Preregistration document for the color-discrimination project with different adapting points

Experimenters: Fangfang Hong, Cameron Zhang, Kara Wang, David Brainard
Affiliation: University of Pennsylvania, Department of Psychology

Purpose

This study examines how the adapting point (e.g. as set by the background) affects chromatic discrimination in the isoluminant plane. The methods are similar to those used in our measurements of the chromatic discrimination field in the isoluminant plane (Hong et. al, 2025). Here we will measure the field for two subjects and use two different backgrounds. We view this as an initial study, and after completion may choose to extend it with adjustments to the methods, additional adapting conditions, or more subjects. At the same time, unless we encounter unexpected difficulties with the measurements, our intent is that these data will be reported in an eventual paper on how adaptation affects color discrimination. This document was finalized before data collection commenced.

Methods

Apparatus

The experiment will be conducted using an Alienware computer (Aurora 11) running Windows 11, equipped with Intel® Core™ i7-10700K processor and NVIDIA GeForce RTX 3080 GPU. The display is a DELL U2723QE monitor (59.8 cm width, 33.6 cm height, 3480 x 2160 resolution, 60 Hz refresh rate, achieving 12-bit effective color depth. This bit depth is achieved via the spatial dithering implicit in the rendering of the stimuli, rather than because of the intrinsic bit depth of the video pipeline at individual pixels. The monitor will be positioned 130 centimeters from the chinrest, subtending a visual angle of 25.9 x 14.7 degrees of visual angle. Monitor color and luminance measurements were obtained using a Klein K-10A colorimeter and a CR-250 spectroradiometer. The pixel resolution of the display is approximately 140 pixels/deg, above the typical human foveal resolution limit.

The Alienware computer will be used solely for stimulus presentation, whereas adaptive sampling of the stimuli is performed on a separate custom-built PC with a high-performance Gigabyte motherboard (X299X aorus master), an NVIDIA GeForce RTX 5070 GPU and a 12-core Intel i9-10920X processor. This computer also runs Windows 11. The two computers communicate via a shared network disk, using a custom protocol based on text files that both computers could read and write. It is possible that we will change the system that performs adaptive sampling, with the goal of reducing the compute time required for adaptive trial selection. We will post an addendum if this is done.

A USB speaker (3 Watts output power, 20k Hz frequency response) is used for playing auditory feedback, and a gamepad controller (Logitech Gamepad F310) is used for registering trial-by-trial response.

Stimulus

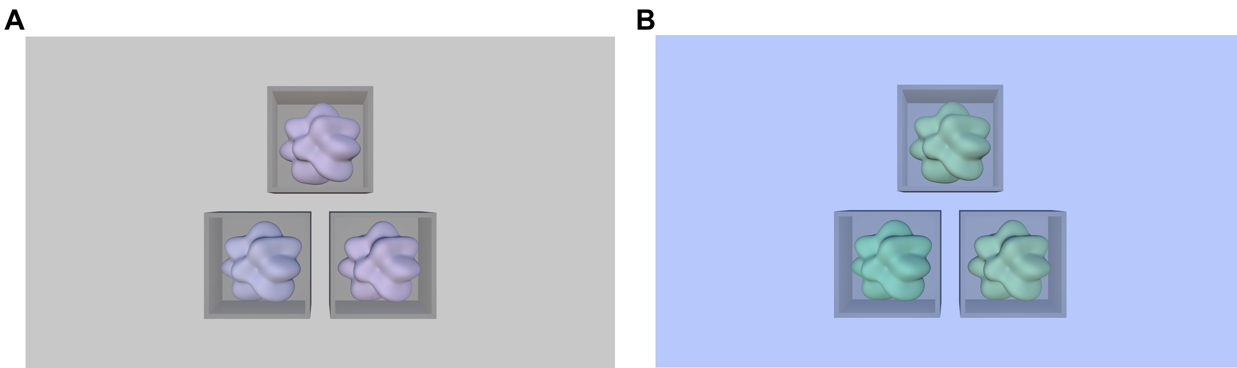

Figure 1. Example scenes for the two adaptation conditions: neutral gray and high-CCT blue (~12,000 K along the Planckian locus).

(x, y, Y)	Gray adaptation	Blue adaptation
Cubic room	0.3125, 0.3294, 84.3396	0.2649, 0.2811, 85.6234
Background	0.3118, 0.3313, 149.2435	0.2684, 0.2829, 147.9283

Table 1. Chromaticity and luminance of the cubic room and the uniform background across conditions.

The visual scene (**Fig. 1**) was constructed in Unity (v2022.3.24f1) and rendered using its standard shader. The scene consisted of three identical blobby 3D objects, each created in Blender (v4.0) with a matte, non-reflective surface. On each trial, the surface color of the blobby objects is varied by adjusting their RGB values in Unity. The three blobby objects (2.5 x 2.5 dva; 203,900 pixels each) are arranged in a triangular configuration. Each blobby object is centered and floating inside its own cubic room (3.3 x 3.3 dva). The centers of the blobby objects are 3.7 dva apart. Each room, along with the blobby stimulus inside it, is illuminated exclusively by an achromatic spotlight positioned in front of the object and set to maximum intensity (R = G = B = 1). The three rooms were presented against a spatially uniform background.

To examine how adaptation affects color discrimination, we will vary the surface color of both the cubic room and the properties of the uniform background. We will use two adapting conditions: (1) a neutral gray (as in Hong et al., 2025) and (2) a blue selected to lie approximately on the daylight locus at a correlated color temperature of 12,000 K. For each condition, we measured the cubic room's and the background's chromaticity and luminance with the CR-250. We selected separate RGB settings for the room and the background that have closely matched chromaticities within each condition and produced similar luminance levels across conditions. Summary measurements are provided in **Table 1**. Note that there is not theoretical reason why it is critical to have a particular chromaticity or luminance for either the cubic rooms, nor that they be matched to each other within condition. Rather, we aim to set these so as to induce different states of adaptation in the visual system across the gray and blue adapting conditions, with the target chromaticities in the two chosen to resemble but not exactly match variation in daylight and the luminances chosen somewhat arbitrarily but with similar values across conditions.

Calibration and color depth

Before running the pilot experiment, we conducted six calibration tasks using the same stimuli as in the study: (1) used the Klein K-10A to measure the precision (quantization) of our display pipeline; (2) the CR-250 to calibrate the blobby located at bottom right to measure the relationship between input and the actually displayed color; (3) used the CR-250 to verify the agreement between the input and displayed color based on the lookup table constructed from (2); (4–5) repeated tasks (2–3) to calibrate the uniform background separately; and (6) repeated task (3) with the blue cubic room and background to confirm that the same channel properties (channel spectra and gamma functions) for the objects remain valid across adaptation conditions.

In the first calibration task, we used the Klein K-10A to confirm that the output color depth achieved smooth increments via Unity's standard shader with its inherent and implicit spatial dithering. We tested RGB values within the range of 1928 / 4095 to 2128 / 4095, with an increment of 1/4095. Each stimulus was displayed for 5 seconds, and the RGB settings from the first frame of the frame buffer were saved in EXR format. We then compared the average RGB values from the EXR files across the blobby object to the luminance (cd/m²) measured over the entire stimulus presentation. Results confirmed that Unity achieved at least 12-bit depth (**Fig. 2**).

In the second calibration task, we verified several key aspects, including the monitor's gamma function as driven through Unity (sampled in 61 evenly spaced steps from 0 to 1), primary spectral power distributions and their stability, primary chromaticities and their stability, linearity, additivity, as well as the effect of background on the spectral power distribution of the target stimulus (**Fig. 3**). In previous experiment (Hong et al., 2025), we also made the same measurements for the other two blobby stimuli, and verified the consistency across screen locations.

The same gamma calibration (primary spectra, gamma correction) will be applied to all three objects during the main experiment, using the calibration data from the bottom right blobby stimulus. Specifically, we interpolated gamma table for 4,096 RGB values using a combination of linear and polynomial fits, which were then used to derive the inverse gamma function for gamma correction in Unity (**Fig. 4A**). In the third calibration task, we validated the gamma correction by repeating color calibration with the correction applied in Unity; results showed excellent alignment with the identity line (**Fig. 4B**).

In the fourth and fifth tasks, we aimed the CR-250 at the uniform background to measure the input–output relationship and regenerate a background-specific gamma correction (**Fig. 5A**). This separate correction was necessary because the background is treated as a special case by Unity and through a different pipeline than the cubic rooms and blobby objects (**Fig. 5B**). Based on the background measurements, we computed a separate lookup table for the background (**Fig. 5C**). Notably, after applying the gamma correction, the input-output relationship is not perfect across primaries. It is none-the-less acceptable for our purposes, as precise control of the blobby-stimulus display is the most important requirement and we only require approximate control of the background (see note on this above).

The sixth task replicated task 3, except the cubic room and background were blue. We compared the input–output functions, chromaticities, and primaries with those from task 3 and found near-perfect agreement (**Fig. 6**), confirming that the same channel properties and gamma correction may be used for the two adaptation conditions.

Design

The main difference between the current design and Hong et al. (2025) is that we will not longer interleave validation trials, as our earlier work validates the use of the Wishart Process Psychophysical Model (WPPM) to characterize discrimination thresholds. Other than this, the remaining design will be very similar to the design described by Hong et al.

We will again restrict our stimuli (both reference and comparison) to be on the isoluminant plane in the DKL space. Stimuli are presented in RGB space, while trial placement and model fitting are conducted in the model space for mathematical convenience, as it aligns with the Chebyshev basis functions used in the model (see Hong et al., 2025). Consequently, transformation matrices are required to convert between these spaces. The transformation scheme across spaces remains the same as in the previous experiment reported in Hong et al. (2025). However, the gamut of the isoluminant plane is dependent on the monitor's spectral properties, which we remeasured for this preregistered study. Based on these updated measurements, we recomputed the transformation matrices to ensure accurate color space conversions (**Tables 2-3**). Any comparison of data across this study and Hong et al. (2025) will take this difference in the physical instantiation of the model space into account by transforming into a device-independent colorimetric representation (e.g., cone contrast space, DKL space, etc.).

For each adaptation condition, we will use AEPsych to determine both the reference and comparison stimuli adaptively. The reference stimuli are bounded between -0.75 and 0.75 in the model space, which itself is bounded between -1 and 1. Comparison stimuli are generated by adding a delta value along each of the two dimensions, with delta values constrained between -0.25 and 0.25 to ensure that stimuli remain within the [-1, 1] boundary. By varying both the reference stimuli (2D) and the delta values (2D), the task is now effectively a 4D psychometric field characterization problem. For each adaptation condition, subjects will complete 5,000 AEPsych-based trials, consisting of 900 Sobol-sampled trials and 4,100 trials using expected absolute volume change (EAVC) sampling. The Sobol-sampled trials will be scaled by one of three factors (1/4, 2/4, 3/4), with an equal distribution across these scalers. Smaller scaling factors increase task difficulty, and to counterbalance this effect, the scalers will be pseudo-randomized in the first session.

Compared with Hong et al. (2025), we reduced the number of AEPsych trials from 6,000 to 5,000 because simulations indicate this is sufficient to recover the ground truth. Specifically, we defined ground-truth thresholds using CIELAB $\Delta E94$, simulated 6,000 AEPsych trials, and then fit the WPPM to subsets of those data ranging from 1,000 to 6,000 trials. For each fit, we computed the Bures–Wasserstein distance (BWD) between the ground-truth threshold ellipses and the WPPM-predicted ellipses, summed over a 9×9 grid of reference stimuli. The BWD as a function of trial number plateaus at approximately 5,000 trials, indicating little additional benefit from collecting more than 5,000 (**Fig. 7**).

To avoid keeping subjects waiting while AEPsych determines for the next trial, we will implement the same fallback trial strategy: if AEPsych exceeds the maximum wait time (2.9 seconds; explained in Hong et al. (2015)), we will slot in Sobol trials, pre-generated with subject- and session-specific seeds (**Table 2**) to keep the experiment moving. The pre-generated Sobol trials will be scaled by one of three factors (2/8, 3/8, 4/8), with an equal distribution across these scalers. Then, 5% of these trials will be replaced by easy catch trials with the largest absolute delta values, i.e., [-0.25, -0.25], [-0.25, 0.25]. [0.25, -0.25] or [0.25, 0.25]. The catch trials allow us to detect observer lapses in attetion. Because we are not interleaving validation trials in this experiment, we expect the number of these fallback trials to be fairly large. If it is possible to speed up AEPsych trial selection with improved hardware, we may introduce new hardware during the course of the experiment

Each session will be split into two halves, one for each adaptation condition. The order in which each condition is presented will be determined randomly by subject- and session-specific random seeds (**Table 4**). In total, subjects will complete 20 sessions, with 500 AEPsych-based trials per session (250 trials per adaptation condition). Note that the actual number of trials per session will exceed 500 because pre-generated Sobol trials will also be slotted in. Subjects will take a break after every 125 AEPsych trials, resulting in 3 breaks per session.

Procedure

Subjects will perform a three-alternative forced-choice (3AFC) oddity task. Each trial begins with a fixation cross displayed in the center of the three cubic rooms for 0.5 seconds, followed by a blank screen for 0.2 seconds. Then, three blobby stimuli will appear in the middle of the cubic rooms for 1 second. After the stimulus, a response probe ("< ^ >" indicating the three possible responses) will appear, prompting subjects to determine which one is the odd stimulus, with no time constraint. Once subjects make a response using a gamepad controller, a blank screen will appear for 0.2 seconds, followed by visual and auditory feedback on accuracy—"correct" or "incorrect," accompanied by a beep or buzz tone. The inter-trial interval will be 1.5 seconds. Subjects will be instructed to move their eyes during stimulus presentation and try to fixate on each object.

We will run a 40-trial practice block at the start of each session. The practice is split into two halves—one for each adaptation condition—to familiarize subjects with the mid-block background change in the main experiment. To ensure adequate adaptation, prior to the first trial of each condition within a session, subjects will fixate on the screen with the adaptation background and cubic room displayed during a countdown (120 seconds for the main experiment; 10 seconds prior to the practice trials).

Data analysis

All analyses of color calibration data were conducted in MATLAB 2023b. The calibration scenes and the experiment will be created and run in Unity 2022.3.24f1, programmed in C#, and behavioral data will be analyzed in Python 3.11.

During color calibration, gamma functions were measured to compute the inverse gamma lookup table and the transformation matrix needed for converting Wishart space values into RGB values. These tables are then exported in CSV format to Unity, enabling gamma correction and the conversion of AEPsych-determined value pairs into RGB values.

For initial analysis, we will partition the data by adaptation condition. Within each condition, we will fit the WPPM jointly to all trials (AEPsych plus slotted in Sobol). Fitting of this model is described in Hong et al. (2025) and we will use the same WPPM parameters as described in that paper. We will then bootstrap to obtain confidence intervals for the model predictions. Finally, we will compare predictions across conditions to quantify how the most sensitive color shifts with adaptation. We will be interested in qualitative descriptions of how adaptation affects the discrimination field, and also expect that the data will constrain future models of color discrimination.

Order of data collection

We will begin by collecting a full dataset for two subjects, and analyze this data before enrolling additional subjects. Our hope is that data from these two subjects will not reveal the need for any major modifications to the experimental protocol, in which case we will include the data in reports of this experiment. We will post a follow-up preregistration document listing any changes made before running additional subjects, indicating whether these two subjects' data will be retained, and specifying the number of subjects to be run in the remainder of the experiment.

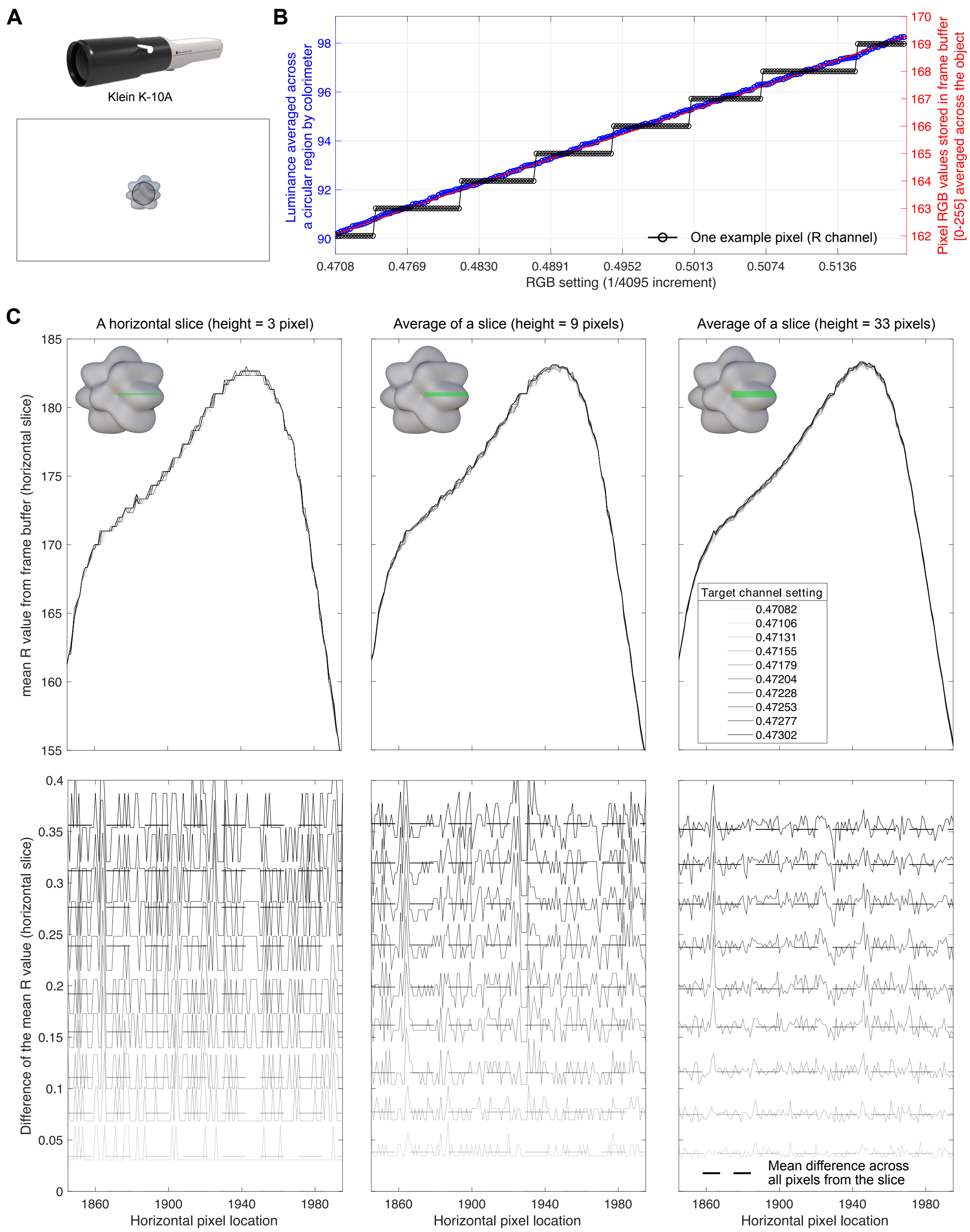

Figure 2. Evidence of spatial dithering by Unity's standard shader when the stimulus surface texture is modified. (A) Calibration task 1. The Klein K-10A colorimeter was positioned near the stimulus to measure luminance increments. The stimulus was centered on the screen, the cubic room removed, and the background set to white. This configuration facilitated extracting stimulus pixels from saved images for analysis. (B) Spatial dithering is suggested by comparing Klein K-10A luminance (averaged over a circular region on the blobby object) with RGB values in the frame buffer. Measured luminance shows small incremental changes as RGB increases in 1/4095 steps. These measurements match averages over pixels in a saved .exr frame-buffer image. The averaged pixel values exhibit 12-bit quantization even though individual pixels exhibit 8-bit quantization. (C) Top: mean R-channel values averaged vertically within a horizontal slice of the blobby object. Bottom: differences in R-channel values between the minimum target R setting and each of the others. Different grays denote different target R settings. For illustration, only part of the slice is shown; solid lines in the bottom row are scaled by 0.1. Dashed lines: mean difference averaged across all pixels in each slice.

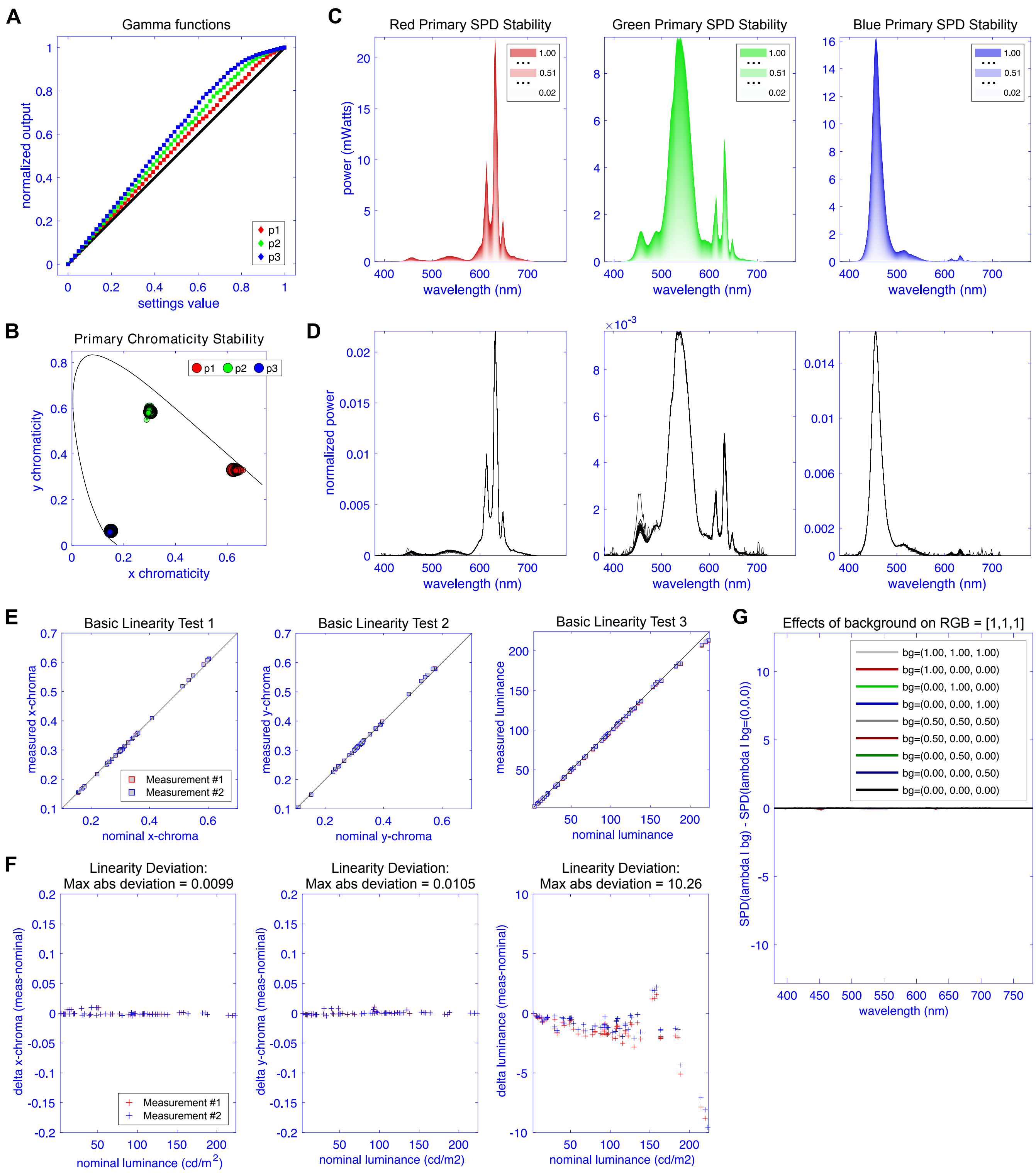

Figure 3. Color calibration results for the blobby stimulus in the top position of the triangular arrangement. (A) Gamma functions for the three primary colors (red, green, blue). (B) Chromaticity coordinates of each primary color in CIE color space at different intensity levels. (C) Spectral power distributions (SPDs) of the three primary colors at different intensity levels. (D) Normalized SPDs for each primary. (E) Linearity tests for chromaticity and luminance: comparison of nominal (predicted) and measured chromaticity (x and y) and luminance values across two separate measurements (F) Linearity deviation. (G) The effects of background (the surface color of the cubic room) on the SPD of the blobby stimulus.

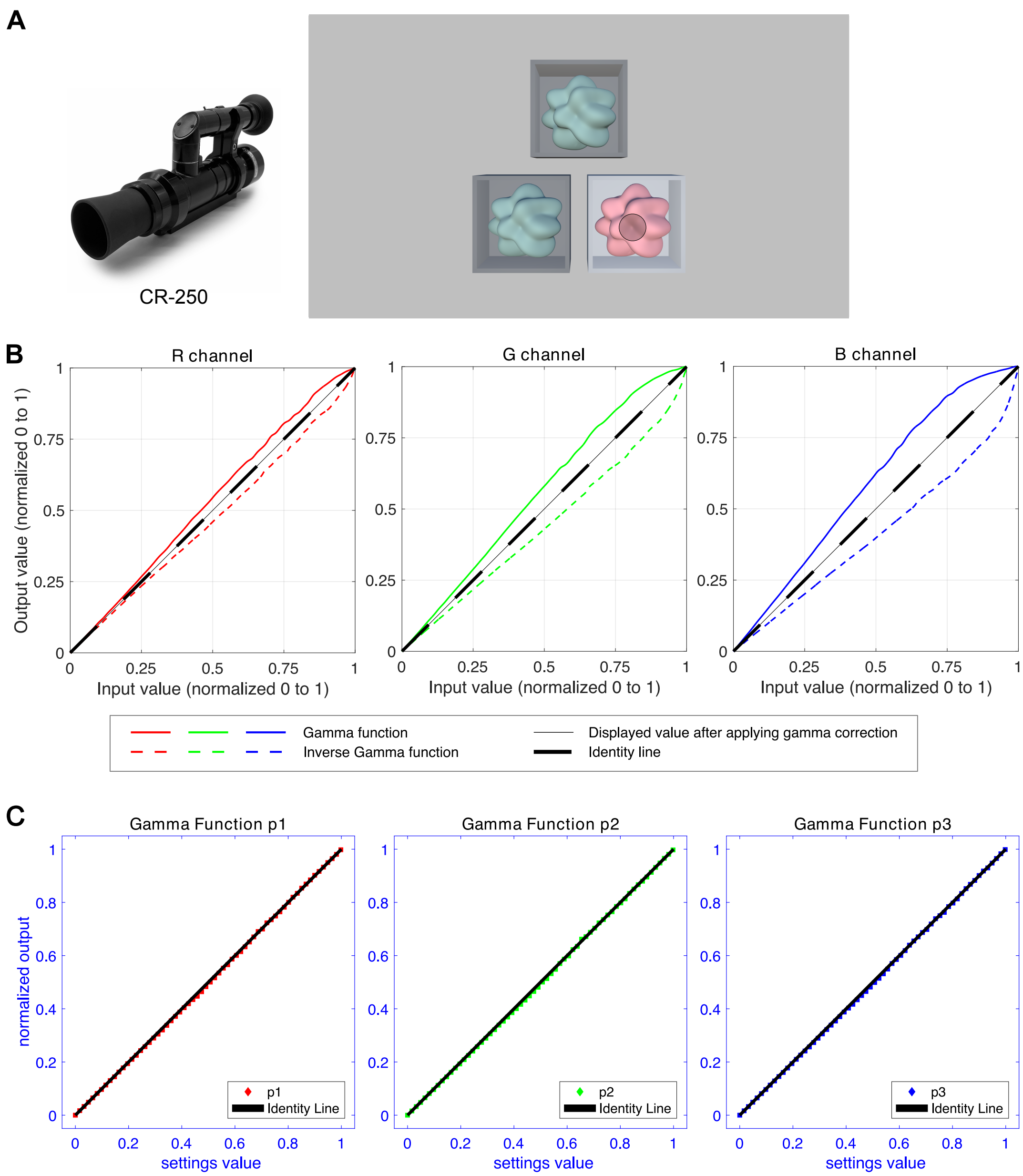

Figure 4. Gamma correction for the blobby stimulus. (A) A calibration scene. The CR-250 spectroradiometer was positioned at the same distance as the participant's chinrest to replicate experimental viewing conditions. The shaded gray circular region on the object indicates the measurement area captured by the radiometer's lens. The surface color of the cubic room and the blobby stimulus (illustrated as the bottom right stimulus) varied across trials for color calibration. (B) Gamma functions are first interpolated at 4096 levels (solid color lines), which are then used to derive inverse gamma functions (dashed color lines). The theoretically expected RGB values post-gamma correction align closely with the identity line. (C) Gamma functions measured after applying gamma correction, showing near-perfect alignment with the identity line across all primary colors.

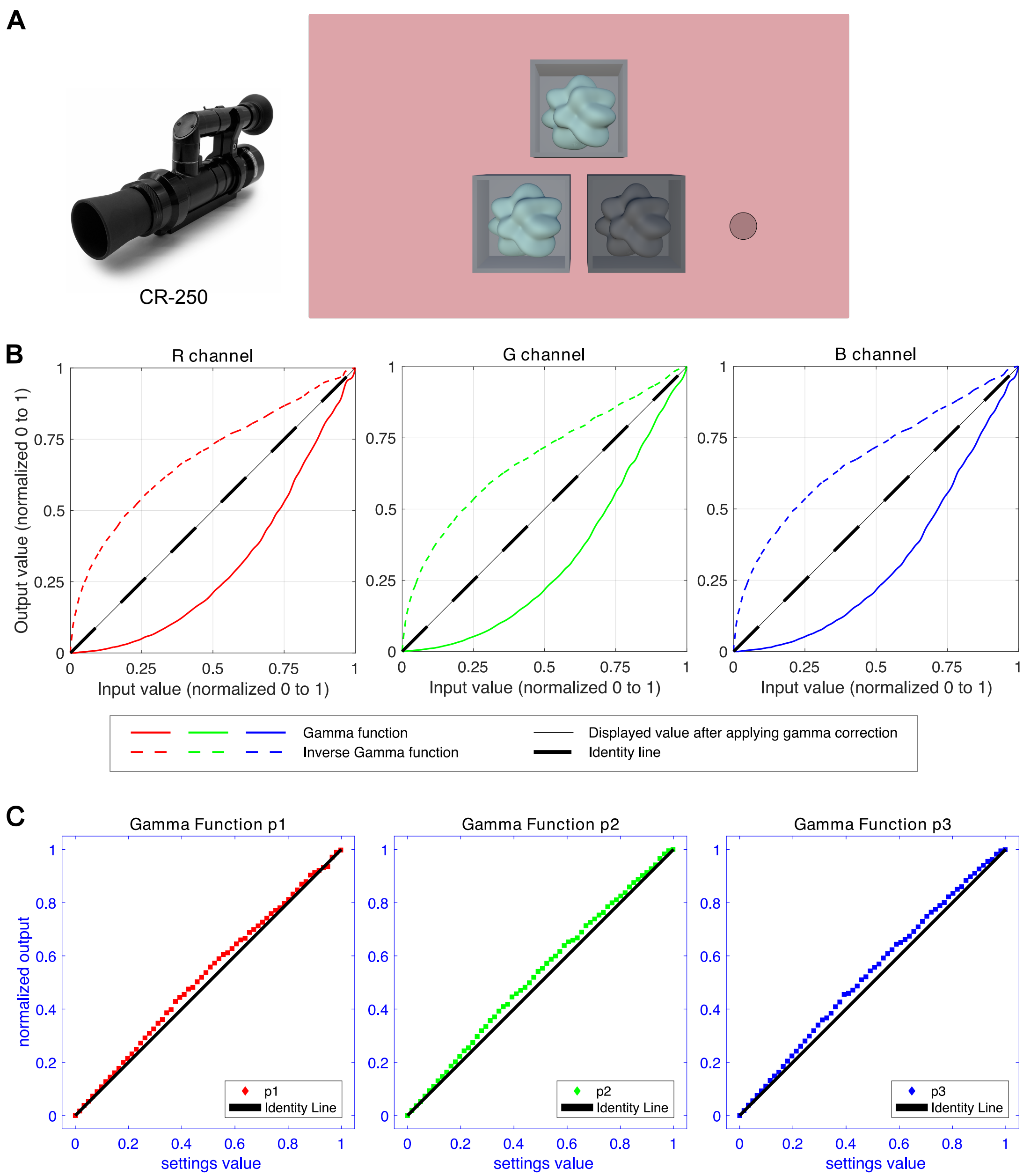

Figure 5. Gamma correction for the uniform background. (A) The uniform background was varied using this calibration task. The gray dot marks the approximate position and aperture of the CR-250. (B) Gamma functions are first interpolated at 4096 levels (solid color lines), which are then used to derive inverse gamma functions (dashed color lines). The theoretically expected RGB values post-gamma correction align closely with the identity line. (C) Gamma functions measured after applying gamma correction. Alignment with the identity line is not perfect across primaries but is acceptable for our purposes, as precise control of the blobby-stimulus display is the primary requirement.

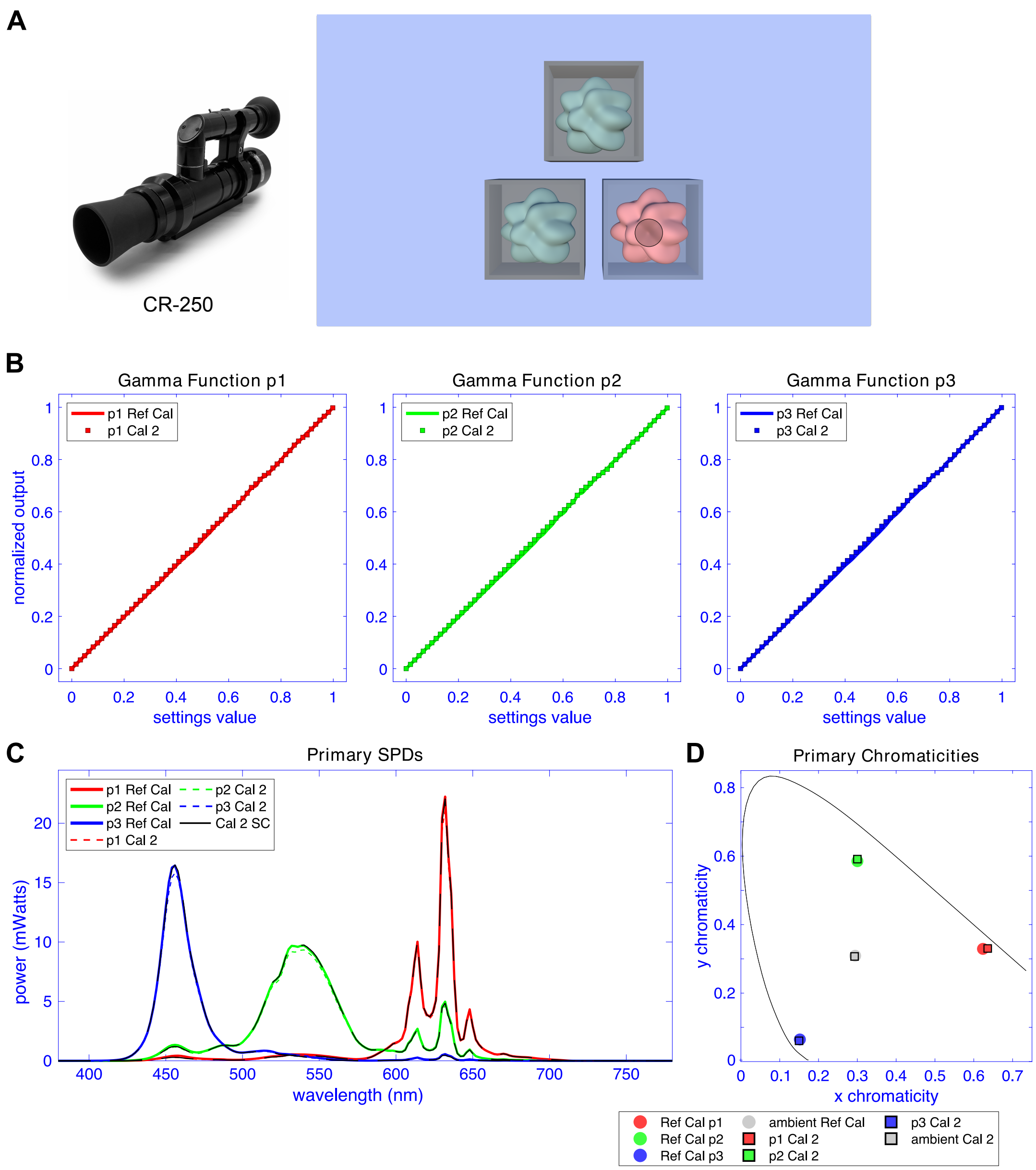

Figure 6. Calibration comparison between gray and blue adaptation conditions. (A) Task 3 repeated with the blue cubic room and background; the gray dot marks the CR-250 position/aperture. (B) Gamma functions for each primary overlaid for both conditions. (C) Spectral power distributions of the three primaries under each condition. (D) CIE xy chromaticity coordinates of each primary, compared across conditions.

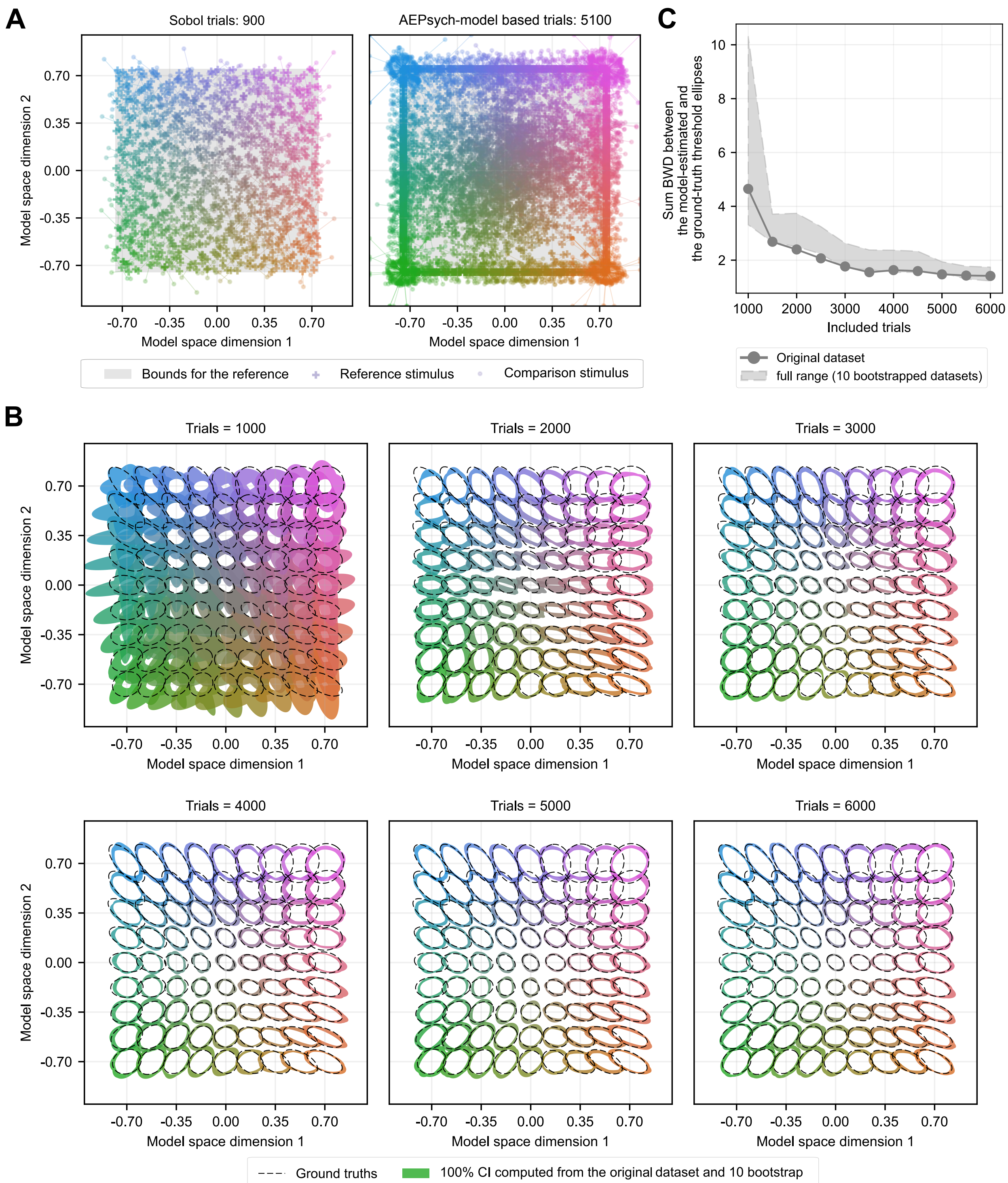

Figure 7. Trial efficiency. (A) Simulation of 6,000 AEPsych trials using ground-truth thresholds defined by CIELAB ΔE94. (B) WPPM fits to subsets of the data ranging from 1,000 to 6,000 trials; shaded bands show bootstrapped confidence intervals and the dashed line indicates ground truth. (C) Efficiency metric: for each reference in a 9 × 9 grid, the Bures–Wasserstein distance between ground truth and WPPM-predicted threshold ellipses is computed and summed across references.

Corner	DKL_{L-M}	DKL_S	DKL_{Lum}	R	G	B	W_{dim1}	W_{dim2}
1	-0.126	-0.816	0	0	0.735	0	-1	-1
2	0.182	-0.839	0	1	0.403	0	1	-1
3	-0.182	0.839	0	0	0.597	0	-1	1
4	0.126	0.816	0	1	0.265	1	1	1

Table 2. Corner vertices of the isoluminant plane in the DKL, RGB and model spaces.

$M_{DKL\text{->}W}$			$M_{RGB\text{->}W}$		
6.502	0.221	0	1.549	-1.361	-0.188
0.092	1.212	0	-0.451	-1.361	1.812
0	0	1	0.451	1.361	0.188

Table 3. Transformation matrices between DKL, RGB and the model spaces.

	Source	Practice	Experiment
Adaptation condition order	Python	sub# x 10,000 + session#	sub# x 100 + session#
Pre-generated Sobol trials	Python	N/A	sub# x 100 + cond# x 50 + session#
Sobol scalers	Python	sub# x 10,000 + cond#	sub# x 100 + cond#
Location of the odd stimulus	C#	sub# x 10,000 + session#	sub# x 100 + session#

Table 4. Scheme for selecting seeds for different participants from random-number generators. Four aspects of the experiment require shuffling: (1) the order of adaptation conditions for each session (0: adaptation is exactly the same as in the eLife paper; 1: blue adaptation point); (2) the sequence of pre-generated Sobol trials, which are inserted when AEPsych computations take too long, (3) the order of Sobol scalers (1/4, 2/4, 3/4) applied to Sobol-generated trials to balance task difficulty; and (4) the location of the odd stimulus, which can appear at the top, bottom left, or bottom right. (3) is determined only once during the first session, after which subsequent sessions reference the pre-determined indices. In contrast, the rest are generated separately for each session.