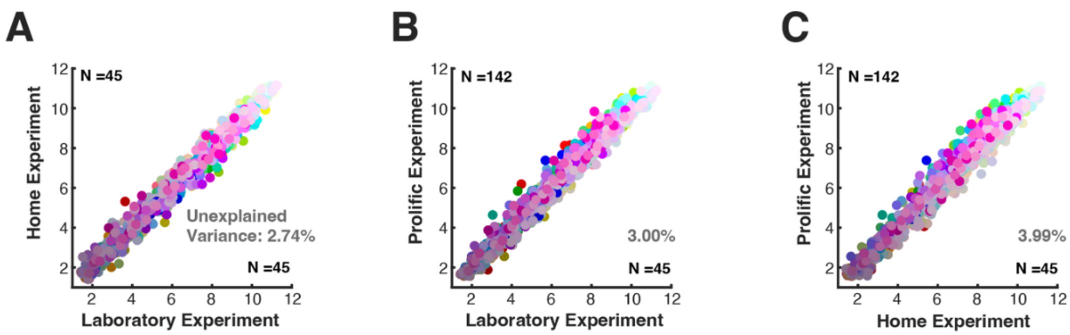

Fig. S7. Mean rankings from three experimental groups and their correlations in the chroma-variation experiment. These groups are: a laboratory test session, a home session (same group as the laboratory test session), and an online session.

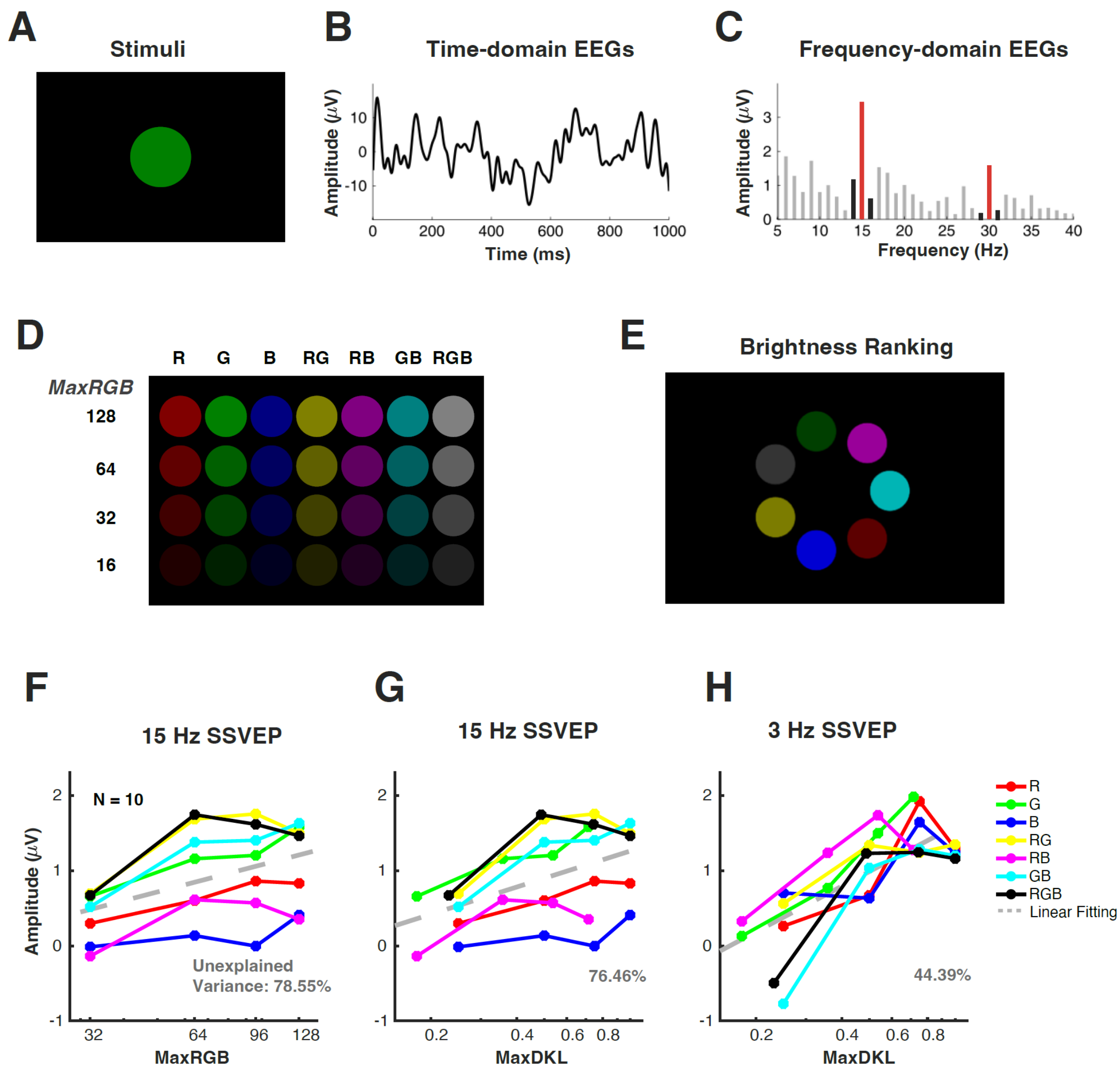

Fig. S8. SSVEP experiment setup. (**A**) A color disc with a radius of 1° was presented to the fovea of human observers. The color disc was modulated sinusoidally between 0% and a certain level of contrast, at a frequency of 3 Hz or 15 Hz. (**B**) The example shows a time-domain EEG trace in response to a 15Hz flicker, and its amplitude spectrum in the frequency domain. (**C**) When calculating the SSVEP amplitude of the 15Hz flicker, we took the average of peak responses at 15Hz and 30Hz (red bars), after discounting the average baseline amplitude of adjacent frequency bins (black bars). (**D**) The stimulus sets of the SSVEP experiment. The disc had one of the primary or secondary colors on the monitor (i.e., R, G, B, RG, RB, GB, RGB), and its intensity was swept from 12.5%, 25%, 37.5% to 50%. (**E**) A shortened version of the brightness ranking was tested on the same group of observers who participated in the SSVEP experiment. The stimuli included the same 7 colors as in the SSVEP task, each presented at 7 linearly spaced intensities ranging from 7% to 50%. In each trial, all colors and intensities were presented. (**F**) SSVEP amplitude at 15 Hz plotted as a function of MaxRGB and (**G**) MaxDKL. (**H**) SSVEP amplitude at 3 Hz plotted as a function of MaxDKL.

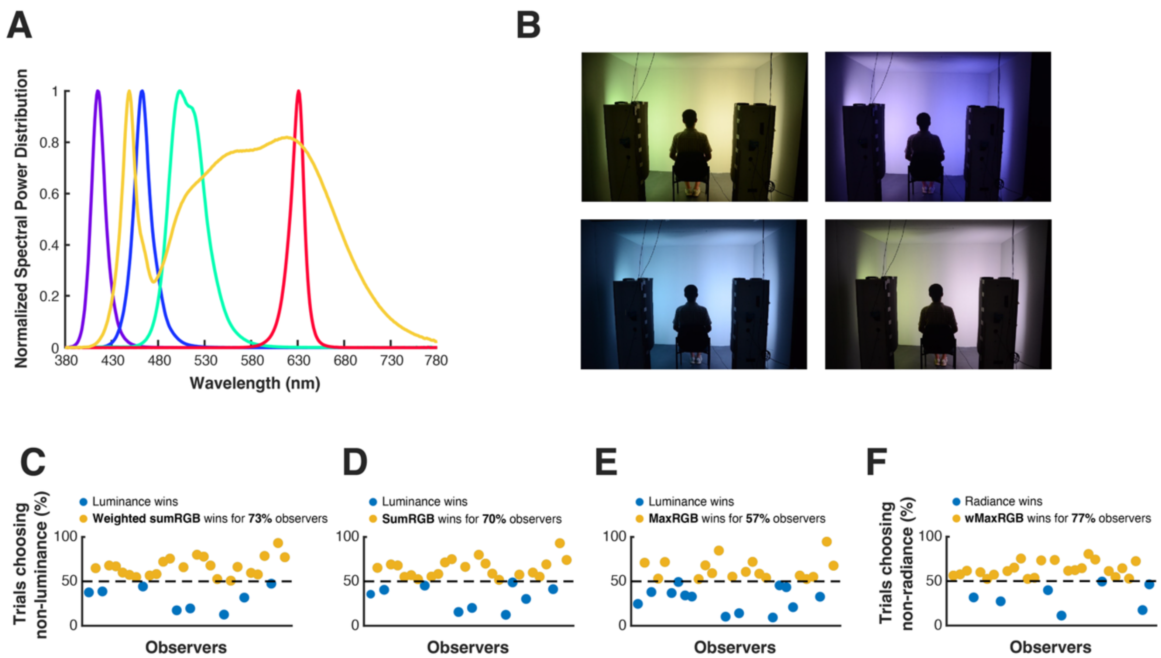

Fig. S9. Real-world lighting experiment and result. (**A**) Normalized spectral power distribution of the five primaries on the LEDs. These channels include four narrow-band channels with peak wavelengths and FWHM values of 415 (16 nm), 463 (20 nm), 503 (43 nm), 631 (15 nm), as well as one broad-band channel featuring two peaks at 449 (18 nm), 618 (188 nm). (**B**) Room and illumination setup: Two LED lamps, each equipped with five controllable primaries, were positioned on either side of the illuminated room. Observers were seated at the center of the room. The four photos show examples of the conflicting lighting scenarios. On each trial, the observer had to choose the brighter side. (**C** to **E**) The percentage of trials where observers chose non-Luminance predictors over Luminance in conflicting scenarios. Each circle represents an observer (blue: Luminance >50%, yellow: other predictors >50%). (**F**) The percentage of trials where observers chose radiance over wMaxRGB in conflicting scenarios. Each circle represents an observer (blue: radiance >50%, yellow: wMaxRGB >50%).

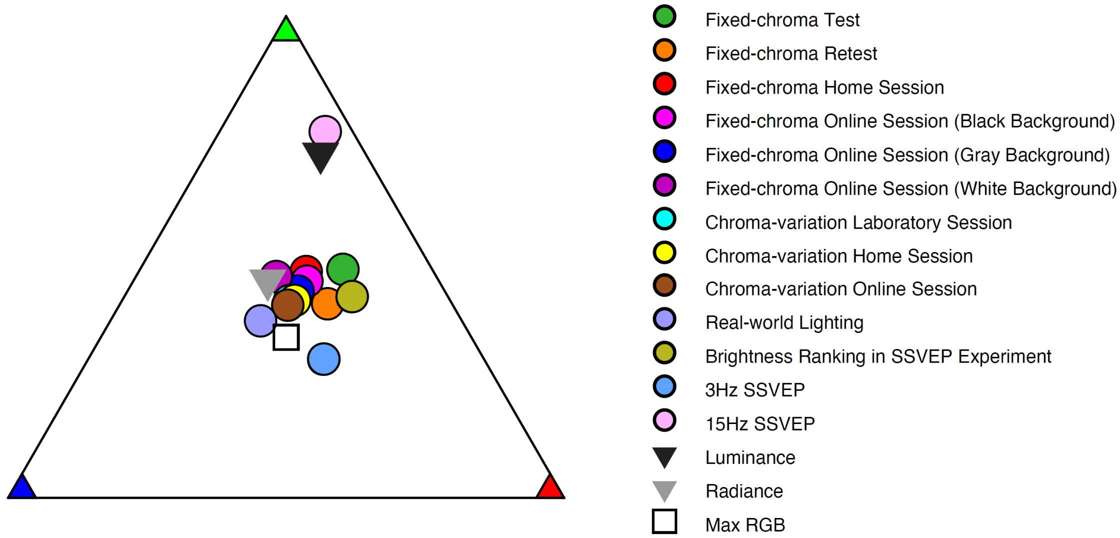

Fig. S10. Optimal weights of the wMaxRGB model fitted across different experiments. The distance from each circle to the side opposite a given primary (e.g., red) represents the weight of that primary. The blue and green triangles represent the relative contributions of the red, green, and blue to Radiance and Luminance on the OLED display, respectively. Weights cluster around MaxRGB, with luminance and 15 Hz SSVEP being the only exceptions.

SI References

1. S. Ishihara, *Ishihara's tests for colour deficiency, 24 plates ed.* (Kanehara & Company, 1997).
2. CIE, *Commission internationale de l'eclairage proceedings, 1931* (Cambridge University Press, Cambridge, 1932).
3. CIE, *Commission internationale de l'eclairage proceedings, 1924* (Cambridge University Press, Cambridge, 1926).
4. CIE, Recommendations on uniform color spaces, color difference equations, psychometric color terms. 1-21 (1978).
5. CIE, Spectral Luminous Efficiency Functions Based upon Heterochromatic Brightness Matching for Monochromatic Point Sources, 2° and 10° Fields. *CIE Pub.* **75** (1988).
6. S. Guth, A new color model. *Color Metrics* (1972).
7. Y. Nayatani, K. Takahama, H. Sobagaki, K. Hashimoto, Color-appearance model and chromatic-adaptation transform. *Color Research & Application* **15**, 210-221 (1990).
8. R. Hunt, Revised colour-appearance model for related and unrelated colours. *Color Research & Application* **16**, 146-165 (1991).
9. C. Li *et al.*, Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS. *Color Research & Application* **42**, 703-718 (2017).
10. N. Moroney, M. Fairchild, R. Hunt, C. Li, The CIECAM02 color appearance model. (2002).
11. M. Li, M. Luo, Simple color appearance model (sCAM) based on simple uniform color space (sUCS). *Optics Express* **32**, 3100-3122 (2024).
12. G. High, P. Green, P. Nussbaum, The Helmholtz-Kohlrausch effect on display-based light colors and simulated substrate colors. *Color Research & Application* **48**, 167-177 (2023).
13. L. Hellwig, D. Stolitzka, M. D. Fairchild, Extending CIECAM02 and CAM16 for the Helmholtz–Kohlrausch effect. *Color Research & Application* **47**, 1096-1104 (2022).
14. G. Seong, Y. Kwak, H. Kim, The CIECAM16-Based Lightness Model Incorporating the Helmholtz–Kohlrausch (H–K) Effect Using Heterochromatic Brightness Matching Data. *Color Research & Application* (2025).
15. Y. Park, Determining the color appearance of Helmholtz-Kohlrausch effect for self-emissive displays. *Journal of Information Display* **23**, 235-242 (2022).
16. Z. Liao, L. Xu, Lightness modifications of the CIECAM16 and CIELAB based on the Helmholtz-Kohlrausch effect. *Optics Express* **32**, 44918-44932 (2024).
17. M. D. Fairchild, E. Pirrotta, Predicting the lightness of chromatic object colors using CIELAB. *Color Research & Application* **16**, 385-393 (1991).
18. A. M. Derrington, J. Krauskopf, P. Lennie, Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of physiology* **357**, 241-265 (1984).
19. J. Krauskopf, D. R. Williams, D. W. Heeley, Cardinal directions of color space. *Vision research* **22**, 1123-1131 (1982).
20. M. Kleiner, D. Brainard, D. Pelli, What's new in Psychtoolbox-3? (2007).
21. J. Chen, M. Valsecchi, K. R. Gegenfurtner, Attention is allocated closely ahead of the target during smooth pursuit eye movements: Evidence from EEG frequency tagging. *Neuropsychologia* **102**, 206-216 (2017).
22. J. Chen, M. Valsecchi, K. R. Gegenfurtner, Enhanced brain responses to color during smooth-pursuit eye movements. *J Neurophysiol* **118**, 749-754 (2017).
23. J. Chen, M. Valsecchi, K. R. Gegenfurtner, Saccadic suppression measured by steady-state visual evoked potentials. *Journal of Neurophysiology* **122**, 251-258 (2019).
24. A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* **134**, 9-21 (2004).

25. J. Liu-Shuang, K. Torfs, B. Rossion, An objective electrophysiological marker of face individualisation impairment in acquired prosopagnosia with fast periodic visual stimulation. *Neuropsychologia* **83**, 100-113 (2016).
26. T. L. Retter, B. Rossion, C. Schiltz, Harmonic amplitude summation for frequency-tagging analysis. *Journal of Cognitive Neuroscience* **33**, 2372-2393 (2021).
27. M. Spitschan, Melanopsin contributions to non-visual and visual function. *Current opinion in behavioral sciences* **30**, 67-72 (2019).
28. J. Koenderink, A. v. Doorn, K. R. Gegenfurtner, Colors and things. *i-Perception* **11**, 2041669520958431 (2020).