1 Supplementary Information

1.1 Supplementary Methods

1.1.1 Supplementary Methods 1: Replication of experiment under the illuminant 3000 K

The main experiment conducted under a 6500 Kelvin (K) illuminant was replicated using an alternative light source to evaluate the generalizability of the findings. Since physical gamut theory is predicated on the observer's internal estimate of the physical gamut associated with the relevant illuminant, it is noteworthy to validate the results under a secondary lighting condition. This replication allows for an assessment of whether perceptual judgements adapt accordingly and whether the high degree of correlation remains consistent. For this purpose, a 3000 K illuminant was selected, representing a significantly warmer, orangish spectral distribution.

We generated the theoretical physical gamut, the stimuli, and the real-world physical gamut following the same protocol as the main experiment, while substituting the illuminant during spectral integration with the spectral power distributions of Planckian radiators at 3000 K.

The experimental design and procedure mirrored those of the main experiment, subject to several specific modifications. Stimulus generation was conducted using a step size of 0.04 within the CIE 1931 (x, y) chromaticity plane, with luminance levels (Y) spanning 2 to 25 cd/m^2 across 15 increments of 1.5 cd/m^2 (the last increment being rounded to 25 cd/m^2). To account for the reduced data resolution in comparison to the main experiment, the ellipse radii utilized for smoothing observers' scalings (see Supplementary Methods 2) were increased to 0.08 for the (x, y) plane and 3.0 for the luminance (Y) axis. These adjustments resulted in a total of 5,256 trials, a reduction from the 30,348 trials used in the primary study. Seven observers participated, each completing two sessions consisting of 375 or 376 trials, resulting in a total of 750 or 751 trials per participant.

Figure 1 in the present material and its interactive 3D counterpart Supplementary Fig. S3 online (or Supplementary Fig. S6 online for a representation in the CIE 1976 $u'v'Y$ color space) illustrate the observers' judgements results and the theoretical physical gamut. Similarly, Fig. 2 in the present material and the associated interactive version Supplementary Fig. S3 online (or Supplementary Fig. S6 online for a representation in the CIE 1976 $u'v'Y$ color space) present the observers' judgements compared against the real-world physical gamut.

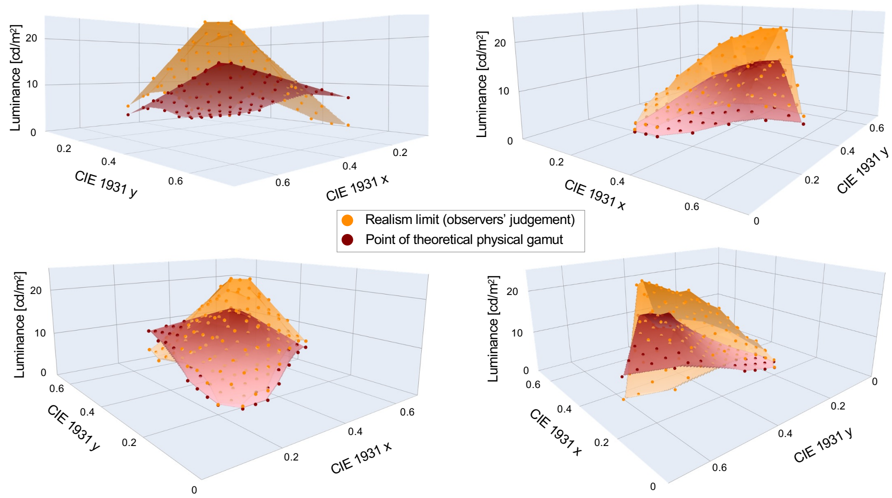

Figure 1: Observers' judgements of realism and theoretical physical gamut under illuminant 3000 K. The four panels show the same figure rendered from different viewing angles to facilitate its three-dimensional interpretation. The theoretical physical gamut (in red) and the observers' judgements of realism (in orange) are plotted in the CIE 1931 xyY color space for visual comparison. The points represent the actual data, while the surfaces are linear interpolations between these points intended to provide better visualization.

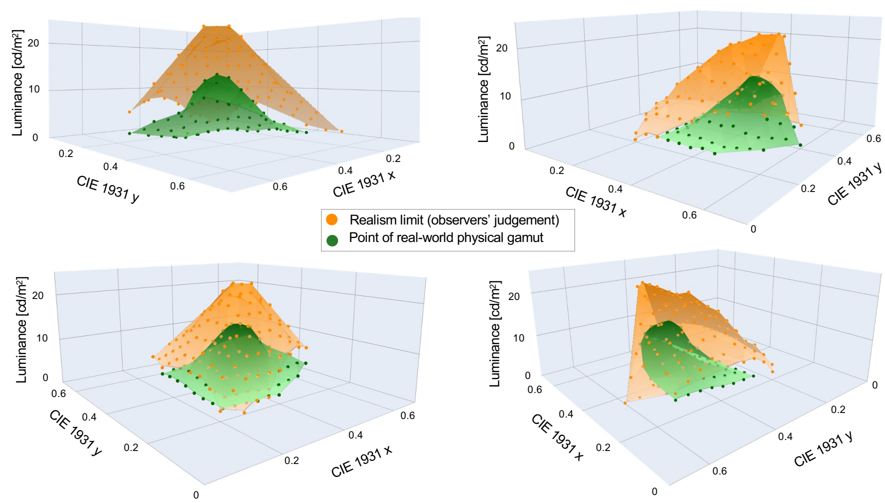

Figure 2: Observers' judgements of realism and real-world physical gamut under illuminant 3000 K. The four panels show the same figure rendered from different viewing angles to facilitate its three-dimensional interpretation. The real-world physical gamut (in green) and the observers' judgements of realism (in orange) are plotted in the CIE 1931 xyY color space for visual comparison. The points represent the actual data, while the surfaces are linear interpolations between these points intended to provide better visualization.

Following the analytical protocol established in the primary experiment, we calculated Pearson correlation coefficients to quantify the relationship between observers' judgements and the respective gamuts. The correlation between observers' judgements and the theoretical physical gamut was $r = 0.68$ ($p < 0.001$, [0.62, 0.72]). A stronger correlation was observed between observers' judgements and the real-world physical gamut, yielding $r = 0.82$ ($p < 0.001$, [0.80, 0.86]). Furthermore, when evaluating the theoretical physical gamut using only the data points corresponding to the real-world comparison, the calculation of the correlation yielded $r = 0.64$ ($p < 0.001$, [0.60, 0.74]).

In conclusion, the findings and trends observed under the 3000 K illuminant are highly consistent with those recorded under the 6500 K condition. This alignment provides further empirical support for our primary hypothesis and suggests that the observed phenomena are robust across varying illuminants. These results indicate that our main findings seem to generalize to other lighting conditions, remaining valid even when the spectral power distribution of the light source is significantly altered. In accordance with the physical gamut theory, observers appear capable of accurately estimating the corresponding internal reference to infer their perception and subsequent judgements.

1.1.2 Supplementary Methods 2: Smoothing of observers' judgements

This supporting information describes the mathematical explanation of how the observers' judgements were modeled and smoothed.

To smooth the realism judgements provided by observers, we employ a locally weighted linear regression model across the CIE 1931 xyY color space. For any target point $\mathbf{q} = [x_q, y_q, Y_q]$, the estimated realism score $\hat{j}$ is obtained by fitting a local linear model that prioritizes neighboring observations through a Gaussian weighting kernel.

The distance d_i between a target point $\mathbf{q}$ and an observation $\mathbf{p}_i = [x_i, y_i, Y_i]$ is defined by a normalized ellipsoidal metric:

$$d_i^2 = \frac{(x_i - x_q)^2 + (y_i - y_q)^2}{h_{xy}^2} + \frac{(Y_i - Y_q)^2}{h_Y^2} \tag{1}$$

The weight w_i assigned to each observation is then calculated using a radial basis function:

$$w_i = \exp\left(-\frac{1}{2}d_i^2\right) \tag{2}$$

The local linear model parameters $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3]^T$ are determined by minimizing the weighted least squares objective function:

$$\min_{\boldsymbol{\beta}} \sum_i w_i \left(j_i - (\beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 Y_i)\right)^2 \tag{3}$$

Finally, the smoothed realism judgement $\hat{j}$ at point $\mathbf{q}$ is given by the prediction of the fitted local model:

$$\hat{j} = \beta_0 + \beta_1 x_q + \beta_2 y_q + \beta_3 Y_q \tag{4}$$

The parameters of the model are defined as follows:

- xyY: Spatial coordinates in the CIE 1931 color space (x, y for chromaticity and Y for luminance);
- j: Raw realism judgements collected from observers;
- h_{xy}: Smoothing bandwidth (radius) for the chromaticity axes (x and y; value fixed at 0.04 for the main experiment under the illuminant 6500 K and 0.08 for the replication under the illuminant 3000 K, according to the logic detailed in the main article and in Supplementary Methods 1);
- h_Y: Smoothing bandwidth (radius) for the luminance axis (Y; value fixed at 2.3 for the main experiment under the illuminant 6500 K and 3 for the replication under the illuminant 3000 K, according to the logic detailed in the main article and in Supplementary Methods 1).

1.1.3 Supplementary Methods 3: Leave-One-Out (LOO) analysis

This section details the leave-one-out (LOO) analysis methodology employed to identify the data points that most deteriorate the Pearson correlation between the observers' judgements and the theoretical physical gamut in the main article.

First, we established a baseline Pearson correlation coefficient, r_{all}, using the complete set of chromaticities (x, y). For each judgement i, the correlation was recomputed upon its exclusion, denoted as r_{-i}. We defined the LOO influence of a specific point i as: $\Delta_i = r_{\text{all}} - r_{-i}$. Judgements characterized by more negative Δ_i values were those whose removal yields the greatest improvement in correlation; consequently, these points were identified as the most detrimental to the correlation. To characterize the improvement in correlation as a function of data exclusion, we ranked all points by Δ_i in ascending order (from most to least detrimental). We then defined $r(m)$ as the Pearson correlation coefficient calculated after the exclusion of the top m most influential points. The resulting $r(m)$ curve summarizes the trade-off between improving statistical agreement and preserving dataset coverage (see Fig. 3 in the main article).

To determine a principled threshold for data removal, we applied an elbow criterion to the $r(m)$ curve. Heuristically, the elbow represents the transition point toward diminishing marginal returns: prior to this point, the correlation increases sharply as systematic outliers are removed; beyond it, further exclusions merely simplify the dataset without substantially enhancing agreement. We objectively identified this point using the maximum distance to the chord method. A linear segment (chord) was drawn between the endpoints of the $r(m)$ curve over the evaluated range, and the optimal truncation level m^* was defined as the value of m where the perpendicular distance between the curve and the chord was maximized.

Crucially, the objective of this procedure was not to artificially optimize correlation through arbitrary data pruning, but rather to diagnose systematic model failures. Since each observation is associated with a specific chromaticity coordinate (x, y), mapping the m^* excluded points back onto the chromaticity plane facilitates the spatial localization of model discrepancies. If these influential points cluster within specific color regions (e.g., near the green primary), it provides the suggestion that the observed mismatch might be structured. Such clustering can indicate region-specific limitations of the theoretical model rather than noise.

1.1.4 Supplementary Methods 4: Additional analysis of the gamut of independent spectral databases

The analyses of the USGS [1] and ECOSTRESS [2, 3] spectral libraries presented in the main article appear to reveal an absence of highly saturated green chromaticities in the real world. This section presents an additional analysis conducted on two independent datasets, with the aim of assessing whether the same trend can also be observed in these measurements. The additional datasets analyzed are the *Giessen's hyperspectral images of fruits and vegetables* *database* [4], a hyperspectral database comprising 42 fruits and vegetables for which both the outer skin and the interior of the objects were imaged, and the *Fifty hyperspectral reflectance images of outdoor scenes database* [5], which contains 50 hyperspectral images of natural scenes acquired in the Minho region of Portugal.

The *Giessen's hyperspectral images of fruits and vegetables database* [4] directly provides the CIE 1931 xyY coordinates of each pixel, derived from the spectra measured in each hyperspectral image. These measurements were obtained under a neutral and stable illuminant, whose chromaticity is specified in the dataset. We then converted these coordinates using Bradford chromatic adaptation in order to represent them under a 6500 K illuminant, thereby making them comparable with the results presented in the main article. Although Bradford chromatic adaptation is not perfectly accurate, we considered it sufficiently appropriate for the purpose of this supplementary analysis. We selected the chromaticities corresponding to the exteriors of the 42 fruits and vegetables, as the interiors are more complex to account for within the scope of the present analysis. The chromaticity coordinates were rounded to two decimal places, and the pixels from all images were plotted independently on the CIE 1931 xy chromaticity diagram.

The *Fifty hyperspectral reflectance images of outdoor scenes database* [5] is the same dataset as that used to generate the four backgrounds in the experiments presented in the main article. Please refer to the methodological details provided in the main article for the relevant technical information. Here, we rendered the 50 images under the same 6500 K illuminant used in the main experiment, derived the CIE 1931 xy chromaticity coordinates of all pixels, rounded them to two decimal places, and then plotted the pixels from all images independently on the CIE 1931 xy chromaticity diagram.

The results are presented in Fig. 3. Similar to the analysis conducted on the USGS [1] and ECOSTRESS [2, 3] spectral libraries, these datasets do not contain chromaticities in the saturated green region that includes the points identified as being most detrimental to the correlation between observers' judgements and the theoretical physical gamut presented in the main article. This observation appears to support the hypothesis that these highly saturated green chromaticities are absent from what can be observed in the real world.

However, although this pattern strengthens the analysis and interpretation presented in the main article, these datasets still do not constitute an exhaustive representation of what can be observed in the real world. It is therefore important to emphasize that no conclusive statement can be made regarding the presence or absence of such highly saturated greens in real world.

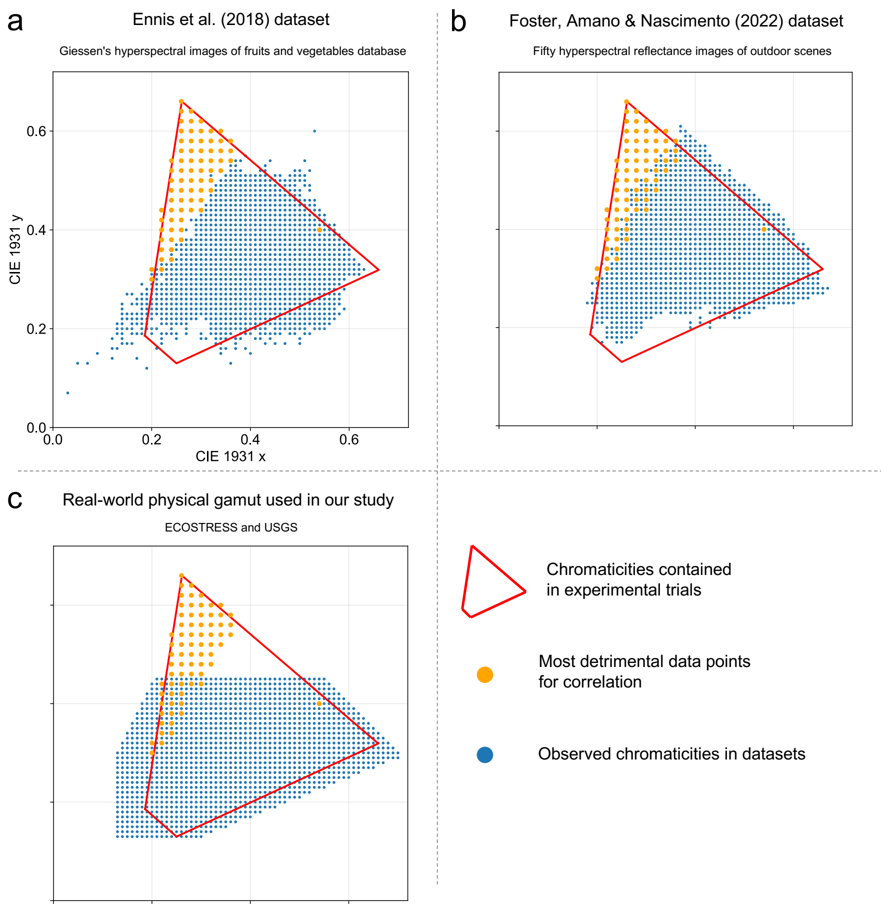

Figure 3: Chromatic distribution of images from the additional analyzed datasets. Panel **a** shows the *Giessen's hyperspectral images of fruits and vegetables database* [4]. Panel **b** shows the *Fifty hyperspectral reflectance images of outdoor scenes database* [5]. Panel **c** shows the real-world physical gamut dataset under the 6500 K illuminant, as used and already presented in the main article, for comparison purposes. The area outlined in red in each panel indicates the chromaticities included in the experiments reported in the main article. The orange points indicate the chromaticities that were most detrimental to the correlations observed in the main article between observers' judgements and the theoretical physical gamut derived from optimal colors.

1.1.5 Supplementary Methods 5: Analysis of real-world gamut and theoretical physical gamut spectra

To account for the reduced performance of the theoretical physical gamut in the region associated with highly saturated greens, the explanation proposed in the main article is that these highly saturated greens may be absent from the real world. Consequently, the human visual system may not have incorporated them into its internal references of physical gamuts. This explanation appears to be supported by the analysis of the chromatic distribution of the real-world physical gamut detailed in the main article. However, it does not directly explain what fundamentally distinguishes the real-world physical gamut from the theoretical physical gamut derived from optimal colors at the spectral level, rather than from the higher-level perspective of the resulting chromaticities. In this section, we provide a qualitative analysis of the spectra observed in these two datasets to gain more detailed understanding of this distinction.

To address this point, we selected spectra associated with chromaticities located at the most saturated boundaries of the theoretical physical gamut and the real-world physical gamut for the green, red, and blue regions. The results are presented in Fig. 4. For the real-world physical gamut, spectra associated with red or blue chromaticities exhibit more pronounced peaks or profiles that resemble step functions. Such profiles are generally characteristic of higher saturation. By contrast, spectra associated with green chromaticities are comparatively flatter, which is generally characteristic of duller and less saturated chromaticities. For the theoretical physical gamut, the spectra correspond to those derived from optimal colors. As explained in the main article, optimal colors are, by definition, characterized by step-function-like spectral profiles. As noted above, such profiles are characteristic of highly saturated chromaticities. It is therefore not surprising that the theoretical physical gamut exhibits more saturated colors than the real-world physical gamut. However, it is noteworthy that the difference between the real-world physical gamut and the theoretical physical gamut is more pronounced for spectra associated with green chromaticities than for those associated with red or blue chromaticities. These observations may therefore provide a spectral-level explanation for the differences in saturation observed between the real-world physical gamut and the theoretical physical gamut in the green region.

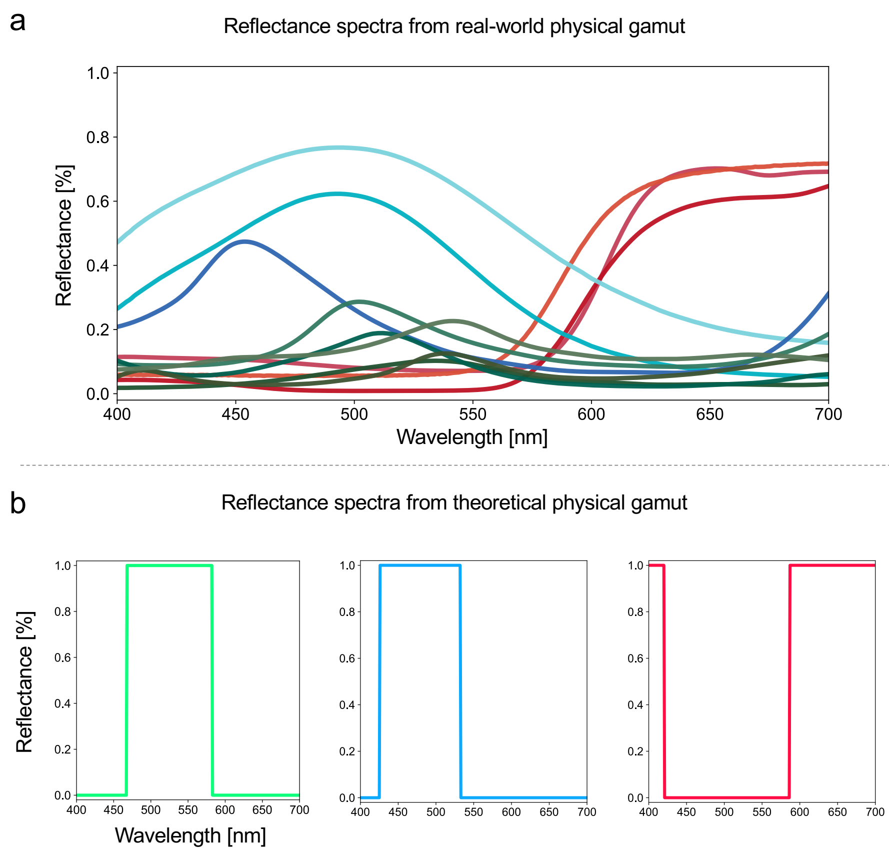

Figure 4: Comparison of spectra associated with chromaticities located at the most saturated boundaries of the theoretical physical gamut and the real-world physical gamut. Panel **a** shows examples of spectra from the real-world physical gamut. Panel **b** shows examples of spectra from the theoretical physical gamut. The spectra from the theoretical physical gamut are displayed in separate plots to facilitate visualization. In each panel, the color of each line provides an indicative sRGB rendering of the color associated with the corresponding spectrum, providing an approximate reference for comparison.

1.1.6 Supplementary Methods 6: Analysis and results of correlation calculations using raw data

The analysis presented in the main article computes the Pearson correlations between the physical gamuts and the smoothed observers' judgements. We explain that, to robustly capture the underlying trends in observers' judgements and mitigate the noise inherent to the large-scale experimental design, a multi-stage smoothing procedure was necessary and was therefore applied to the data.

To quantify this noise and assess the impact of smoothing, this section presents the Pearson correlation coefficients between the physical gamuts and the raw observers' judgements. This supplementary analysis was conducted for both the main experiment under the 6500 K illuminant and the additional experiment under the 3000 K illuminant. The results are reported in Table 1. The first results column presents the correlations obtained with the smoothed observers' judgements, as introduced in the main article, whereas the second results column reports the correlations obtained with the raw observers' judgements, without any smoothing. Apart from the removal of the smoothing procedure, all other aspects of the analysis pipeline and correlation computation remain identical to those described in the main article.

Although the correlations observed with the raw data are lower by 0.16 points on average, this difference can be considered relatively modest in the present context. Consequently, the overall trends, conclusions, and interpretations remain unchanged compared with those obtained using the smoothed data. These results confirm that smoothing the observers' judgements does not artificially generate correlations, but rather serves to reduce noise inherent to the large-scale of our experiments.

Notably, the results obtained under the 3000 K illuminant appear to be more strongly affected by smoothing. This discrepancy can be explained by the fact that, in the replication of the main experiment under the 3000 K illuminant, we substantially reduced the number of conditions and participants. This reduction introduced greater instability into an experimental protocol that relies on a fine-grained coverage of the entire color space, achieved through a large number of experimental trials and participants, in order to reveal stable trends in the results. In this case, smoothing is therefore particularly important for extracting the underlying trend.

Table 1: Pearson correlation coefficients between physical gamuts and observers' judgements. The p-values and 95% confidence intervals (indicated in brackets) were obtained using BCa bootstrapping across observers, as described in the main article.

	Observers' judgements	
	Smoothed data (presented in main article)	**Raw data**
Theoretical physical gamut – 6500 K	0.63 (p < 0.001, [0.54, 0.71])	0.54 (p < 0.001, [0.49, 0.65])
Theoretical physical gamut – 3000 K	0.68 (p < 0.001, [0.62, 0.72])	0.53 (p < 0.001, [0.48, 0.60])
Theoretical physical gamut, without most detrimental points – 6500 K	0.89 (p < 0.001, [0.84, 0.92])	0.78 (p < 0.001, [0.73, 0.82])
Real-world physical gamut – 6500 K	0.89 (p < 0.001, [0.85, 0.92])	0.75 (p < 0.001, [0.72, 0.79])
Real-world physical gamut – 3000 K	0.82 (p < 0.001, [0.80, 0.86])	0.61 (p < 0.001, [0.53, 0.71])
Theoretical physical gamut, same subset as real-world physical gamut – 6500 K	0.80 (p < 0.001, [0.73, 0.85])	0.68 (p < 0.001, [0.67, 0.76])
Theoretical physical gamut, same subset as real-world physical gamut – 3000 K	0.64 (p < 0.001, [0.60, 0.74])	0.37 (p < 0.001, [0.26, 0.52])

1.1.7 Supplementary Methods 7: Correlation analyses across background conditions

This section presents the Pearson correlation coefficients observed between the physical gamuts and observers' judgements in the main experiment, separately for each background condition. The results are presented in Table 2. For reference, the background conditions are illustrated in Fig. 5. The background condition IDs reported in Table 2 match the IDs assigned to the backgrounds in Fig. 5. Apart from the separation by background condition, the analysis strategy and correlation computation remain identical to those described in the main article.

Overall, the Pearson correlations observed across background conditions show relative stability, suggesting that background had only a limited effect on the results and that the human visual system remains relatively consistent in estimating the physical gamut across backgrounds. These results are consistent with previous findings from studies on luminosity thresholds and physical gamuts [6, 7].

Table 2: Pearson correlation coefficients between observers' judgements and physical gamuts per background condition under illuminant 6500 K. The p-values and 95% confidence intervals (indicated in brackets) were obtained using BCa bootstrapping across observers, as described in the main article.

	Background condition ID			
	1	**2**	**3**	**4**
Theoretical physical gamut	0.70 $p < 0.001$, [0.67, 0.76]	0.61 $p < 0.001$, [0.55, 0.70]	0.51 $p < 0.001$, [0.42, 0.61]	0.66 $p < 0.001$, [0.62, 0.71]
Theoretical physical gamut, without most detrimental points	0.89 $p < 0.001$, [0.87, 0.92]	0.87 $p < 0.001$, [0.84, 0.91]	0.80 $p < 0.001$, [0.76, 0.86]	0.88 $p < 0.001$, [0.85, 0.91]
Real-world physical gamut	0.83 $p < 0.001$, [0.80, 0.85]	0.84 $p < 0.001$, [0.81, 0.89]	0.86 $p < 0.001$, [0.81, 0.90]	0.91 $p < 0.001$, [0.88, 0.93]
Theoretical physical gamut, same subset as real-world physical gamut	0.78 $p < 0.001$, [0.75, 0.83]	0.76 $p < 0.001$, [0.70, 0.83]	0.74 $p < 0.001$, [0.64, 0.81]	0.86 $p < 0.001$, [0.83, 0.88]

Figure 5: Stimuli used in our experiments. The backgrounds and the targets are shown, with the target highlighted in blue to clearly illustrate which part of the image was changed throughout the trials.

1.1.8 Supplementary Methods 8: Monitor gamut constraints and observers' judgements

The aim of this section is to demonstrate that the observers' judgements were not constrained by the colorimetric limits of the monitor. Indeed, the background images, as well as the full set of test colors applied to the targets across the experimental trials, were selected in the CIE 1931 xyY color space so as to satisfy the monitor constraints without requiring clipping or incurring any loss of information during the conversion to RGB values. This was made possible by the color calibration procedure described in the main article.

Figure 6 shows the full set of test colors applied to the targets across the experimental trials, together with the limits of perceptual realism derived from the observers' judgements, in the CIE 1931 xyY color space. In addition, the red, green, and blue vertical lines denote the chromaticities of the monitor's red, green, and blue primaries, respectively. The measured monitor primaries were located at $R = (0.677, 0.324)$, $G = (0.145, 0.753)$, and $B = (0.125, 0.053)$ in the CIE 1931 xy chromaticity diagram. For reference, the monitor's maximum luminance was measured at 185 cd/m^2, a limit well above the values used in our experiments.

These results clearly show that the realism limit cannot be explained simply by the monitor gamut, given that the realism extent occupies only a restricted subset of the available displayable colors. Moreover, the set of test colors used in our experiments itself represents only a subset of the monitor's full capabilities, which extend beyond these limits.

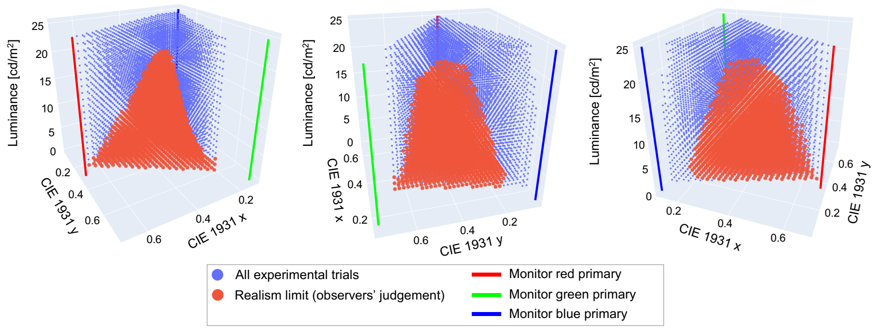

Figure 6: Set of test colors and observers' judgements. The three plots show the same figure rendered from different viewing angles to facilitate its three-dimensional interpretation. The full set of test colors applied to the targets across the experimental trials (in blue) and the limit of perceived realism derived from the observers' judgements (in orange) are plotted in the CIE 1931 xyY color space. The red, green, and blue vertical lines denote the chromaticities of the monitor's red, green, and blue primaries, respectively.

References

[1] Raymond F Kokaly, Roger N Clark, Gregg A Swayze, K Eric Livo, Todd M Hoefen, Neil C Pearson, Richard A Wise, William Benzel, Heather A Lowers, Rhonda L Driscoll, et al. USGS Spectral Library Version 7. *U.S. Geological Survey Data Series*, (1035):61, 2017.

[2] Susan K Meerdink, Simon J Hook, Dar A Roberts, and Elsa A Abbott. The ecostress spectral library version 1.0. *Remote Sensing of Environment*, 230:111196, 2019.

[3] Alice M Baldridge, Simon J Hook, CI Grove, and GJRSoE Rivera. The aster spectral library version 2.0. *Remote sensing of environment*, 113(4):711–715, 2009.

[4] Robert Ennis, Florian Schiller, Matteo Toscani, and Karl R Gegenfurtner. Hyperspectral database of fruits and vegetables. *Journal of the Optical Society of America A*, 35(4):B256–B266, 2018.

[5] David Foster, Kinjiro Amano, and Sérgio Nascimento. Fifty hyperspectral reflectance images of outdoor scenes, 1 2022. Available at: https://figshare.manchester.ac.uk/articles/dataset/Fifty_hyperspectral_reflectance_images_of_outdoor_scenes/14877285.

[6] Takuma Morimoto, Ai Numata, Kazuho Fukuda, and Keiji Uchikawa. Luminosity thresholds of colored surfaces are determined by their upper-limit luminances empirically internalized in the visual system. *Journal of vision*, 21(13):3–3, 2021.

[7] Killian Duay and Takehiro Nagai. Optimal colors can predict luminosity thresholds in natural scenes. *Plos one*, 21(3):e0343984, 2026.