Blue hues don't bring the blues: questioning conventional notions of color–emotion associations

Karen B. Schloss,[1,2,*] **Christoph Witzel,**[3] ⓘ **and Leslie Y. Lai**[4]

[1]*Department of Psychology, University of Wisconsin–Madison, 1202 W. Johnson St. Madison, Wisconsin 53706, USA*
[2]*Wisconsin Institute for Discovery, University of Wisconsin–Madison, 330 N. Orchard St. Madison, Wisconsin 53715, USA*
[3]*Department of Psychology, University of Southampton, England, UK*
[4]*Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, 190 Thayer St. Providence, Rhode Island 02912, USA*
Corresponding author: kschloss@wisc.edu

It is commonly held that yellow is happy and blue is sad, but the reason remains unclear. Part of the problem is that researchers tend to focus on understanding why yellow is happy and blue is sad, but this may be a misleading characterization of color–emotion associations. In this study, we disentangle the contribution of lightness, chroma, and hue in color–happy/sad associations by controlling for lightness and chroma either *statistically* or *colorimetrically*. We found that after controlling for lightness and chroma, colors with blue hue were no sadder than colors with yellow hue, and in some cases, colors with blue hue were actually happier. These results can help guide future efforts to understand the nature of color–emotion associations.   © 2020 Optical Society of America

1. INTRODUCTION

Which hue is happiest? And, which is saddest? You probably answered yellow for happy and blue for sad, and according to several studies on color–emotion associations, you would be correct [1–5]. Yellow–happy and blue–sad associations are also found in popular culture. The Disney film "Inside Out" portrays orbs storing happy memories as yellow and sad memories as blue. And, phrases like "I feel blue" are synonymous with feeling sad [6]. However, the origin of these associations is unknown, and researchers have been perplexed about why adults would associate blues with sadness [7], given that blues are typically among the most preferred colors [8–13].

Accounts of color–emotion associations [7,14–16] fall into four main categories: physiological responses (e.g., face flushing when angry), emotional responses (e.g., feeling aggressive when seeing red), language conventions (e.g., "feeling blue"), and experiences of colored objects in the world (e.g., associations between red and blood) [14]. The latter account, that color–emotion associations are based on experiences with color-related objects, is similar to the ecological valence theory of color preferences: i.e., color preferences are determined by preferences for all concepts associated with those colors [10,17].

To evaluate theories of color–emotion associations, it is important to have a precise description of the pattern of data the theories aim to explain. However, concerns have been raised

that reports on color–happy/sad associations are misleading [18]. Reported differences in yellow versus blue may be an epiphenomenon of variations in lightness and chroma of the experiment stimuli [18]. In the following sections, we lay out an account for why we predict that yellow hues may be no happier than blue hues when lightness and chroma are controlled. We then present evidence that supports this prediction. Our focus is on semantic associations between colors and emotion words, which are distinct from the effects of color on emotion [19–21].

A. Pertinent Issues in Color Perception and Cognition

Three main issues in color perception and cognition underlie the prediction that happy/sad associations are dominated by lightness and chroma rather than yellow/blue hue.

(1) **Perceptual color spaces are asymmetric:** The most saturated colors of yellowish hues are much lighter than the most saturated colors of bluish hues. There are no highly saturated yellowish hues that are dark or highly saturated bluish hues that are light (Fig. 1) [18,22–24].

(2) **Category prototypes differ in lightness:** The "best" examples [25], or prototypes [26], of chromatic color categories such as YELLOW and BLUE are maximally saturated [27–29]. Given that maximally saturated yellow hues are light and maximally saturated blue hues are dark (Fig. 1),

the most prototypical member of the category YELLOW is light, and the most prototypical member of the category BLUE is dark.

(3) **Color names typically refer to prototypes.** Without further specification, color names most likely refer to category prototypes. Given that chromatic category prototypes are highly saturated, and maximal saturation occurs at different lightness levels, chromatic color terms (e.g., "yellow," "blue," "red," and "green") refer to different lightness levels [27,28]. The term "yellow" typically refers to yellowish hues that are light and saturated, whereas the term "blue" typically refers to bluish hues that are dark and saturated. As colors of yellow hue darken, they cross a category boundary from YELLOW to BROWN [26–29].

Moreover, there is a risk of confusing the terminology used to discuss colors, including (1) color terms (e.g., "yellow" color name), (2) specifications of perceptual dimensions (e.g., yellowish hue, at any lightness or chroma level), and (3) relative position along a color opponent axis (e.g., more or less positive on the b^* yellow/blue axis in CIELAB space). Color names are also used to describe the poles of color opponent axes. Colors that are more positive on b^* are yellower and more negative on b^* are bluer, but that does not always mean that colors that are more negative on b^* appear to have a bluish hue. Using opponent axes is useful for revealing patterns of hue variation in data [10,11,30], but it is important to examine the data for individual colors and not just rely on the sign of the weight on the opponent axis to interpret the results.

B. Implications for Understanding Color–Emotion Associations

These issues in color perception and cognition have important implications for interpreting prior work on color–emotion associations. Due to the asymmetry in color space (Fig. 1), studies that tested highly typical (saturated) colors necessarily co-varied yellow/blue hue with lightness [1–3]. When those studies showed that yellowish colors were happier than bluish colors, it was unclear whether the differences were due to differences in hue or lightness [18]. This issue also relates to studies in which participants selected colors from the entirety of color space [31,32]. Yellowish colors may have been selected more often for joy because of the yellowish hue, or because out of all hues, yellows reach the highest lightness while maintaining high chroma. For studies in which color coordinates were unspecified [4,5,7],
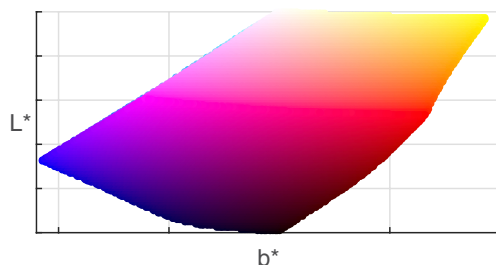
Fig. 1. Plane of L^* (lightness) and b^* (approximately yellow/blue) in CIELAB space. The most saturated yellow is lighter than the most saturated blue; a general property of color perception models (e.g., CIELUV, Munsell, NCS).

it is plausible that researchers chose prototypical colors for each color category, so yellows were likely lighter than blues.

In a study where lightness was held constant, color patches (and faces) were judged as happier when they were redder and yellower (i.e., more positive on a^* and b^*), and sadder when they were greener and bluer (i.e., less positive on a^* and b^*) [16]. However, the colors described as bluer and greener were still positive on a^* and b^* (i.e., appeared grayish rather than greenish or bluish hue).

For studies in which participants judged color names rather than color patches [15,33,34], it is unclear whether they judged colors they imagined when reading color names, concepts of colors, or some other cognitive representation. However, it seems plausible that participants evaluated category prototypes and, if so, their cognitive representation of the term "yellow" likely had greater lightness than their cognitive representation of "blue." Thus, it is unclear whether "yellow" is associated with happiness (or joy) and "blue" with sadness because of the hue or lightness attributed to the terms "yellow" and "blue."

Evidence suggests lightness and chroma are important dimensions in color–happy/sad associations [18,31,32,35], and studies reporting that blue was sad also reported that dark desaturated colors were also sad (e.g., blacks, browns, and grays) [2,3,5]. Most relevant, D'Andrade and Egan [18] tested happy/sad associations for colors that were controlled for chroma and value in Munsell space. They reported that yellowish colors were not happier than blues, and if anything bluish-greens might have been most happy. Unfortunately, they did not show their data or present corresponding statistical tests. Also relevant to the present study, D'Andrade and Egan [18] suggested that hue variations in happy and sad judgments may have been due to perceived variation in chroma, despite holding Munsell chroma constant. Munsell space is an approximate model of color appearance and cannot guarantee complete control of the perceptual dimensions, especially chroma [24].

C. Present Study

We tested whether yellow hues were happier than blue hues when lightness and chroma were controlled, either *statistically* (Experiment 1) or *colorimetrically* (Experiment 2). We contrast our results for happy/sad associations with angry associations, for which we predicted hue would still drive the associations (i.e., redness with anger) after controlling lightness and chroma.

2. EXPERIMENT 1

In this experiment, we controlled for lightness and chroma statistically, using the color–emotion association data from Palmer *et al.* [1]. Our hypothesis was that yellows were judged as happier than blues because of differences in lightness, rather than differences in hue. Palmer *et al.* [1] originally used these data to study color–music associations. Our results do not bear on their key finding that music–color associations are mediated by emotion associations. Our results just provide a deeper understanding about the nature of color–emotion associations.

A. Methods

The dataset included color–emotion association ratings for the Berkeley Color Project 37 (BCP-37) colors and the emotion terms happy, sad, and angry ($n = 48$ participants). We combined the happy/sad ratings into a single dimension given they were strongly negatively correlated [1]. The BCP-37 colors included eight hues [Red, Orange, Yellow, cHartreuse, Green, Cyan, Blue, and Purple] sampled at four different saturation/lightness levels (referred to as "saturated," "light," "muted," and "dark"), plus five achromatic colors (black, three shades of gray, and white) (Fig. 2A). The colors for each hue were sampled by first determining the color with the highest Munsell chroma that could be rendered within the monitor gamut (saturated set), and scaling the light, muted, and dark colors with respect to the saturated color for that hue (see Ref. [1]). Given that it is impossible to hold lightness constant while producing highly saturated, prototypical yellow and blue (Section 1.A), it was necessary to vary lightness across hues within each set to include highly saturated yellow and blue. For the present analyses, we converted from CIE xyY to CIELAB coordinates using a white point near CIE Illuminant C ($x = .312$, $y = .318$, $Y = 116$ cd/m^2; the chromaticity of the gray background with luminance of BCP white). The color–emotion association data and color coordinates are on github.com/SchlossVRL/coloremoblues.

B. Results and Discussion

Happy/sad. Figure 2B shows the mean happy/sad ratings. We first compared the most typical yellow and blue within the set (saturated-yellow and saturated-blue; marked with gray circles). Mean happy/sad ratings for saturated-yellow were significantly greater than for saturated-blue [$t(47) = 7.86$, $p < .001$, $d = 1.14$]. However, saturated-yellow was lighter and had higher chroma than saturated-blue, so we next tested whether this yellow versus blue difference would persist if we controlled lightness and chroma.

We controlled lightness and chroma statistically by (1) conducting a multiple linear regression (MLR) on the mean ratings using L^* and C^* from CIELAB space as predictors; (2) using the resulting linear equation to calculate the predicted ratings for each color based only on L^*, C^*, and the constant; and (3) subtracting those predictions from each participant's ratings to obtain the residuals. For this, and all subsequent MLR models, we verified that standard assumptions were met. This model accounted for 74% of the variance in happy/sad ratings ($F(2, 34) = 48.29$, $p < .001$, $HS_{\text{pred}} = -81.53 + 0.80L^* + 1.08C^*$). As shown in Fig. 2D, not only was saturated-yellow no longer happier than saturated-blue, but the mean residual for saturated-blue was actually happier than for saturated-yellow [$t(47) = 2.7$, $p = .011$, $d = .38$].

We next tested whether the hue opponent dimensions of yellow/blue (b^* in CIELAB) or red/green (a^* in CIELAB) accounted for additional variance in mean happy/sad ratings beyond L^* and C^* by including all four predictors in a forward-stepping MLR. Although four dimensions are more than needed to represent a 3D space, we included Cartesian (L^*, a^*, b^*) and polar (C^*) coordinates because these dimensions capture distinct variance in patterns of judgments across
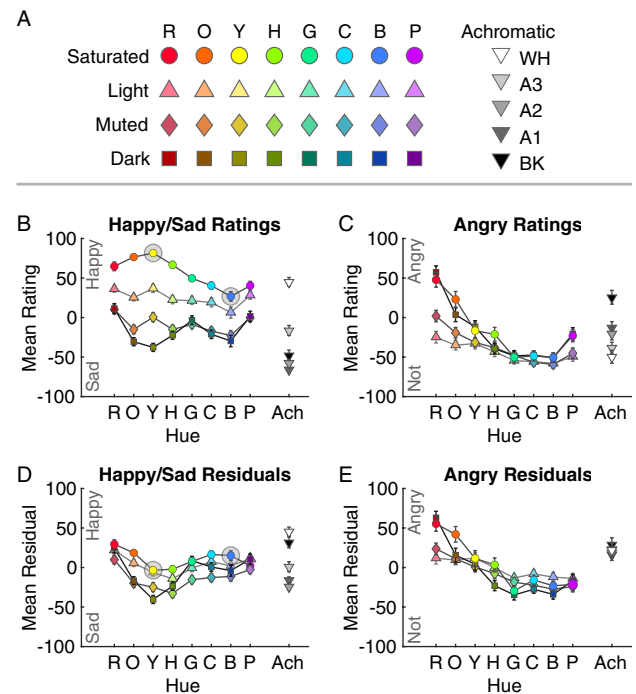
colors (see Ref. [30]). Only L^* and C^* explained significant variance, so the resulting model was the same as reported above. C^* was entered first, accounting for 55% variance with higher chroma being happier, and L^* accounting for an additional 19% variance, with lighter colors being happier. The finding that b^* was not entered into the model provides further evidence that yellower colors were not happier than bluer colors when lightness and chroma were controlled.

Angry. Figure 2C shows mean angry ratings, and Fig. 2E shows mean angry residuals after controlling L^* and C^*. The regression model with only L^* and C^* accounted for only 29% of the variance in angry ratings ($F(2, 34) = 6.98$, $p = .003$, $A_{\text{pred}} = -1.79 - 0.71L^* + 0.44C^*$). With comparably little variance explained by L^* and C^*, the pattern of data in the residuals is similar to the original ratings, with redder colors being most angry. When we included L^*, C^*, a^*, and b^* in a forward-stepping MLR, the model explained 79% of the variance with a^* (40% variance), b^* (+19% variance), and L^* (+20% variance) entered in that order ($F(3, 33) = 40.10$, $p < .001$, $A_{\text{pred}} = 6.86 + 0.61a^* + 0.54b^* - 0.68L^*$). Colors that were redder, yellower, and darker were rated as more angry.

The results of this experiment suggest that the canonical yellow/blue difference reported for happy/sad judgments [1–5] was due to differences in lightness and chroma rather than hue, but the canonical red–angry association [2,3,10,16,33,36] is due to redness of hue, not to lightness or chroma. By analogy, these results also suggest that happiness (or joy) may be more associated with the color name "yellow" than the color name "blue" because of lightness rather than hue. If it were possible to statistically control for lightness in the cognitive representations

Fig. 2. A, BCP-37 colors tested in Ref. [1] and the legend for B–E; B, mean happy/sad ratings from [1]; C, mean angry ratings from [1]; D, mean happy/sad residuals; E, mean angry residuals (see text for details). Error bars represent SEM. B and C are adapted from the Supplementary Materials in Ref. [1].

of "yellow" and "blue," we predict that the word "yellow" would no longer be more associated with happiness (or joy) than the word "blue," but it is unclear how that could be done.

3. EXPERIMENT 2

In Experiment 2, we tested for yellow/blue hue differences when controlling for lightness and chroma colorimetrically, by holding L^* and C^* constant. Doing so required using somewhat desaturated colors, given the asymmetry of perceptual color space (Fig. 1). Therefore, to ensure we could observe hue effects where we expected them, we collected additional data on angry/not-angry associations. We also assessed color appearance (red/green, yellow/blue, light/dark, saturated/desaturated), primarily to check that yellow hues appeared yellowish and blue hues appeared bluish even though the colors were desaturated.

A. Methods

1. Participants

We tested 20 participants (14 females, 6 males, mean age = 23.85 years, all with normal color vision (screened with the H.R.R. Pseudo-Isochromatic Plates [37]). All gave informed consent, and the Brown University IRB approved the protocol.

2. Design, Displays, and Procedure

The participants judged emotional associations (happy/sad and angry/not-angry) and color appearance (yellow/blue, red/green, light/dark, and saturated/desaturated) for 32 colors.

As shown in Fig. 3, the 32 colors included all combinations of 8 hues [Red, Orange, Yellow, cHartreuse, Green, Cyan, Blue, and Purple] × 2 lightness levels [light, dark] × 2 chroma levels [low, high]. The colors were defined relative to the light colors in the BCP-37 [1]. The eight hues had the hue angle (h in CIELCh; polar representation of CIELAB) of the BCP-37 light colors (R = 28.5, O = 64.3, Y = 96.3, H = 116.2, G = 154.7, C = 184.2, B = 245.3, P = 313.7). The luminance of the present light colors was the mean luminance of the BCP-37 light colors ($Y = 66.41$ cd/m^2). The Michelson luminance contrast between the light colors and the background was +55%, so we defined the dark colors to have a corresponding decrement in contrast from the background (−55%; $Y = 5.65$ cd/m^2). The mean chroma ($C^* = 38$) of the BCP-37 light colors could not be rendered for all hues within the monitor's gamut, so chroma was defined as $C^* = 30$ for the high chroma set and $C^* = 15$ for the low chroma set. The background was a medium gray ($L^* = 51$, $Y = 19.26$ cd/m^2) with chromaticity at CIE Illuminant C (CIE $x = .310$, $y = .316$). Conversions between CIE xyY and CIELAB spaces were calculated using CIE Illuminant C as the white point ($Y = 100$ cd/m^2). A Konica Minolta CS-200 Color Meter was used to characterize the monitor and manually check the color production accuracy. The maximal deviation between rendered colors and target colors was < .01 for x and y and <1 for Y in CIE xyY space. Files with color coordinates are posted at github.com/SchlossVRL/coloremoblues.
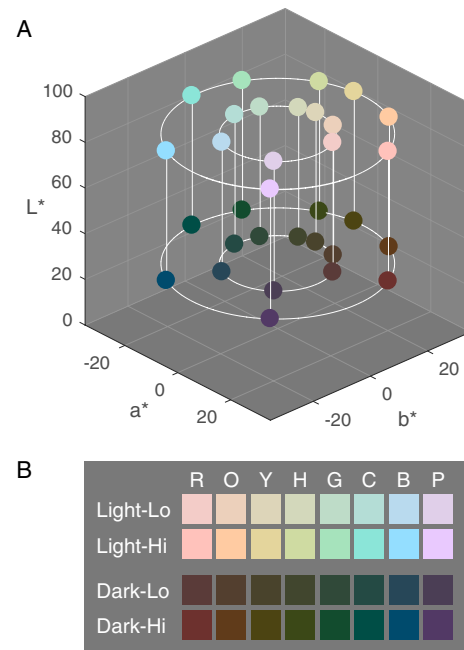

Fig. 3. Thirty-two colors tested in Experiment 2, shown A in CIELAB space and B as colored patches (see text for details).

The displays consisted of a colored square (100 pixels × 100 pixels; 2.6° × 2.6°) centered on the screen, with a 400-pixels-long (10.3°) response scale centered below the square. The middle of the scale was demarcated with a tick mark. The end points of the scale were (right/left) labeled according to the rating dimension: Happy/Sad, Angry/Not Angry for color–emotion associations, and Yellow/Blue, Red/Green, Light/Dark, and Saturated/Desaturated. Data were scaled to range from −100 (left end point) to +100 (right end point). We used a single scale for happy/sad ratings given that happy and sad ratings are strongly inversely correlated [1]. The software Presentation (https://www.neurobs.com) was used to generate and display test items during the experiment. The monitor was an Asus ProArt PA246Q monitor (1920 × 1200 pixels resolution, 51.9 cm wide × 32.5 cm tall), viewed from a distance of approximately 60 cm in a dark booth.

All participants first completed the color–emotion associations rating task, then the color appearance rating task in one testing session. At the start of the color–emotion associations task, participants were shown the full set of colors. They were told they would see each color one at a time, and their task was to rate the color on the dimensions of Happy versus Sad and Angry versus Not-Angry. They then completed an anchoring task so they knew what the end points of the rating scales meant for them in the context of these colors. While viewing the full set of colors (Fig. 3B), they pointed to the colors that were the happiest, saddest, most angry, and least angry. They were instructed to rate those colors near the end points of the scale and to use the full range of the scale.

During the experiment, the order of the emotion dimensions was randomized across participants. For each emotion dimension, each of the 32 colors was presented twice, in a blocked randomized design (all colors appeared once in a random order
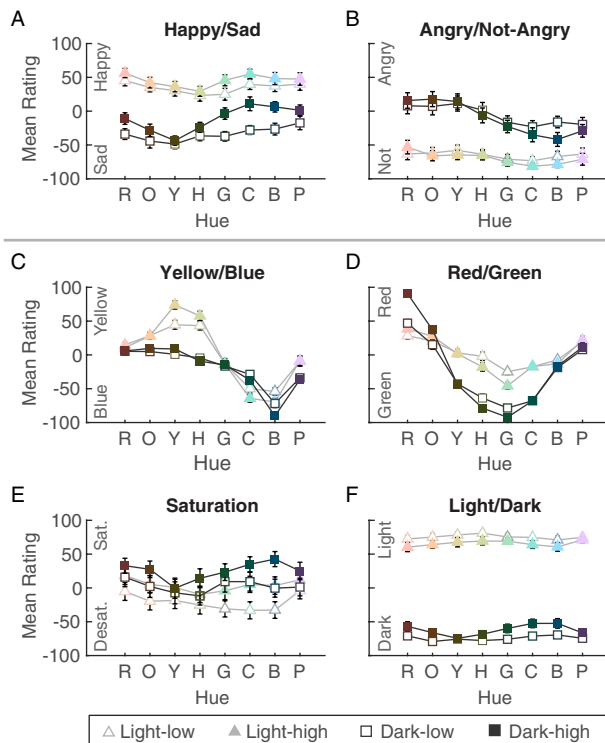
Fig. 4. Mean ratings for light colors (triangles) and dark colors (squares) of low chroma (open symbols) and high chroma (filled symbols): A, happy/sad; B, angry/not-angry; C, yellow/blue; D, red/green; E, saturated/desaturated; F, light/dark dimensions. Error bars represent standard errors of the means (SEM).

before any colors were repeated). To respond, participants used the mouse to slide the cursor along the response scale and clicked to record their response. The colors remained on the screen until a response was made, and trials were separated by a 500 ms inter-trial-interval displaying only the gray background.

The color appearance task was the same, except participants rated the colors on four color appearance dimensions. For anchoring in the instructions, all dimensions were described by their end points (i.e., "Red versus Green," "Yellow versus Blue," "Light versus Dark") except for saturation, where we added further explanation "Desaturated (close to gray) versus Saturated (very vivid)."

B. Results and Discussion

Figure 4 shows the mean color–emotion association and color appearance ratings, averaged over repetitions and participants.

1. Color–Emotion Associations

Happy/sad. To describe the pattern of mean happy/sad ratings (Fig. 4A), we conducted a forward-stepping MLR using L^*, a^*, b^*, and C^* as predictors, as in Experiment 1. The model accounted for 94% of the variance with L^* (82% variance), C^* (+6% variance), and b^* (+6% variance) entered in that order (a^* was not entered into the model) [$F(3, 28) = 133.91$, $p < .001$, $HS_{pred} = -79.03 + 1.10L^* + 1.23C^* - 0.48b^*$]. Colors were rated as happier when they were lighter, higher chroma, and bluer. The relative weights and variance

explained for L^* and C^* suggest L^* was most prominent followed by C^*, which is the opposite order from Experiment 1. This difference may be because the BCP-37 colors varied more in chroma than colors in the present set. Caution should be used when interpreting the relative strength of predictors for a given stimulus set because that may depend on variation of colors within the set.

We next conducted a linear mixed-effect regression model (LMER), which enabled us to test for systematic effects while accounting for individual subject variability (software R 3.4.1, lme4 1.1-13). We included fixed effects for L^*, C^*, a^*, and b^*; all interactions within L^*, C^*, and a^*; all interactions within L^*, C^*, and b^*; and by-subject random intercepts and by-subject random slopes for each fixed effect. We did not include interactions involving a^* and b^* because our primary focus was on how effects of each color-opponent dimension might interact with lightness and chroma. We z-scored the data and predictors to center them and represent them on the same scale. We set the covariances to zero because correlations between predictors were near zero [38].

The model (Table 1) showed significant positive weights on L^* and C^*, and negative weight on b^*, indicating that lighter, higher chroma, and bluer colors were rated as happier. An $L^* \times b^*$ interaction indicated there was a larger yellow/blue difference for dark colors than for light colors (Fig. 4A). An $L^* \times C^*$ interaction indicated that the degree to which higher chroma colors were rated as happier was greater for dark colors than light colors. A $C^* \times a^*$ interaction indicated that the degree to which greener colors were rated as happier was greater for high chroma than for low chroma colors.

The negative weight on b^* for both regression analyses suggests that bluer hues were rated as happier than yellower hues. We next specifically compared blue and yellow hues using a three-way repeated measures ANOVA [2 hues (B, Y) $\times$ 2 lightness (light, dark) $\times$ 2 chroma (high, low)]. A main effect of hue indicated that blue hues were happier [$F(1, 19) = 8.54$, $p = .009$, $\eta_p^2 = .310$], but hue interacted with lightness [$F(1, 19) = 6.48$, $p = .020$, $\eta_p^2 = .254$] and chroma [$F(1, 19) = 7.52$, $p = .013$, $\eta_p^2 = .284$], and there was a three-way interaction [$F(1, 19) = 8.83$, $p = .008$, $\eta_p^2 = .317$].

To understand these interactions, we did follow-up two-way ANOVAs within each lightness level. For light colors, there was no effect of hue or hue $\times$ chroma interaction ($Fs < 1$). For dark colors, a main effect of hue indicated that blue was happier than yellow [$F(1, 19) = 16.59$, $p = .001$, $\eta_p^2 = .466$], and a hue $\times$ chroma interaction indicated this difference was larger for high chroma colors [$F(1, 19) = 11.24$, $p = .003$, $\eta_p^2 = .372$]. Thus, when colors were light, we found no significant difference in yellow versus blue hues, and when colors were dark, blue hues were happier than yellow hues. When yellow hues darken, they cross a category boundary from YELLOW to BROWN, so this result aligns with reports that brown is associated with sadness [3,5,18,39].

Angry/not-angry ratings. We conducted the same forward-stepping MLR and LMER analyses for the angry/not-angry ratings (Fig. 4B). The forward-stepping MLR model accounted for 95% of the variance with L^* (81% variance), b^* (+8% variance), a^* (+5% variance), and C^* (+1% variance) entered in that order [$F(4, 27) = 121.90$, $p < .001$,

Table 1.   LMER Models for Happy/Sad, Angry/Not-Angry, Saturation, and Lightness Ratings[a,b]

	Pred.	β	Df	t	p
Happy/Sad	Int	0.00	19.1	−0.1	0.947
	L	0.63	19.0	7.5	***
	C	0.19	19.2	7.0	***
	a	0.03	31.1	0.9	0.398
	b	−0.17	22.5	−3.3	0.003*
	$L:C$	0.07	19.2	−2.5	0.036*
	$L:a$	0.05	31.5	1.6	0.121
	$C:a$	−0.05	473.3	−2.1	0.032*
	$L:b$	0.08	30.4	2.4	0.020*
	$C:b$	0.02	473.3	0.8	0.427
	$L:C:a$	0.00	473.3	−0.1	0.931
	$L:C:b$	0.00	473.3	0.1	0.946
Angry/Not	Int	0.00	19.0	0.0	0.998
	L	−0.57	19.0	−6.8	***
	C	−0.06	19.2	−2.9	0.009**
	a	0.14	27.3	4.7	***
	b	0.18	19.4	3.7	0.001**
	$L:C$	0.01	462.6	0.8	0.402
	$L:a$	−0.04	24.5	−1.2	0.226
	$C:a$	0.00	462.6	−0.1	0.916
	$L:b$	−0.12	24.3	−3.2	0.004**
	$C:b$	0.00	19.5	−0.1	0.899
	$L:C:a$	0.00	462.6	0.0	0.975
	$L:C:b$	0.01	462.6	0.5	0.605
Saturation	Int	0.00	19.1	0.0	1.000
	L	−0.19	19.0	−1.1	0.301
	C	0.02	19.1	6.0	***
	a	0.09	21.2	1.9	0.076
	b	−0.05	22.3	−1.1	0.272
	$L:C$	0.01	19.1	0.3	0.776
	$L:a$	0.08	37.6	3.5	0.001**
	$C:a$	−0.05	474.0	−2.8	0.006**
	$L:b$	0.03	28.9	1.2	0.256
	$C:b$	−0.03	474.0	−1.4	0.168
	$L:C:a$	−0.04	474.0	−2.0	0.048*
	$L:C:b$	0.00	474.0	0.2	0.854
Lightness	Int	0.00	19.1	0.0	0.971
	L	0.95	19.0	21.0	***
	C	0.01	19.2	0.7	0.513
	a	−0.01	31.4	−0.9	0.354
	b	−0.01	32.1	−0.9	0.350
	$L:C$	−0.08	19.2	−5.2	***
	$L:a$	0.00	37.7	0.1	0.901
	$C:a$	0.00	475.6	−0.2	0.805
	$L:b$	0.04	31.0	2.8	0.008**
	$C:b$	−0.01	475.6	−0.9	0.372
	$L:C:a$	0.01	475.6	0.7	0.479
	$L:C:b$	−0.01	475.6	−1.0	0.336

[a]Predictors included the intercept (Int), L^*, C^*, a^*, and b^* and their interactions ("*" notation was excluded to avoid confusion with the interaction symbol ":").

[b]*$p < .05$, **$p < .01$, ***$p < .001$.

$A_{pred} = 28.17 − 1.04L^* + 0.57b^* + 0.46a^* − 0.40C^*$]. Colors were rated as more angry when they were darker, yellower, redder, and lower chroma. The LMER model showed similar results, with significant negative weights on L^*, C^*,

and positive weights on a^* and b^* (Table 1). The only interaction was between L^* and b^*, which indicated that the blueness/yellowness difference in angry associations was stronger when colors were darker.

2. Color Appearance

Yellow/blue and red/green ratings. Our main purpose for collecting color appearance data was to check that the yellowish hues were judged as yellow and the bluish hues were judged as blue, given that the colors were desaturated. Yellow/blue ratings were strongly correlated with the corresponding dimension b^* ($r(30) = .78$, $p < .001$). As shown in Fig. 4C, the yellowish hues were judged as yellow as long as they were light, and the bluish hues were judged as blue regardless of the lightness level. This pattern reflects the color categories of the end points of the yellow/blue dimension (e.g., Fig. 4 in Ref. [28]; Fig. 1 in Ref. [40]); light yellowish hues were judged as yellow, but dark yellowish hues were neither judged as yellow nor blue because they are typically referred to by the term "brown".

Similarly, red/green ratings were correlated with the corresponding dimension a^* ($r(30) = .81$, $p < .001$). As shown in Fig. 4D, only the dark red, high chroma, and dark green colors were strongly rated as red and green, respectively, because membership of the RED category depends on lightness and chroma (i.e., the term for dark reddish colors is "red," but the term for light reddish colors is "pink") and membership of the GREEN category depends on lightness (i.e., dark greenish colors are more typical for the term "green").

Saturation and light/dark ratings. As shown in Figs. 4E and 4F, saturation and lightness ratings did not only vary with chroma and lightness, respectively, but they appeared to also vary with hue. To understand these patterns, we conducted the forward-stepping MLR and LMER analyses used for the color–emotion association data.

Saturation ratings. The forward-stepping MLR on mean saturation ratings explained 80% of the variance with C^* (39% variance), L^* (+33% variance), b^* (+4% variance), and a^* (+4% variance) entered in that order [$F(4, 27) = 26.38$, $p < .001$, $S_{pred} = −10.96 + 1.68C^* − 0.39L^* − 0.23b^* + 0.24a^*$]. Colors were rated as more saturated when they were higher chroma, darker, bluer, and redder. The LMER model showed an effect of C^*, but also two-way interactions between C^* and a^*, and between L^* and a^*, and a three-way interaction between L^*, C^*, and a^* (Table 1). That is, variations in saturation ratings across hue differ between low chroma and high chroma and between dark and light colors. As seen in Fig. 4E, dark colors with high chroma peak around greenish and bluish hues, with a trough at yellow hue (corresponding to brown), whereas light desaturated colors peak around purple with a trough around green and blue.

These differences in saturation ratings for colors of equal chroma (C^*) resonate with D'Andrade and Egan's [18] speculation that blue–green hues may appear more saturated than other hues, even when the colors were equated in Munsell chroma. Our results are also consistent with previous reports that subjective estimations of saturation vary across hue when controlling for saturation through measures of chroma and saturation, such as C^* in CIELAB space [41,42].

Lightness ratings. The lightness ratings in Fig. 4F were largely determined by L^*. A forward-stepping MLR accounted for 99% of the variance in lightness ratings using only L^*, with no other predictors entered [$F(1, 30) = 2750.64$, $p < .001$, $L_{pred} = -138.00 + 2.45L^*$]. However, Fig. 4F shows that lightness ratings also slightly varied depending on chroma and hue. The LMER model showed an effect of L^*, plus two-way $L^* \times C^*$ and $L^* \times b^*$ interactions (see Table 1). The $L^* \times C^*$ is due to more extreme differences between light and dark colors for low chroma than for high chroma colors (Fig. 4F). The $L^* \times b^*$ interaction is due to a peak around green–blue hues and a trough around yellow hue for dark colors but not for light colors. Moreover, lightness ratings were correlated with saturation ratings across the 32 colors ($r(30) = -.56$, $p < .001$). These results align with previous reports that lightness estimations vary across hue for colors of equal L^* [43], and that perceived lightness depends on saturation (Helmholtz–Kohlrausch effect for spectral colors [44]).

3. Effects of Color Appearance on Happy/Sad Ratings?

Based on the results reported above, it is possible that differences in happy/sad ratings across hues within a given C^* and L^* level were due to perceived differences in saturation and lightness, rather than hue. To test this possibility, we included the mean color appearance ratings [lightness (LD), saturation (Sat), yellow/blue (YB) and red/green (RG)] as predictors in a forward-stepping MLR along with L^*, C^*, a^*, and b^*. The logic was that hue variations in lightness and saturation ratings might reduce the role of blueness/yellowness differences (captured by either b^* or YB ratings) in the model. This model accounted for 97% of the variance with lightness ratings (85% variance), saturation ratings (+9% variance), and yellow/blue ratings (+3% variance) entered in that order [$F(3, 28) = 256.93$, $p < .001$; $HS_{pred} = 4.86 + 0.57LD + 0.60 Sat - .16YB^*$]. Colors were rated as happier when they were lighter, more saturated, and bluer. The findings suggest that subjective ratings of lightness, saturation, and yellow/blue ratings capture happy/sad judgments better than L^*, C^*, and b^* given that (1) the model entered subjective ratings instead of L^*, C^*, and b^*, and that (2) YB accounted for somewhat less variance here (3%) than b^* accounted for in the model with L^*, C^*, and b^* reported above (6%).

In summary, the results of this experiment showed that happy/sad judgments were largely determined by perceived lightness and saturation. When colors were dark with high chroma, bluer hues were rated as happier than yellower hues, which may have been due to dark bluish hues appearing more saturated and lighter than dark yellowish hues, even though C^* and L^* were constant. In the Appendix A of this paper, we report on a supplementary experiment that replicated the key findings of Experiment 2, and explored perceptually matching colors for saturation and assessing effects on happy/sad judgments.

4. CONCLUSION

This study showed that when we controlled lightness and chroma statistically (Experiment 1) and colorimetrically

(Experiment 2), yellow hues were no happier than blue hues, and in some cases, blue hues were actually happier. Our results suggest that previous evidence that yellow is happier than blue [1–5] is likely due to variations in lightness between colored stimuli rather than variations in hue. We contrast these findings with associations between redness and anger, which appear really to be due to redness of hue.

Still, yellow colors have the greatest potential to be happiest, not because of their hue, but because they can simultaneously maximize lightness and chroma (Fig. 1). As shown by Jonauskaite *et al.*, when participants had free range of color space, they matched light, high chroma yellows to joy, and dark, desaturated grays to sadness.

Although yellow hues have the greatest potential to be associated with happiness, like any other hue, they become associated with sadness when they are dark and desaturated [31,39]. This point is not only relevant for color cognition, but also for applying color in design (e.g., graphic or interior design). We hope that our approach of isolating the key dimensions underlying happy/sad judgments will help constrain and advance theories about the origins of color–emotion associations.

APPENDIX A

We conducted a supplementary experiment with two aims. First, we aimed to test whether the variation across hue in happy/sad ratings would diminish if we perceptually matched colors of different hues for perceived saturation. We focused on matching saturation rather than lightness because saturation ratings (Fig. 4E) varied more strongly across hue than lightness ratings did (Fig. 4F). Second, we aimed to replicate the Experiment 2 results.

We tested 20 new participants in the color matching task (15 females, 5 males, mean age = 23.4) and 43 other new participants in the color–emotion association task (31 females, 12 males, mean age = 20.1). We piloted our matching procedure on 18 other participants while resolving color specification details. All had normal color vision (tested using [37]) and gave informed consent. The Brown University IRB approved the protocol.

A1. Matching Colors for Perceived Saturation

We used a saturation matching task [45] to produce colors with equal subjective saturation, which we subsequently used to assess color–emotion associations. In the saturation matching task, participants saw pairs of colors that had two different hues but the same luminance. They adjusted the chroma of one color (the "comparison" color on the right), to match the chroma of another color (the "test" color on the left). When luminance is equal, adjusting chroma corresponds to adjusting saturation. A benefit of this approach is that observers do not need to understand the concepts of chroma and saturation to do the task—it can be accomplished by making the colors as similar as possible in color appearance.

Trials included all pairwise combinations of the eight hues ($8 \times 8 = 64$) within each of the two lightness levels (L^*) from Experiment 2 (128 trials). The colored squares were presented side by side (each 2.2 cm, 2°). Each pair was presented twice

(left/right balanced) so each color served as the test or the comparison. We only tested one chroma level ($C^* = 20$) because gamut limitations made matching impossible for the high chroma colors ($C^* = 30$) of Experiment 2. The initial chroma of the comparison for each trial was $C^* = 20$, the same as the test color [45]. The upper bound of adjustment was fixed to $C^* = 35$ to stay within the monitor gamut. We used equal bounds for all hues to avoid biases due to gamut shape. The white point was set to equal energy white [CIE1931 $x = 0.33$, $y = 0.33$, $Y = 80$ cd/m^2]. The background was a medium gray (CIE1931 $x = 0.33$, $y = 0.33$, $Y = 15.4$ cd/m^2), which corresponded to the same lightness ($L^* = 51$) as in Experiment 2 given the white point.

Participants were instructed to use the number pad to adjust the comparison color so that it was most similar to the test color (1 = increase, 2 = decrease, 9 = done). They were told that they could only adjust "saturation," which is "how colorful (vivid) a color appears." Before the experiment, participants completed practice trials. Measurements for dark and light colors were done in separate blocks, randomized over participants. After this task, participants also completed a saturation ranking task, which will be the subject of a separate study.

Participants viewed the stimuli through a black tunnel to control color adaptation, presented on a 20 inch (50.8 cm) Sony GDM-20SE2T monitor (8 bits per channel, 1024 × 768 pixels). We used a CRT monitor in these measurements to help ensure calibration was reliable across continuous variations in color. We used a slightly different white point to allow for reproducing the same CIELAB colors within the gamut of this monitor as on the monitor used in the main experiments. The CIE1931 xyY coordinates of the monitor primaries were R = (0.6641, 0.3157, 24.62), G = (0.2383, 0.6787, 67.28), and B = (0.1446, 0.0722, 9.65). Gamma corrections without bit loss were applied based on the measured gamma curves of the monitor primaries. Experiments were written in MATLAB (The MathWorks Inc., 2007) with the Psychophysics toolbox Version 3 extensions [46].

For each test hue at a given lightness level, we calculated mean match across all comparison hues. A repeated measures ANOVA for eight hues within each lightness level showed that the average matches differed significantly across hues [light: $F(7, 133) = 8.3$, $p < 0.001$, $\eta_p^2 = .0304$; dark: $F(7, 133) = 7.7$, $p < 0.001$, $\eta_p^2 = 0.288$]. Figure 5A shows the C^* values for each hue, which should produce colors appearing equal in saturation within each lightness level. We obtained these values for each comparison hue within a lightness level by calculating the mean C^* for that hue across the C^* values obtained from matches with each of the eight test hues (including itself). At this level of chroma ($C^* = 20$), discrimination thresholds for differences in chroma are roughly about 3 to 8 CIELAB units (cf. detection thresholds and Weber fractions in Figs. 8 and 9 of Ref. [40]). Thus, the matches did not differ strongly across test hues, varying by ∼1 discrimination threshold.
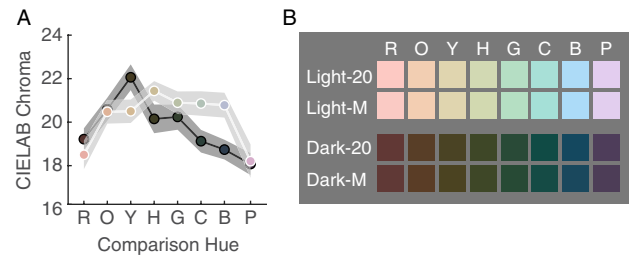
Fig. 5. A, mean chroma (C^*) matches showing equal subjective saturation. Error regions represent SEM. B, matched colors ("M") and unmatched colors in which $C^* = 20$ ("20").

A2. Rating Color–Emotion Associations and Color Appearance

Participants rated color–emotion associations and color appearance for the 16 saturation matched colors (2 lightness levels × 8 hues; matched around $C^* = 20$), 16 unmatched colors (2 lightness levels × 8 hues; $C^* = 20$), and the 32 colors from Experiment 2. We refer to data on the matched and $C^* = 20$ colors as the "matched dataset" and data on the Experiment 2 colors as the "replication" dataset. The colors referred to as "light" and "dark" had the same lightness levels as Experiment 2.

The methods were the same as in Experiment 2, except (1) we included the additional set of colors, (2) we counterbalanced task order [color–emotion task first ($n = 20$), or color appearance task first ($n = 23$)], which doubled our sample size, (3) participants only completed one block per color rather than two, (4) we provided further description of saturation "saturated colors are very colorful and vivid, whereas desaturated colors are very muted and grayish". The monitor and calibration procedure were the same as in Experiment 2. Files with color coordinates and data are on github.com/SchlossVRL/coloremoblues.

A2.1 Matched Dataset

Color–emotion association ratings. As shown in Fig. 6, saturation matching had little effect on happy/sad or angry/not-angry ratings. We conducted similar LMER models as in Experiment 2, but we coded the "chroma" factor (matched versus unmatched) as −1 versus 1, and used the a^* and b^* values for the unmatched colors for both stimulus sets. This enabled us to test whether the fit of a^* and b^* interacted with the chroma factor. The chroma factor did not have an effect, and it did not interact with any other factor for happy/sad or angry/not-angry ratings (Table 2). For happy/sad ratings, there were positive weights on L^* and a^* and a negative weight on b^*, indicating colors were rated as happier when they were lighter, redder, and bluer. For angry/not-angry ratings, there was a negative weight on L^* and positive weights on b^* and a^*, indicating that colors were rated as angrier when they were darker, yellower, and redder.

We directly tested for a yellow versus blue hue difference using a repeated measures ANOVA [2 hues (B, Y) × 2L^* (light, dark) × 2C^* (matched, unmatched)]. Hue had no effect ($F < 1$), and although it did interact with lightness
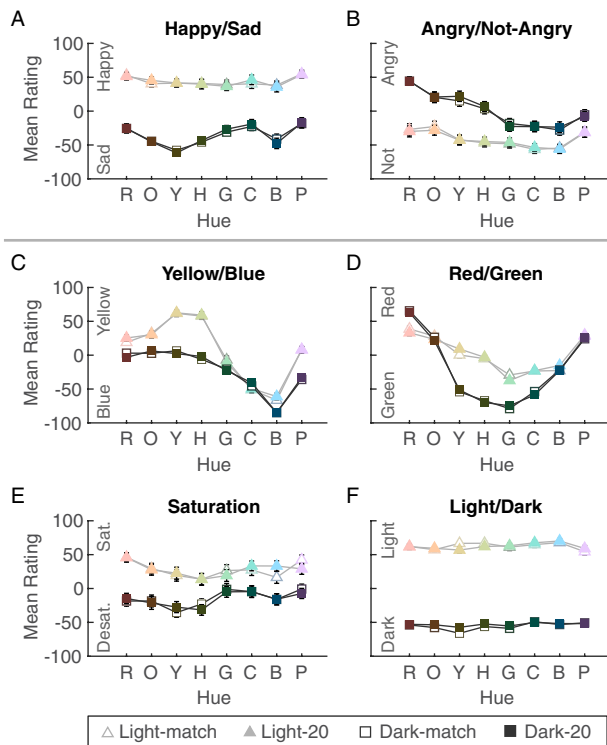
Fig. 6. Mean ratings for matched dataset in the supplementary experiment. The data are plotted in the same manner as in Fig. 4 except open symbols represent colors matched for saturation and filled symbol represent $C^* = 20$.

Table 2. LMER Models for Happy/Sad, Angry/Not-Angry, Saturation, and Lightness Ratings in the Matched Dataset[a]

		β	df	t	p
Happy/Sad	Int	0.00	42.0	0.0	1.000
	L	0.70	42.0	10.5	***
	C	0.00	1112.0	0.1	0.932
	a	0.05	42.0	2.3	0.026*
	b	−0.07	42.0	−2.4	0.023*
	$L:C$	0.00	1112.0	0.1	0.953
	$L:a$	0.03	42.0	1.8	0.085
	$C:a$	−0.01	1112.0	−0.5	0.636
	$L:b$	0.05	42.0	2.4	0.022*
	$C:b$	0.01	1112.0	0.5	0.600
	$L:C:a$	0.00	1112.0	0.2	0.831
	$L:C:b$	0.00	1112.0	0.0	0.993
Angry	Int	0.00	42.0	0.0	1.000
	L	−0.40	42.0	−4.9	***
	C	0.00	1112.0	0.1	0.910
	a	0.25	42.0	7.9	***
	b	0.16	42.0	5.5	***
	$L:C$	−0.01	1112.0	−0.5	0.642
	$L:a$	−0.07	42.0	−3.4	0.002**
	$C:a$	−0.01	1112.0	−0.7	0.499
	$L:b$	−0.10	42.0	−4.7	***
	$C:b$	0.00	1112.0	0.2	0.857
	$L:C:a$	−0.01	1112.0	−0.6	0.545
	$L:C:b$	0.00	1112.0	−0.3	0.746
Saturation	Int	0.00	42.0	0.0	1.000
	L	0.38	42.0	3.4	0.001**
	C	0.00	1030.0	−0.3	0.794
	a	0.04	42.0	2.0	0.056
	b	−0.08	42.0	−4.2	***
	$L:C$	0.01	1030.0	0.5	0.635
	$L:a$	0.07	42.0	3.9	***
	$C:a$	−0.01	1030.0	−0.8	0.441
	$L:b$	0.03	42.0	1.6	0.118
	$C:b$	0.00	1030.0	−0.2	0.851
	$L:C:a$	−0.02	42.0	−1.0	0.300
	$L:C:b$	−0.01	42.0	−0.6	0.541
Lightness	Int	0.00	42.0	0.0	1.000
	L	0.91	42.0	23.8	***
	C	0.01	1029.0	1.2	0.235
	a	−0.02	42.0	−2.3	0.024*
	b	−0.03	42.0	−2.2	0.030*
	$L:C$	−0.01	42.0	−1.2	0.229
	$L:a$	−0.02	42.0	−2.0	0.049*
	$C:a$	0.00	1029.0	0.0	0.991
	$L:b$	0.01	42.0	1.2	0.241
	$C:b$	0.00	1029.0	−0.4	0.667
	$L:C:a$	0.00	42.0	0.2	0.810
	$L:C:b$	−0.02	1029.0	−2.6	0.008*

[a]In this model, C coded matched versus unmatched colors.

$[F(1, 42) = 5.30, p = .026, \eta_p^2 = .112]$, tests within each lightness level showed no hue difference for light colors ($F < 1$) or dark colors $[F(1, 42) = 3.17, p = .082, \eta_p^2 = .070)$. Chroma did not interact with any of the other factors ($Fs < 1$).

These results show that the specific yellow hue versus blue hue difference in happy/sad ratings that we aimed to cancel out with our matching procedure did not occur for the matched or unmatched colors in this dataset. Part of the problem is that it was necessary to use low chroma colors in order to implement saturation matching within the monitor gamut. Yet, Experiment 2 showed that the finding of blue being happier than yellow was predominantly for (dark) *high* chroma colors.

The lack of effect of saturation matching may be explained by the small difference in C^* between the matched and unmatched colors (Fig. 6). Thus, we next asked whether saturation matching had any effect on saturation ratings.
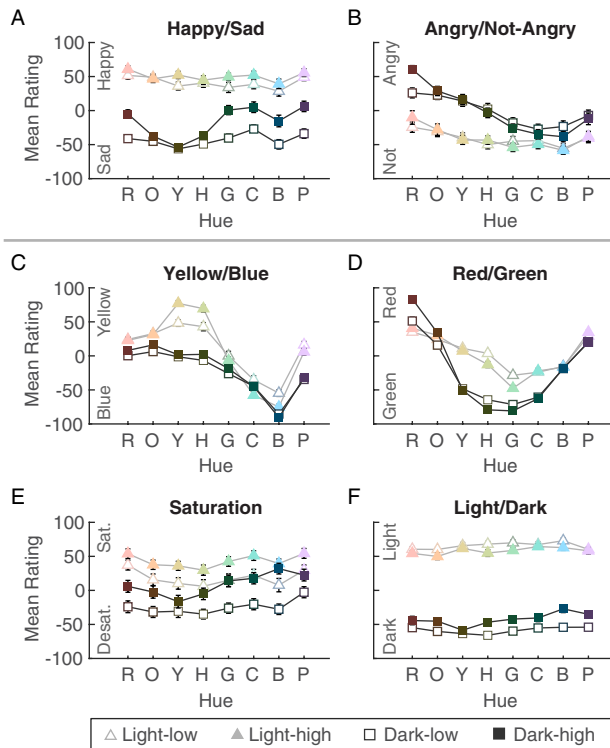
Saturation ratings. The saturation ratings in Fig. 6E suggest participants still reported differences in saturation across hues for matched colors. We conducted the same LMER model to predict saturation ratings as the happy/sad ratings, and found a significant negative weight on b^* and an interaction between a^* and L^* (Table 2). These weights indicated that colors were rated as more saturated if they were bluer overall, and if they were greener, especially in the dark set (Fig. 6). With no effect of matched versus unmatched colors and no interactions between that factor with any others (factor labeled C in Table 2), we conclude that saturation matches did not control perceived variation in saturation over hues, as measured by the saturation ratings. Saturation matches underestimated the magnitude of

variation across hue, which may arise if response noise in the matches is equally distributed across hue and hence decreases the variation in C^* of the matches across hue. However, there is no reference standard for the measurement of subjective

Fig. 7.   Mean ratings in the replication dataset (data plotted in the same manner as in Fig. 4).

saturation, and finding such a reference standard is a separate research question.

A2.2. Replication Dataset: Color–Emotion Associations

Happy/sad. The mean happy/sad ratings (Fig. 7A) were strongly correlated with the corresponding ratings from Experiment 2 (Fig. 4A) ($r(30) = .98$, $p < .001$). A forward-stepping MLR model with L^*, C^*, a^*, and b^* accounted for 90% of the variance with L^* (85% variance) and C^* (+5% variance) entered in that order ($F(2, 29) = 124.59$, $p < .001$, $HS_{pred} = -94.15 + 1.33L^* + 1.17C^*$). Like Experiment 2, mean responses were dominated by L^* and C^*, but unlike Experiment 2, b^* was not entered. However, when we included mean color appearance ratings and these colorimetric predictors, the model was similar to Experiment 2: lightness and saturation ratings replaced L^* and C^*, and yellow/blue ratings were also included. This model accounted for 96% of the variance with lightness ratings (89% variance), saturation ratings (+6% variance), and yellow/blue ratings (+1%) entered in that order ($F(3, 28) = 216.32$, $p < .001$; $HS_{pred} = -1.30 + .43LD + .65 Sat + .09 YB$).

As in Experiment 2, the LMER model for happy/sad showed effects of L^* and C^*, and interactions of $L^* \times C^*$ and $L^* \times b^*$ (Table 3). The effect of b^* and $C^* \times a^*$ interaction reported in Experiment 2 were in the same direction here, but did not reach significance. There were also new interactions of $L^* \times a^*$, $C^* \times b^*$, and $L^* \times C^* \times a^*$. These interactions reflect a pronounced peak around greens and blues for the dark, high chroma colors, which is evident here (Fig. 7A) and

Table 3.   LMER Models for Happy/Sad, Angry/Not-Angry, Lightness, and Saturation in the Replication Dataset

		β	df	t	p
Happy/Sad	Int	0.00	42.3	0.1	0.954
	L	0.66	42.1	11.2	***
	C	0.16	42.3	6.8	***
	a	0.05	57.1	1.9	0.060
	b	−0.06	48.9	−1.7	0.095
	$L:C$	−0.08	42.6	−3.9	***
	$L:a$	0.06	995.4	3.9	***
	$C:a$	−0.03	57.2	−1.8	0.078
	$L:b$	0.09	62.1	4.2	***
	$C:b$	−0.03	995.4	−2.0	0.047*
	$L:C:a$	−0.03	995.4	−2.0	0.043*
	$L:C:b$	0.02	61.9	0.9	0.392
Angry	Int	0.00	42.2	0.1	0.956
	L	−0.36	42.1	−4.8	***
	C	0.00	42.5	−0.2	0.832
	a	0.28	56.8	8.9	***
	b	0.20	56.8	6.5	***
	$L:C$	0.01	42.9	0.4	0.664
	$L:a$	−0.10	72.3	−4.1	***
	$C:a$	−0.02	991.0	−1.1	0.285
	$L:b$	−0.11	72.2	−4.9	***
	$C:b$	−0.03	991.0	−1.7	0.083
	$L:C:a$	0.01	991.0	0.6	0.574
	$L:C:b$	0.02	71.5	0.8	0.437
Saturation	Int	0.00	42.5	−0.1	0.927
	L	0.33	42.0	3.1	0.003**
	C	0.26	42.2	6.7	***
	a	0.07	86.4	3.6	***
	b	−0.12	76.3	−5.9	***
	$L:C$	−0.04	42.6	−2.1	0.040*
	$L:a$	0.04	81.3	1.9	0.056
	$C:a$	−0.06	1000.0	−3.5	***
	$L:b$	0.06	64.2	2.9	0.004**
	$C:b$	0.00	1000.0	−0.2	0.835
	$L:C:a$	0.00	1000.0	−0.2	0.839
	$L:C:b$	0.00	63.7	0.3	0.787
Lightness	Int	0.00	42.3	0.0	0.977
	L	0.89	42.1	23.7	***
	C	0.04	42.6	2.6	0.013*
	a	−0.03	97.0	−2.3	0.025*
	b	−0.06	75.9	−4.5	***
	$L:C$	−0.10	42.7	−7.6	***
	$L:a$	−0.04	100.3	−3.4	0.001**
	$C:a$	0.01	1021.0	0.6	0.520
	$L:b$	0.03	76.2	2.4	0.017*
	$C:b$	0.00	1021.0	−0.2	0.828
	$L:C:a$	0.02	1021.0	2.1	0.037*
	$L:C:b$	0.00	1021.0	0.1	0.922

in Experiment 2 (Fig. 4A), but these higher-order interactions were more likely to reach significance here because of the larger sample size.

The ANOVA directly comparing yellow versus blue hues [2 hues (B, Y) × 2 lightness (light, dark) × 2 chroma (high, low)] showed no main effect of hue [$F(1, 42) = 1.01$,

$p = .321$, $\eta_p^2 = .023$], but hue interacted with lightness [$F(1, 42) = 25.72$, $p < .001$, $\eta_p^2 = .380$] and marginally interacted with chroma [$F(1, 42) = 3.73$, $p = .060$, $\eta_p^2 = .081$] and there was a three-way interaction [$F(1, 42) = 10.45$, $p = .002$, $\eta_p^2 = .199$]. As in Experiment 2, two-way ANOVAs within each lightness level showed that for light colors, there was no effect of hue [$F(1, 42) = 3.52$, $p = .067$, $\eta_p^2 = .077$] or hue $\times$ chroma interaction ($F < 1$), but for dark colors, a main effect of hue indicated that blue was happier than yellow [$F(1, 42) = 7.96$, $p = .007$, $\eta_p^2 = .159$], and a hue $\times$ chroma interaction indicated this difference was especially large for high chroma colors [$F(1, 42) = 12.83$, $p = .001$, $\eta_p^2 = .234$].

Taken together, this experiment and Experiment 2 consistently showed that lighter, higher chroma colors were rated as happier. However, the role of yellow versus blue hue was more tenuous in this experiment, and it is unclear why. Nonetheless, neither experiment provided evidence that blue hues were *sadder* than yellow hues within a given lightness and chroma level.

Angry/not-angry. The mean angry/not-angry ratings (Fig. 7B) were strongly correlated with the corresponding ratings from Experiment 2 (Fig. 4B) ($r(30) = .90$, $p < .001$). A forward-stepping MLR model accounted for 87% of the variance with L^* (48% variance), a^* (+27% variance), and b^* (+12% variance) entered in that order ($F(3, 28) = 63.38$, $p < .001$, $A_{\text{pred}} = 17.52 - 0.70L^* + 0.93a^* + 0.61b^*$). Like Experiment 2, angrier colors were lighter, yellower, and redder, but unlike Experiment 2, C^* was not entered into the model. The LMER model showed effects of L^*, a^*, and b^*, and an $L^* \times b^*$ interaction as in Experiment 2, but there was no effect of C^*. New here, there was also an $L^* \times a^*$ interaction due to a peak of angriness for dark, saturated colors of reddish hue here, but not Experiment 2.

Despite some differences between Experiment 2 and this replication, the overall patterns of ratings were highly similar (Figs. 4 and 7), and the key components of the analyses replicated. Our observation that color–emotion associations are stable across experiments aligns with reports of [34]. Others have suggested color–emotion associations do not replicate, but they focused on comparisons between the top few associated colors with particular emotions, rather than the overall pattern across colors [47].

Color appearance ratings. Mean color appearance ratings were strongly correlated with those of Experiment 2 for light/dark ($r(30) = .995$, $p < .001$), red/green ($r(30) = .99$, $p < .001$), and yellow/blue ($r(30) = .98$, $p < .001$), but not for saturation ($r(30) = .07$, $p = .696$). We repeated the forward-stepping MLR analyses for lightness and saturation as in Experiment 2. For lightness ratings, 98% of the variance in mean lightness ratings was accounted for by L^* (98% variance) and b^* (+less than 1% variance) in that order ($F(2, 29) = 776.01$, $p < .001$, $L_{\text{pred}} = -105.70 + 1.98L^* - 0.22b^*$). Similar to Experiment 2, lightness ratings were dominated by L^*, but there was a small effect of b^* with bluer colors rated as lighter. For saturation ratings, 89% of the variance was accounted for by L^* (52% variance), C^* (+30% variance), and b^* (+7% variance) ($F(3, 28) = 78.59$, $p < .001$, $S_{\text{pred}} = -71.39 + 0.68L^* + 2.06C^* - 0.43b^*$). Like Experiment 2, there was a positive weight on C^* and negative weight on b^*, but

unlike Experiment 2 there was a positive rather than negative weight on L^*, and no weight on a^*. These results are supported by LMER models (Table 3)—saturation was rated higher for dark than light colors in Experiment 2, but this relationship reversed here (Fig. 7E). The difference between saturation ratings between the present data and Experiment 2 seems to be due to slight variations in the way saturation was defined for participants. Defining saturated as "very colorful and vivid" and desaturated as "muted and grayish" here (as opposed to "very vivid" versus "close to gray" in Experiment 2) seems to have caused observers to emphasize lightness more in their saturation ratings. Considering how subtle changes in definitions affect saturation judgments may be relevant for interpreting other studies that relied on observers' judgments of saturation (e.g., [41,42]).

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. S. E. Palmer, K. B. Schloss, Z. Xu, and L. R. Prado-León, "Music–color associations are mediated by emotion," Proc. Natl. Acad. Sci. USA **110**, 8836–8841 (2013).
2. N. Kaya and H. Epps, "Relationship between color and emotion: a study of college students," Coll. Student J. **38**, 396–405 (2004).
3. M. R. Zentner, "Preferences for colours and colour–emotion combinations in early childhood," Dev. Sci. **4**, 389–398 (2001).
4. G. L. Collier, "Affective synesthesia: extracting emotion space from simple perceptual stimuli," Motiv. Emotion **20**, 1–32 (1996).
5. L. B. Wexner, "The degree to which colors (hues) are associated with mood-tones," J. Appl. Psych. **38**, 432–435 (1954).
6. K. A. Barchard, K. E. Grob, and M. J. Roe, "Is sadness blue? The problem of using figurative language for emotions on psychological tests," Behav. Res. Meth. **49**, 443–456 (2017).
7. M. M. Terwogt and J. B. Hoeksma, "Colors and emotions: preferences and combinations," J. Gen. Psychol. **122**, 5–17 (1995).
8. H. J. Eysenck, "A critical and experimental study of colour preferences," Am. J. Psychol. **54**, 385–394 (1941).
9. J. P. Guilford and P. C. Smith, "A system of color-preferences," Am. J. Psychol. **72**, 487–502 (1959).
10. S. E. Palmer and K. B. Schloss, "An ecological valence theory of human color preference," Proc. Natl. Acad. Sci. USA **107**, 8877–8882 (2010).
11. A. C. Hurlbert and Y. Ling, "Biological components of sex differences in color preference," Curr. Biol. **17**, R623–R625 (2007).
12. K. Yokosawa, K. B. Schloss, M. Asano, and S. E. Palmer, "Ecological effects in cross-cultural differences between us and Japanese color preferences," Cogn. Sci. **40**, 1590–1616 (2016).
13. P. Sorokowski, A. Sorokowska, and C. Witzel, "Sex differences in color preferences transcend extreme differences in culture and ecology," Psychon. Bull. Rev. **21**, 1195–1201 (2014).

14. C. Soriano and J. Valenzuela, "Emotion and colour across languages: implicit associations in Spanish colour terms," Social Sci. Inf. **48**, 421–445 (2009).

15. D. Jonauskaite, A. M. Abdel-Khalek, A. Abu-Akel, A. S. Al-Rasheed, J.-P. Antonietti, Á. G. Ásgeirsson, K. A. Atitsogbe, M. Barma, D. Barratt, V. Bogushevskaya, M. K. B. Meziane, A. Chamseddine, T. Charernboon, E. Chkonia, T. Ciobanu, V. Corona, A. Creed, N. Dael, H. Daouk, N. Dimitrova, C. Doorenbos, S. Fomins, E. Fonseca-Pedrero, A. D. Gaspar, A. Gizdic, Y. Griber, G. Grimshaw, A. H. Aya, H. Jelena, M. Hirnstein, B. Karlsson, J. Kim, N. Konstantinou, E. Laurent, M. Lindeman, B. Manav, L. Marquardt, P. Mefoh, A. Mroczko-Wąsowicz, P. Mutandwa, S. Muthusi, G. Ngabolo, D. Oberfeld, M. Papadatou-Pastou, C. M. Perchtold-Stefan, A. Pérez-Albéniz, N. Pouyan, T. R. Soron, M. Roinishvili, L. Romanyuk, A. Salgado-Montejo, A. Sultanova, R. Tau, M. Uusküla, S. Vainio, V. Vargas-Soto, E. Volkan, G. Wasowicz-Kirylo, S. Zdravković, M. Zhang, and C. Mohr, "The sun is no fun without rain: physical environments affect how we feel about yellow across 55 countries," J. Environ. Psychol. **66**, 101350 (2019).

16. C. A. Thorstenson, A. J. Elliot, A. D. Pazda, D. I. Perrett, and D. Xiao, "Emotion-color associations in the context of the face," Emotion **18**, 1032–1042 (2018).

17. K. B. Schloss and S. E. Palmer, "An ecological framework for temporal and individual differences in color preferences," Vis. Res. **141**, 95–108 (2017).

18. R. D'Andrade and M. Egan, "The colors of emotion," Am. Ethnologist **1**, 49–63 (1974).

19. P. Valdez and A. Mehrabian, "Effects of color on emotions," J. Exp. Psychol. **123**, 394–409 (1994).

20. L. Wilms and D. Oberfeld, "Color and emotion: effects of hue, saturation, and brightness," Psychol. Res. **82**, 896–914 (2018).

21. H.-J. Suk and H. Irtel, "Emotional response to color across media," Color Res. Appl. **35**, 64–77 (2010).

22. G. Wyszecki and W. S. Stiles, *Color Science* (Wiley, 1982), Vol. **8**.

23. A. Munsell, *A Grammar of Color* (Van Nostrand Reinhold, 1921).

24. M. D. Fairchild, *Color Appearance Models* (Wiley, 2013).

25. B. Berlin and P. Kay, *Basic Color Terms: Their Universality and Evolution* (University of California, 1991).

26. E. H. Rosch, "Natural categories," Cognitive Psych. **4**, 328–350 (1973).

27. C. Witzel, "The role of saturation in colour naming and colour appearance," in *Progress in Colour Studies: Cognition, language and beyond*, L. MacDonald, C. Biggam, and G. Paramei, eds. (John Benjamin, 2018), pp. 41–58.

28. C. Witzel and K. R. Gegenfurtner, "Color perception: objects, constancy, and categories," Annu. Rev. Vis. Sci. **4**, 475–499 (2018).

29. M. Olkkonen, C. Witzel, T. Hansen, and K. R. Gegenfurtner, "Categorical color constancy for real surfaces," J. Vis. **10**(9):16, 1–22 (2010).

30. K. B. Schloss, L. Lessard, C. Racey, and A. C. Hurlbert, "Modeling color preference using color space metrics," Vis. Res. **151**, 99–116 (2018).

31. D. Jonauskaite, B. Althaus, N. Dael, E. Dan-Glauser, and C. Mohr, "What color do you feel? Color choices are driven by mood," Color Res. Appl. **44**, 272–284 (2019).

32. N. Dael, M.-N. Perseguers, C. Marchand, J.-P. Antonietti, and C. Mohr, "Put on that colour, it fits your emotion: colour appropriateness as a function of expressed emotion," Q. J. Exp. Psychol. **69**, 1619–1630 (2016).

33. E. M. Karp and H. Karp, "Color associations of male and female fourth-grade school children," J. Psychol. **122**, 383–388 (1988).

34. D. Jonauskaite, J. Wicker, C. Mohr, N. Dael, J. Havelka, M. Papadatou-Pastou, M. Zhang, and D. Oberfeld, "A machine learning approach to quantify the specificity of colour–emotion associations and their cultural differences," R. Soc. Open Sci. **6**, 190741 (2019).

35. B. Wright and L. Rainwater, "The meanings of color," J. Gen. Psychol. **67**, 89–99 (1962).

36. R. B. Hupka, Z. Zaleski, J. Otto, L. Reidl, and N. V. Tarabrina, "The colors of anger, envy, fear, and jealousy: a cross-cultural study," J. Cross-Cultural Psychol. **28**, 156–171 (1997).

37. L. H. Hardy, G. Rand, M. C. Rittler, J. Neitz, and J. Bailey, *HRR Pseudoisochromatic Plates* (Richmond Products, 2002).

38. M. Brauer and J. J. Curtin, "Linear mixed-effects models and the analysis of nonindependent data: a unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items," Psych. Methods **23**, 389–411 (2018).

39. B. Manav, "Color-emotion associations and color preferences: a case study for residences," Color Res. Appl. **32**, 144–150 (2007).

40. C. Witzel, J. Maule, and A. Franklin, "Red, yellow, green, and blue are not particularly colorful," J. Vis. **19**(14):27, 1–26 (2019).

41. F. Schiller and K. R. Gegenfurtner, "Perception of saturation in natural scenes," J. Opt. Soc. Am. A **33**, A194–A206 (2016).

42. F. Schiller, M. Valsecchi, and K. R. Gegenfurtner, "An evaluation of different measures of color saturation," Vis. Res. **151**, 117–134 (2018).

43. J. Koenderink, A. van Doorn, and K. Gegenfurtner, "Color weight photometry," Vis. Res. **151**, 88–98 (2018).

44. R. L. Donofrio, "The Helmholtz-Kohlrausch effect," J. Soc. Inf. Disp. **19**, 658–664 (2011).

45. C. Witzel and A. Franklin, "Do focal colors look particularly 'colorful'?" J. Opt. Soc. Am. A **31**, A365–A374 (2014).

46. M. Kleiner, D. Brainard, D. Pelli, A. Ingling, R. Murray, and C. Broussard, "What's new in Psychtoolbox-3," Perception **36**, 1–16 (2007).

47. J. M. B. Fugate and C. L. Franco, "What color is your anger? Assessing color-emotion pairings in English speakers," Front. Psychol. **10**, 206 (2019).