SUPPLEMENTARY MATERIAL


MEASUREMENTS .................................................................................................................................................................. 2
Discrimination ...................................................................................................................................................................... 2
Raw JNDs for each observer (fig. S1) ............................................................................................................................ 2
Differences across staircases (fig. S2) ........................................................................................................................... 4
Individual JNDs (fig. S3) ................................................................................................................................................. 5
JNDs calculated as Euclidean differences (fig. S4) ........................................................................................................ 5
Ellipses (fig. S5) .............................................................................................................................................................. 6
Categorization ..................................................................................................................................................................... 7
Color naming results (fig. S6) .......................................................................................................................................... 7
Individual differences in color naming (tab. S1) ........................................................................................................... 7
Prototype adjustments (fig. S7) ....................................................................................................................................... 7
Alternative naming task ................................................................................................................................................... 9
Combination (fig. S8) ........................................................................................................................................................ 10

RESULTS .............................................................................................................................................................................. 11
Relative JNDs for individual observers ............................................................................................................................. 11
Relative JNDs for isoluminant colors (fig. S9) ................................................................................................................ 11
Relative JNDs for dark and light colors (fig. S10) ........................................................................................................... 12
Aggregated JNDs and consensus categories (fig. S11) ................................................................................................... 14
Individual JNDs and categories ......................................................................................................................................... 14
Individual Boundary Tests (fig. S12) ............................................................................................................................. 14
Other individual tests (fig. S13) ..................................................................................................................................... 15

DISCUSSION ........................................................................................................................................................................ 17
Categorical sensitivity and perceptual mechanisms (fig. S14) ......................................................................................... 17
Generalization to other color spaces (fig. S15) ............................................................................................................... 17
MEASUREMENTS

Discrimination

Raw JNDS for each observer (fig. S1)

a.) isoluminant
Figure S1. Raw JNDs. JNDs resulting from azimuth increasing (green) and azimuth decreasing (red) staircases are shown separately for each observer. Format as in fig. 3 of the main article. Panels a-c show data for isoluminant, dark, and light colors, respectively. For all participants, the profile of both JND curves is very similar.
Differences across staircases (fig. S2)

Figure S2 illustrates the differences between JNDs for increasing and decreasing staircases. The green and the red curves show the JNDs obtained through the azimuth increasing and decreasing staircases, respectively. They were averaged across repeated measurements and across participants (see Figure S1 for single observers).

The average increasing and decreasing JNDs were highly correlated across the n = 72 isoluminant (r = 0.76, p < 0.001, n = 10), the 36 dark (r = 0.64, p < 0.001, n = 4), and the 12 light hues (r = 0.90, p < 0.001, n = 2). The congruence between increasing and decreasing JNDs was high for every single participant (cf. fig. S1). When calculating these correlations for each observer separately, only the correlation for 1 of the 4 participants with dark colors was not significant (f3), and 1 (f8) of 10 with isoluminant colors was only marginally significant. All other pairwise correlations at all lightness levels were significant. These results show that both kinds of staircases were highly consistent with each other.

However, we also observed differences between increasing and decreasing JNDs. Figure S2 shows that the curves for the increasing and decreasing measurements were slightly shifted away from each other by about 5 to 10 deg along the x-axis. These differences correspond to about 1 JND. That the curves of increasing and decreasing JNDs are shifted by about 1 JND (5-10 deg) away from each other is to be expected. The midpoints of a test and comparison color at threshold are half a JND away from the test color in the direction of measurement. For two opposite measurement directions this implies an overall shift of 1 JND, such as the one we observed.

Moreover, in a paired t-test across participants we found that for isoluminant colors increasing staircases yielded on average 0.36° higher JNDs than decreasing staircases (t(9) = 2.9, p < 0.02). When tested for each participant by a t-test across hues (red line - green line in fig. S2), the difference between increasing and decreasing JNDs was significant for 4 of 10 observers with isoluminant colors (cw, m2, f3, and f7), for 3 of 4 with dark colors (cw, f2, f3), and for both participants with light colors (cw & f1). This last result is surprising since increasing and decreasing differences refer to opposite hue directions, not intensities.

However, these differences only concern the absolute value of the JNDs. The high correlations show that the two kinds of staircases measure overall the same profiles of JNDs.

Figure S2. JND variability. JNDs resulting from azimuth increasing (green) and azimuth decreasing (red) staircases are shown separately. Axes as in fig. 3. The thick lines show the JNDs as averaged across repeated measurements and observers. Pale shaded areas depict SEM across participants. Panels a-c show data for isoluminant, dark, and light colors, respectively. The profiles of both JND curves are very similar, but shifted away from each other by about 1 JND.
**Individual JNDS (fig. S3)**

a.) Dark

![Graph showing individual JNDS for dark colors.](image)

b.) Light

![Graph showing individual JNDS for light colors.](image)

Figure S3. Individual variation of JNDS for dark and light colors. Format as in fig. 3. Panel a shows data for dark, panel b for light colors. As for isoluminant colors, JND profiles were highly similar across observers.

**JNDS calculated as Euclidean differences (fig. S4)**

![Graph showing aggregated JNDS calculated as Euclidean distances.](image)

Figure S4. Euclidean and azimuth differences. Aggregated JNDS for isoluminant colors are calculated as Euclidean distances (red line) and as differences in azimuth (black line) in DKL-space. Format as in fig. 3 of the main article. The left y-axis refers to the JNDS calculated as Euclidean distances, the right axis to those calculated as azimuth differences. Note that the curves overlap almost completely, indicating that there is barely any difference in profile between the two ways to calculate JNDS.
Figure S5. JND-ellipses for dark and light colors. Panel a. and b. show the fitted ellipses for the JNDs of the dark, panels c. and d. for the light colors. Format as in figure 5 of the main article. Note that dark and light colors do not yield the same patterns as isoluminant colors. JNDs for dark colors follow the ellipse; but unlike with isoluminant colors residuals do not show pronounced minima at the DKL-axes. The JNDs for light colors result in stronger deviations from the ellipse and do not involve minima at the axes.
Categorization

**Color naming results (fig. S6)**

![Figure S6. Color naming for dark and light colors. Format as in fig. 6 of the main article. Panel a shows results for dark, panel b for light colors.](image)

**Individual differences in color naming (tab. S1)**

**a.**

<table>
<thead>
<tr>
<th>Dark Category</th>
<th>Boundaries</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ε</td>
<td>df1</td>
<td>df2</td>
<td>F</td>
<td>p</td>
</tr>
<tr>
<td>Pink</td>
<td>0.25</td>
<td>1.5</td>
<td>6.0</td>
<td>3.4</td>
<td>0.11</td>
</tr>
<tr>
<td>Red</td>
<td>0.34</td>
<td>2.0</td>
<td>8.1</td>
<td>4884 ***</td>
<td></td>
</tr>
<tr>
<td>Green</td>
<td>0.23</td>
<td>1.8</td>
<td>7.3</td>
<td>10.9 **</td>
<td></td>
</tr>
<tr>
<td>Blue</td>
<td>0.35</td>
<td>2.8</td>
<td>11.2</td>
<td>29.3 ***</td>
<td></td>
</tr>
<tr>
<td>Purple</td>
<td>0.38</td>
<td>3.0</td>
<td>12.1</td>
<td>10.4 ***</td>
<td></td>
</tr>
<tr>
<td>Brown</td>
<td>0.33</td>
<td>2.6</td>
<td>10.4</td>
<td>33.3 ***</td>
<td></td>
</tr>
</tbody>
</table>

**b.**

<table>
<thead>
<tr>
<th>Light Category</th>
<th>Boundaries</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ε</td>
<td>df1</td>
<td>df2</td>
<td>F</td>
<td>p</td>
</tr>
<tr>
<td>Pink</td>
<td>0.24</td>
<td>1.9</td>
<td>7.7</td>
<td>2.9</td>
<td>0.12</td>
</tr>
<tr>
<td>Orange</td>
<td>0.36</td>
<td>2.9</td>
<td>11.6</td>
<td>16.9 ***</td>
<td></td>
</tr>
<tr>
<td>Yellow</td>
<td>0.34</td>
<td>1.7</td>
<td>11.0</td>
<td>7.4 **</td>
<td></td>
</tr>
<tr>
<td>Green</td>
<td>0.33</td>
<td>2.7</td>
<td>10.7</td>
<td>9.9 **</td>
<td></td>
</tr>
<tr>
<td>Blue</td>
<td>0.35</td>
<td>2.8</td>
<td>11.2</td>
<td>14.2 ***</td>
<td></td>
</tr>
<tr>
<td>Purple</td>
<td>0.41</td>
<td>3.3</td>
<td>13.0</td>
<td>13.5 ***</td>
<td></td>
</tr>
</tbody>
</table>

**Table S1. Individual differences of category boundaries for dark and light colors.** Format as in tab. 1 of the main article. Part a: Dark colors. Analyses included 5 repeated measurements for 9 participants (without f5), except for pink and red, where only 7 participants had boundaries for 5 repeated measurements (without cw & f6; cf. fig. S6.a). Part b: Light colors. Analyses for all categories included 5 repeated measurements for all 9 participants (without f5, cf. fig. S6.b).

**Prototype adjustments (fig. S7)**

**METHOD.** Before each adjustment, target descriptions were shown in black on the isoluminant gray background. Hue could be shifted clockwise and counterclockwise along the isoluminant color circle by pressing the right and the left cursor key, respectively. Each key press shifted the hue by 1 deg azimuth. This resolution allowed for a sufficiently fine-grained adjustment of the colors given the height of JNDs (cf. “Discrimination” section).

In the condition with adjustable lightness, lightness could be changed by pressing the up and down cursor keys.
This, however, did not change the luminance of the stimuli. For the same reasons as described before (“General method” section), pressing the up key would decrease the luminance of the background until it is black (DKL value of –1). In contrast, pressing the down key would increase the luminance of the background until it is white (DKL value of +1). When arriving at the limits of the gamut (–1 or +1), an announcement informed the observer that they had reached the maximum in the respective direction. While the start hue was random, the start lightness was isoluminant with the stimuli.

Participants could also press the shift key while pushing one of the adjustment keys to continuously change the color. This option allowed them to quickly shift the color to the approximate region of the prototype before fine-adjusting it to the exact prototype. They pressed “enter” to confirm their adjustment and to start the next trial.

a.)

![Dark colors (White background)](image)

b.)

![Light colors (Black background)](image)

**Figure S7. Color categories and their prototypes for dark and light colors.** Format as in figure 7 of the main article. Panel a shows results for dark, panel b for light colors. Note the consistency between categories and the two kinds of prototype adjustments.

**RESULTS.** We assessed the validity of the prototype adjustments in general, and tested which prototype measurement is more representative for the category sets across lightness levels (cf. disks and triangles in figure 7 and S7).

If prototype adjustments were sensible they should lie within the respective categories. In a first approach, we counted how many prototypes lay outside the categories with each adjustment method. For isoluminant colors, 4 of 54 prototypes (orange, 2 yellow, purple) were outside the respective category borders when luminance was fixed, and 2 (both blue) when luminance could be adjusted. For dark colors there were 9 of 47 prototypes outside the categories when luminance was fixed (3 pink, 3 red, 1 orange, 1 yellow, & 1 purple), and 9 of 55 with adjustable luminance (3 pink, 2 red, 1 yellow, 2 blue, 1 brown). For light colors, both conditions yielded 3 of 55 prototypes outside the boundaries (fixed luminance: 2 orange, 1 purple; adjustable: 1 red, 2 blue). Averaging the two measurements (i.e. the x-values of the disks and triangles in fig. 7 and S7) did not reduce the amount of prototypes outside the category borders (isoluminant: 3, dark: 8, black background: 3).

To further evaluate the differences between the two kinds of adjustments, we calculated a paired t-test for each category that tested across the 9 participants whether the differences between the adjustments were systematically different from zero. For the blue category, the isoluminant adjustments were on average closer to red by 8.3 deg (t(8) = 2.52, p = 0.04). The shift of red towards pink in the isoluminant adjustments was marginally significant (mean 10.4 deg, t(8) = -2.52, p = 0.07). There was no significant difference for any other category (all other |t(8)| < 1.21; p > 0.26).

In order to assess the relationship between hue shifts and luminance, we tested whether luminance adjustments were in general lower than isoluminance and whether they were shifted to a particular hue direction compared to isoluminant adjustments. We applied a paired t-test across all categories and all participants. We only excluded brown because it did not exist for isoluminant adjustments. Luminance adjustments were on average 13% (in DKL units) darker than isoluminance, which was only marginally significant (t(69) = –1.7, p = 0.09). At the same time, isoluminant adjustments tended to be shifted by on average -3.9 deg away from the adjustments with variable luminance t(62) = 2.6, p = 0.01). We therefore tested for each category whether the differences between the two kinds of adjustments correlated with the luminance adjustments in the condition with adjustable luminance. Only for the adjustments of red, there was a marginally significant correlation across the 9 participants (r = –0.63, p = 0.06; all other p > 0.12).

We evaluated whether there were systematic variations of prototypes across participants. Since few measurements were available, we calculated correlations between the two measurements across the 9 participants. If the variation of adjustment differs systematically across individuals there should be positive correlations. Hence, the test of signifi-
cance was one-tailed. Correlation coefficients were positive for all categories. However, only for green and purple, these correlations were significant (r = 0.66, p = 0.03, and r = 0.71, p = 0.02); for blue the correlation was only marginally significant (r = 0.50, p = 0.08; all other p > 0.12).

Finally, we calculated correlations of the prototypes with each of the two boundaries of the respective category. We expected positive correlations since a shift of any of the two category borders should imply a shift of the prototype in the same direction. However, this pattern did not appear consistently. To illustrate this, we report two-tailed statistics (all n = 9). Most significant correlations occurred for the adjustments with fixed luminance. However, there was only one significant positive correlation of the blue prototype with the light upper boundary (r = 0.83, p = 0.006), and a marginally significant correlation of the red prototype with the upper dark boundary (r = 0.62, p = 0.08). The other 4 correlations were significantly negative: purple at isoluminant (r = -0.69, p = 0.04) and light lower boundary (r = -0.68, p = 0.04), yellow at isoluminant upper boundary (r = -0.70, p = 0.04), and orange at the light upper boundary (r = -0.65, p = 0.06). For adjustments with variable luminance there were only significant correlations for the brown category, which were negative for the lower (r = -0.87, p = 0.005) and positive for the upper dark boundary (r = 0.73, 0.04; both n = 8). Averaging the two kinds of adjustments should reduce measurement noise. However, no correlation reached significance when using average adjustments.

**DISCUSSION.** In sum, prototype adjustments were generally in line with the categories in that almost all adjusted prototypes lay within the corresponding categories. Dark colors seem to yield more adjustments outside the categories (8 or 9). Half of these (4–5) occurred for pink, orange and yellow, which are not typical categories for dark colors.

However, there was little evidence that adjustments varied as a function of overall category shifts, as would be indicated by positive correlations with the respective boundaries. There were some such correlations, in particular for blue in the adjustments with fixed luminance (69% of total variance), and for brown in the condition with variable luminance (76% & 53%). However, given the multiple testing, the presence of only a few such correlations may just occur due to statistical noise.

Positive correlations across participants indicate that the variation of prototypes across participants is systematic. This was at least the case for green (44% of total variance) and purple (50%), and to some extend also for blue (25%). All the other categories also yielded positive correlation coefficients. Even though these were not statistically significant, they show that the correlations for green, purple, and blue are not the result of multiple testing. This suggests that different observers have different prototypes for green, purple and blue.

The comparison between the two conditions of prototype adjustments provided some evidence that prototypes change systematically depending on lightness (on average about 3.9 deg). This was mainly the case for blue, and maybe for red (differences of 8.3 and 10.4 deg). At the same time, the correlations across participants also indicate that the two measurements agree with each other, at least for green and purple. Finally, the correlations with the category borders did not provide convincing evidence that one or the other adjustment condition yielded more reliable results. Taken together, it seems that the average prototypes from both adjustment conditions are the most reliable and representative because averaging reduces measurement noise.

**Alternative naming task**

Three of the 10 participants took part in the alternative naming task.

The hue of the presented stimulus color was adapted through a staircase procedure. The initial color of a staircase was located at the center of one of the two categories adjacent to the boundary that was to be measured. Depending on the participant’s answer the test color was shifted along the hue circle either towards one or the other color. In this way, the test color converged towards the boundary during the staircase. The background was achromatic and isoluminant to the hue circle. Stimulus presentation was the same as in the main naming task (cf. section 2.3).

We employed a 2-Alternative-Forced Choice (2AFC) task. In one trial, participants were shown the uniformly colored disks. They had to indicate whether the color belongs to one or the other category by pressing one or another key. A 1-up-1-down staircase was employed to obtain the boundary where the color belongs to 50% to one or the other category. The boundary between pink and red as well as between red and orange was measured. For each boundary, two staircases were measured, each starting in one and the other adjacent category.
Figure S8. Individual data for dark and light colors. Format as in fig. 8 of the main article. For observers whose JNDs were not measured but interpolated, ids are shown in brackets. The right y-axis is scaled from 0 to 25. For original (non-interpolated) datasets SEM across the 4 staircases are shown by the gray shaded area. Panel a presents the results for the dark colors. Thresholds were interpolated for participants m2, f4-8. Panel b shows the results for the light colors. Here, thresholds were interpolated for all participants except cw and f1.
RESULTS

Relative JNDs for individual observers

Relative JNDs for isoluminant colors (fig. S9)

Figure S9. Relative JNDs of each observer for isoluminant colors. Format as in fig. 11 of the main article, with the exception that boundaries are shown as black vertical lines instead of dots. These relative JNDs were used for the individual categorical perception tests in fig. 12, and panels a, d, and g in S13.
Relative JNDs for dark and light colors (fig. S10)
Figure S10. Relative JNDs of each observer for dark and light colors. Format as in fig. S9. Panels a-j show the results for dark colors, panels k-t those for light colors. These relative JNDs were used for the individual categorical perception tests in fig. S12 and S13.
Aggregated JNDs and consensus categories (fig. S11)

Figure S11. Relative JNDs for dark and light colors. Format as in fig. 11 of the main article. Panels a and b correspond to the data for the dark, and light colors, respectively.

Individual JNDs and categories

Individual Boundary Tests (fig. S12)

Figure S12. Individual boundary tests for dark and light colors. Format as in fig. 12 of the main article. Panels a and b correspond to the data for the dark, and light colors, respectively. Apart from green results were inconsistent across participants.
Other individual tests (fig. S13)

a.)

b.)

c.)

d.)

e.)

f.)
Figure S13. Individual categorical perception tests. Panels a-c illustrate the boundary tests for frequencies for isoluminant (panel a), dark (panel b), and light colours (panel c). The height of the colored bars represents the relative frequency of category JNDs above the boundary line. The minimum and maximum number of category JNDs in each category are given on top of the bars. The horizontal black line illustrates chance level (50%). Symbols at the base of the bars refer to p-values of binominal tests for difference from chance level. **Bars above the 0.50 line are in line with boundary effects.** Panels d-f illustrate the individual focal tests. The bars show the difference between category and prototype JNDs for isoluminant (panel d), dark (panel e), and light colors (panel f). Apart from that, format as in fig. 12 & S12. **Bars above zero indicate a prototype effect.** Panels g-i illustrate the individual triangle tests for isoluminant (panel g), dark (panel h), and light colors (panel i). The bars show correlations between the relative category JNDs and the location of colors between boundaries and prototypes. The height of bars (y-axis) represents the size of the correlation coefficients. Each bar corresponds to one side of the prototype for a particular category, and a particular participant. Saturated bars refer to the side with a lower-azimuth than the prototype, unsaturated bars represent the upper-azimuth side. Apart from that, format as in fig. 12. **Both kinds of bars (saturated and unsaturated) must be above zero to confirm a category effect.**
DISCUSSION

Categorical sensitivity and perceptual mechanisms (fig. S14)

Figure S14. Residuals for dark and light colors. The graphics compare the residuals of the ellipse fit to the color categories for dark (panel a) and light colors (panel b). Format as in fig. 14. Unlike the original JNDS, the residuals for dark colors do not show the categorical pattern for green, but categorical patterns for brown and blue. The residuals for light colors completely contradict any categorical pattern.

Generalization to other color spaces (fig. S15)

Figure S15. Tests in CIELUV and CIELAB. The graphics illustrate the categorical perception tests for the JNDS in CIELUV (panel a) and CIELAB (panel b). Format as in fig. 13, with the only difference that JNDS are calculated as Euclidean distances. Apart from yellow in CIELAB, results were similar to those in DKL-space.