



# Ensemble coding of color and luminance contrast

Siddhart Rajendran<sup>1,2</sup> · John Maule<sup>3</sup> · Anna Franklin<sup>3</sup> · Michael A. Webster<sup>1</sup>

Accepted: 30 August 2020  
© The Psychonomic Society, Inc. 2020

## Abstract

Ensemble coding has been demonstrated for many attributes including color, but the metrics on which this coding is based remain uncertain. We examined ensemble percepts for stimulus sets that varied in chromatic contrast between complementary hues, or that varied in luminance contrast between increments and decrements, in both cases focusing on the ensemble percepts for the neutral gray stimulus defining the category boundary. Each ensemble was composed of 16 circles with four contrast levels. Observers saw the display for 0.5 s and then judged whether a target contrast was a member of the set. False alarms were high for intermediate contrasts (within the range of the ensemble) and fell for higher or lower values. However, for ensembles with complementary hues, gray was less likely to be reported as a member, even when it represented the mean chromaticity of the set. When the settings were repeated for luminance contrast, false alarms for gray were higher and fell off more gradually for out-of-range contrasts. This difference implies that opposite luminance polarities represent a more continuous perceptual dimension than opponent-color variations, and that “gray” is a stronger category boundary for chromatic than luminance contrasts. For color, our results suggest that ensemble percepts reflect pooling within rather than between large hue differences, perhaps because the visual system represents hue differences more like qualitatively different categories than like quantitative differences within an underlying color “space.” The differences for luminance and color suggest more generally that ensemble coding for different visual attributes might depend on different processes that in turn depend on the format of the visual representation.

**Keywords** Color and light · Color · Color and light · Contrast, color and light · Lightness/brightness

## Introduction

In complex images, observers are often more sensitive to the gist of the scene than to the individual items composing the scene (Alvarez, 2011; Ariely, 2001; Whitney & Yamanashi Leib, 2018). These summary percepts are called *ensemble coding*, and have been demonstrated for a number of stimulus features including features like size (Ariely, 2001), motion (Watamaniuk & Duchon, 1992), and orientation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), as well as high-level attributes like faces (Elias, Dyer, & Sweeny, 2017; Haberman, Harp, & Whitney, 2009; Haberman & Whitney, 2007), biological motion (Sweeny, Haroz, & Whitney, 2013), or “lifelikeness” (Leib, Kosovicheva, & Whitney, 2016). In each of these cases observers can readily

extract the average of the stimulus set even when the mean level is not included as part of the set.

Here we examined ensemble coding in color vision. Estimating the average chromaticity of the scene could play an important role in processes like color constancy, for example to estimate the color of a global illuminant. Previous studies have examined ensemble coding for nearby hues, and have examined how sensitivity to the average varies with the range of the color differences (Maule & Franklin, 2015; Webster, Kay, & Webster, 2014). Researchers have also tested for categorical biases in ensemble processing, for example to see if the perceived mean is shifted toward a category boundary (Maule, Witzel, & Franklin, 2014), and have explored how the ensemble percept depends on the number of elements or variance of the set (Maule & Franklin, 2016; Rajendran & Webster, 2020; Virtanen, Olkkonen, & Saarela, 2020). In general, however, it remains unknown how and how well the visual system could “compute” the average of a set of colors, and what rules determine this averaging.

These previous studies focused primarily only on the dimension of hue, whereas color also varies along the dimensions of saturation and lightness. Incorporating these dimensions provides a richer test of the capacity for ensemble coding

---

✉ Michael A. Webster  
mwebster@unr.edu

<sup>1</sup> Department of Psychology, University of Nevada, Reno, NV, USA

<sup>2</sup> LV Prasad Eye Institute, Hyderabad, India

<sup>3</sup> School of Psychology, University of Sussex, Falmer, Sussex, UK

within color space. To explore the limits of this coding, we examined stimulus sets that varied in chromatic (saturation) or luminance (lightness) contrast relative to a neutral gray. Gray is special in color processing for a number of reasons. First, ensemble coding in color has generally not shown a categorical effect (Maule et al., 2014). However, the categories tested have been between adjacent hues (e.g., the blue-green boundary) and thus vary in the relative proportions of two hues (Chetverikov, Campana, & Kristjansson, 2017; Maule et al., 2014; Maule & Franklin, 2015). Gray offers a stronger test of categorical coding because transitions through the gray point result in completely distinct complementary colors. Thus, gray represents the strongest instantiation of a categorical color boundary. Second, gray represents a null or norm in color processing. Specifically, many stimulus dimensions (e.g., color (Webster & Leonard, 2008), faces (Valentine, Lewis, & Hills, 2015), blur (Elliott, Georgeson, & Webster, 2011), or aspect ratio (Elias & Sweeny, 2020)) can be modelled in a perceptual space in which individual variations of the stimulus appear to vary relative to a unique norm. The norm itself appears neutral, and the physical stimulus corresponding to the norm may elicit a response null within the encoding mechanism. Thus the norm has been theorized to hold a special perceptual status in visual coding (Webster, 2015). Third, it is not clear how ensemble percepts would incorporate this norm-based representation for color (or other norm-based stimulus attributes). For example, suppose an observer is exposed to a set of equal-contrast hues that uniformly sample the color circle. What would the perceived average of this set be? A metrical averaging of the chromaticities would yield an average of zero and thus the neutral gray. Yet this average would have a very different saturation than any of the color elements, and it is possible that observers might instead assume that the “average” had the same saturation as the elements, but differed in hue. That is, observers might instead compute the average independently for hue and saturation, and the nature of the ensemble percepts might therefore point to the representation of color at the level at which ensemble coding occurs. Finally, color can vary not only in hue and saturation but also lightness. A zero-contrast gray is also a categorical boundary for lightness differences, since it demarcates the transition from increments to decrements (white to black). However, it is not clear whether the complements of light and dark behave in the same way as, for example, red and green.

To examine these questions we measured ensemble percepts for sets composed of the same hue (e.g., different saturations of red) versus different hues (ensembles with different levels of both red and the complementary hue), or that varied in chromatic contrast versus luminance contrast. Our aim was to examine how the visual system summarizes information within versus between different perceptual categories, and the implications of this encoding for the perceptual representation of color.

## Methods

### Participants

A total of 30 unique observers participated in the study, with different participants tested in different subsets of the conditions, and some in more than one experiment (ten (six female) total participants for Experiment 1; ten (nine female) for Experiment 2; 18 (13 female) for Experiment 3). With the exception of one participant, all were recruited from the University of Nevada, Reno, student subject pool and were naïve to the specific aims of the experiment. Participants had normal color vision as assessed using the Cambridge Color Test, and gave informed consent following the protocols approved by the university’s Institutional Review Board.

### Stimuli

Stimuli were presented on a CRT monitor controlled by a Cambridge Research Systems Visage graphics system. Chromaticities and luminances on the display were calibrated with a PR655 spectroradiometer. For all the conditions, ensembles were made of 16 randomly positioned circles arranged in a 4 x 4 irregular grid. At a testing distance of 100 cm, each circle subtended 2° and their centers were separated by 4° with a random jitter of  $\pm 0.5^\circ$ . They were shown on a neutral gray background with the chromaticity of Illuminant C and a mean luminance of either 5 cd/m<sup>2</sup> (Experiment 1) or 20 cd/m<sup>2</sup> (Experiments 2 and 3). Throughout we use the term “gray” to refer to this zero-contrast, achromatic background level. Depending on the condition, the individual circles either varied in their relative saturation or in their relative luminance, which was defined photometrically and thus not adjusted for individual observers. In the saturation condition, ensembles varied along a randomly chosen hue axis or between complementary hues, while in the luminance-varying condition, on each trial the elements had the same randomly chosen chromaticity.

### Experiments

We conducted three separate experiments that differed primarily in how the color and luminance variations in the ensemble elements were defined.

**Experiment 1: Chromatic ensembles varied in a uniform color space** We used two different measures of chromatic contrast. The first experiment was based on distances in the CIELAB color space, which is designed to be perceptually uniform so that equal distances in the space represent equal perceptual differences. On each trial the stimuli had a fixed randomly chosen hue angle within the space and equal steps of contrast values ranging from up to -60 to 60 relative to the gray point,

corresponding to the two complementary poles of the axis. The magnitude of the differences along the axis corresponded to Delta E values, where a Delta E of 1 corresponds roughly to a just noticeable color difference. The luminance of the elements was 20 cd/m<sup>2</sup> while the background was 5 cd/m<sup>2</sup>. This was done so gray targets in the ensemble would also differ from the background and thus could not be classified simply on the basis of an “absence” of a stimulus.

Nine different ensembles were tested, composed of different sets of contrasts. These are listed in Table 1 along with the rationale for their sets. The first three had equal contrast along the two complementary colors (two levels of each) and thus had a mean of gray. These were included to examine whether observers would misperceive the mean gray to be part of the set even though it did not share the “hue” of any of the members. They differed in the contrast range spanned by the ensemble, and thus the distance from gray. Sets C4–C6 had a higher contrast along a given hue angle compared to the complement. These were used to test for possible interactions between hue and saturation. Finally, in C7–C9 the colors were restricted to only one side of the chromatic axis and thus contained only one hue, with (C7) or without (C8 and C9) a gray element. These were used to test for potential within-hue categorical effects. For all of the ensembles we measured reported membership rates for the same nine targets that included the four ensemble members and five non-members. In the experiment the three subsets were run in separate sessions with the ensemble order counterbalanced across participants.

**Experiment 2: Chromatic ensembles varied in a cone-opponent space** Because CIELAB only approximates uniformity, we also conducted a second experiment where the color contrast steps were empirically determined to yield equal perceptual differences. For this experiment the chromatic contrasts were based on a scaled version of the MacLeod-Boynton chromaticity diagram (MacLeod & Boynton, 1979), which represents a plane of constant luminance defined

by opposing signals in the long- and medium-wavelength cones (LvsM) or signals in the short-wavelength (S) cones opposed by the L and M cones. Contrasts along each axis were based on a scaled version of the space designed to roughly equate threshold sensitivity along the LvsM and SvsLM cardinal axes, based on a previous study (Webster, Miyahara, Malkoc, & Raker, 2000). The specific conversion between the scaled space and the MB space is given by:

$$\text{LvsM contrast} = (r_{mb} - 0.6568) * 2754$$

$$\text{SvsLM contrast} = (b_{mb} - 0.01825) * 4099$$

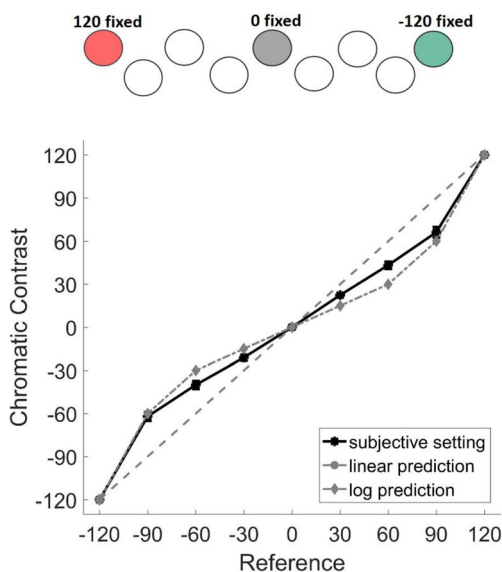
where LvsM is the reported contrast level, and  $r_{mb}$  and  $b_{mb}$  are the chromaticity coordinates of the stimuli in the MacLeod-Boynton color space.

Within this space we empirically evaluated equal perceptual contrast differences in order to try to equate the perceived differences between adjacent contrast levels in the ensembles. To do this we used a scaling task, in which contrast levels along the LvsM axis of the space were displayed as a row of elements with the same dimensions of the ensemble stimuli (Fig. 1). The ends of this series were shown fixed at -120 and +120 chromatic contrast and the center element at 0 contrast. Seven participants (four females) adjusted the remaining six intermediate contrast levels until they appeared to increase uniformly in saturation. Results were based on the mean of five repeated settings per participant. Figure 1 shows that in the cone-opponent space the required scaling is nonlinear, and is biased toward, but less than, a constant ratio scaling. We therefore adopted constant ratio steps to reassess the ensemble percepts.

For this experiment we also modified the stimuli so that the elements had a luminance of 20 cd/m<sup>2</sup> (equivalent to the background luminance) and thus differed from the background only in chromatic contrast. So that they remained clearly visible from the equiluminant background, in this case the elements were delimited from the gray background by narrow black borders.

**Table 1** Chromatic (C) ensembles used for Experiment 1. Contrast magnitudes refer to the distance (Delta E) from gray in CIELAB, with positive and negative values corresponding to opposite poles of the color axis

Ensemble	Contrast level	Feature of the ensemble
C1	-30,-15,15,30	Unbiased ensembles with gray average
C2	-45,-15,15,45	
C3	-60,-30,30,60	
C4	-15,0,15,30	Biased ensembles to test for interactions between hue and saturation
C5	-30,-15,45,60	
C6	-30,0,30,60	
C7	0,15,30,45	Ensembles that span or do not span the gray boundary to test for categorical effects at gray
C8	15,30,45,60	
C9	30,45,60,75	
Targets	-60,-45,-30,-15,0,15,30,45,60	



**Fig. 1** Chromatic contrast scaling task. **Top:** Observers were shown fixed extremes of the chromatic axis and then adjusted the intermediate levels to produce perceptually equal contrast steps. **Bottom:** The subjective contrast of the ensemble stimuli estimated from the contrast scaling. Note reference values of +30, 60, or 90 indicate the target steps (.25, .5, or .75 of the 120 max) that observers adjusted the chromatic contrasts for. Data points are the mean settings across observers  $\pm 1$  standard error. Lines show the scaling predicted by a linear (dashed) or (suprathreshold) log contrast response (diamonds)

The four ensembles tested for this experiment are listed in Table 2, and were tested as part of a single session. The first two of these again had equal contrast along the two complementary colors (two levels of each) and thus had a mean chromaticity of gray, and were again used to test whether observers would misperceive the mean gray to be part of the set even though it did not share the “hue” of any of the members. They differed in the contrast range spanned by the ensemble, and thus the distance from gray. The third was biased to have a higher contrast along a given hue angle compared to the complement and again tested for interactions between hue and saturation. Finally, the fourth was restricted to one hue angle and the gray point, without the complementary axis, to compare performance for single hues and the effect of the gray boundary. As in the preceding experiment, targets were the same for each set and included the four ensemble levels and five intervening levels.

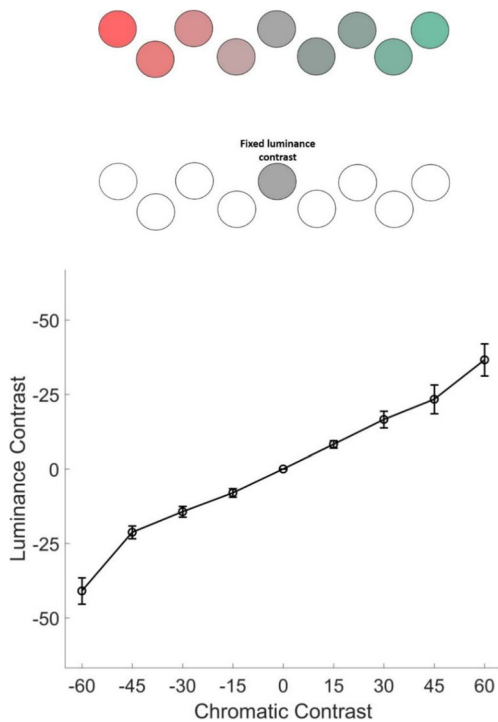
**Experiment 3: Ensembles varying in luminance rather than chromatic contrast** In the third experiment, individual circles in an ensemble had the same chromaticity, but were either all darker or lighter relative to gray or were a combination of increments and decrements. The luminances of the elements are given by the proportional Weber contrast times 100 (e.g., a value of +60 was 1.6 times the background contrast or 32 cd/m<sup>2</sup> while a value of -60 was 8 cd/m<sup>2</sup>). When presenting these ensembles, the background on the monitor was still a neutral gray, but the elements had a fixed chromatic contrast of 30 in the cone-opponent plane, and again varied randomly in hue on each trial. Thus, the set appeared as different lightness levels of a desaturated red or green, etc. The chromatic contrast was added so that the zero-contrast element appeared distinct from the gray background.

In order to compare the results for chromatic versus luminance variation, the differences in perceived lightness versus saturation in the ensembles need to be comparable. Rather than assume the CIELAB scaling this, we again empirically evaluated the lightness scale by asking a set of participants to perform a contrast-matching task between luminance and chromatic contrast, a task that can be performed reliably (Switkes & Crognale, 1999). For this, nine equally spaced contrasts ranging from -60 to 60 Delta E along the LvsM axis were displayed as circles in an upper row, and then luminance levels were adjusted in circles shown in a corresponding lower row until the lightness steps appeared as the same magnitude as the chromatic steps (Fig. 2). In this case only the central gray was fixed, and the four observers varied all eight of the other lightness levels. The result of this experiment showed a roughly linear relationship between luminance and chromatic contrast (Switkes & Crognale, 1999), but indicated that the nominal range of -60 to 60 Delta E for chromatic contrast corresponded to a range of approximately -40 to 40 for luminance contrast (Fig. 2). We therefore used a linear scaling of the CIELAB luminance values adjusted for this range.

The ensembles for the luminance-varying sets and the rationale for using them are listed in Table 3. The first two again had a mean equal to the background luminance level, to test whether this level was less likely to be misperceived as a member of sets with only increments and decrements. The third and fourth again tested for interactions between the

**Table 2** Chromatic contrasts of the ensembles and test stimuli. Contrasts correspond to the levels in the scaled cone-opponent space

Ensemble	Contrast levels	Features of the ensemble
E1	-60,-15,15,60	Unbiased ensemble with gray average
E2	-120,-30,30,120	Unbiased ensemble with gray average
E3	-30,0,30,120	Biased 2-hue ensemble
E4	0,15,30,60	Single-hue ensemble
Targets	-120,-60,-30,-15,0,15,30,60,120	



**Fig. 2 Top:** Contrast matching task for luminance and chromatic contrast. Observers were shown a fixed scale corresponding to equal Delta E steps along the LvsM axis and adjusted the luminance of the lower circles so that the step sizes in luminance contrast appeared equivalent. **Bottom:** Average settings for equating luminance and chromatic contrast. Data points are the mean across observers  $\pm 1$  standard error

magnitude and sign of the luminance of the elements. The final sets consisted of decrement-only (L5–L7) or increment-only (L8–L10) sets to test for categorical effects at the background luminance level separating increments and decrements. Ensembles L1–L4 and L5–L10 were tested in different sessions with the order again randomized across participants.

## Procedure

For each of the experiments participants performed a member identification task where an ensemble was presented for 0.5 s and then was followed after 1 s by the presentation of a single target stimulus (Fig. 3). The observer was given unlimited time to report if the target was a part of the presented ensemble. All ensembles were made of four contrast levels, and with the hues varied randomly on each trial. Specifically, for a given ensemble of chromatic contrasts the contrast levels were fixed, but the hue angle of the set was randomly rotated for each presentation (within the CIELAB or cone-opponent space). The test target included the four contrast levels present in the ensemble and five additional contrasts that included three intermediate levels and two levels outside the ensemble range. During a single session, observers were tested on three or four different ensembles. Each ensemble/target condition was shown in random order for total of 20 repetitions, from which we calculated the proportion of times the observer thought the target level was present.

## Results

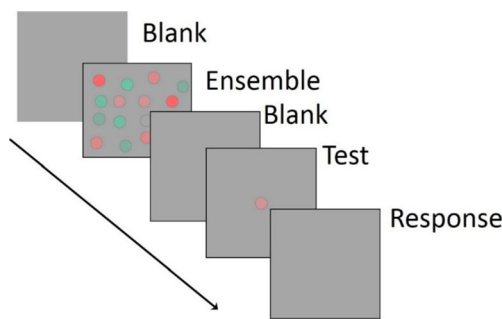
### Experiment 1: Color ensembles in CIELAB space

**Unbiased color ensembles (gray average)** As noted, in the first experiment we examined performance for ensembles that varied in chromatic contrast relative to the gray, defining the contrasts by the colorimetric distances in CIELAB. The results for the first three ensembles are shown in Fig. 4. These were again chosen so that the average of the colors was gray. Contrasts that were higher than the range of presented levels had lower false-alarm rates, while the intervening stimuli within the range were equally likely to be perceived as part of the set whether they were members or not. These false alarms are consistent with ensemble coding for these

**Table 3** Ensembles for luminance contrast

Ensemble	Contrast levels	Feature of the ensemble
L1	-30,-10,10,30	Unbiased ensembles with gray average
L2	-40,-20,20,40	
L3	-50,-30,-10,10	Biased ensembles to test for interactions between contrast polarity and magnitude
L4	-10,10,30,50	
L5	0,-10,-20,-30	Decrement-only ensembles with or without the zero-contrast gray
L6	-10,-20,-30,-40	
L7	-20,-30,-40,-50	
L8	0,10,20,30	Increment-only ensembles with or without the zero-contrast gray
L9	10,20,30,40	
L10	20,30,40,50	
Targets	-40,-30,-20,-10,0,10, 20, 30, 40	

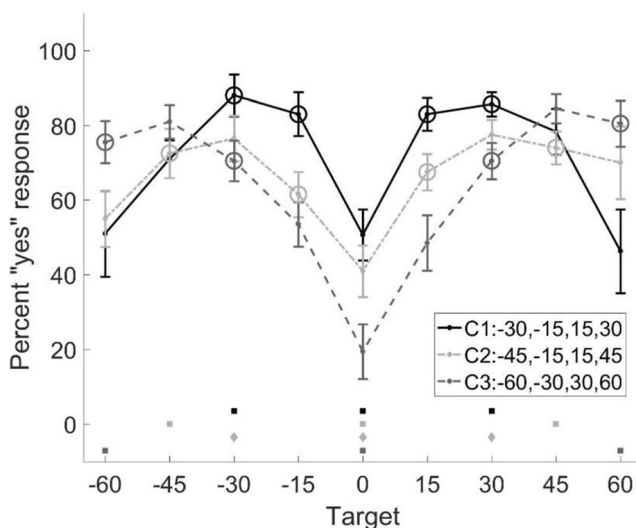




**Fig. 3** Member identification task. On each trial participants were shown an ensemble with four contrast levels shown in 16 randomly spaced elements. The ensemble appeared for 0.5 s followed by a 1-s blank. One of nine target contrasts was then displayed and the observer responded whether the contrast was present in the ensemble

intermediate hues. However, the exception was gray, which had comparatively low false-alarm rates even though it corresponded to the mean of the set. This suggests that observers were not forming a strong ensemble percept of the mean of the contrast, but rather may have averaged within each complementary hue. On the other hand, the proportion of false alarms was still substantial for the gray average. This could be because of the proximity to gray of the low contrast elements in the sets (as suggested by the higher rates for gray for C1 and C2) or because the trials were interleaved with sets that did include gray. Nevertheless, the main point is that gray was misperceived less than the chromatic samples within the gamut even though it was the metric average of the gamut.

The drop in the false-alarm rate for gray differs from the common finding that mean of an ensemble is more likely to be reported as presented than individual members of the set



**Fig. 4** Contrast ensembles with a gray average. Plots show the percent of times each target contrast was reported as a member of the set for the three different ensembles. Data points are the mean across observers  $\pm 1$  standard error. The circled data points show targets that were members of the set. Symbols at the bottom of the plot show the targets that were compared in statistical analyses of the effects (see text)

(Whitney & Yamanashi Leib, 2018). To formally assess this, we compared the false-alarm rate for the achromatic mean with the rate of correctly reporting the most extreme contrasts in each set, using a repeated-measures analysis of variance (RMANOVA) with correction for multiple comparisons. The specific target levels compared are shown by the points near the bottom of Fig. 4 for each ensemble. These analyses are summarized in Table 4. In all but one case the high-contrast member elements were significantly more likely to be reported than the average gray. For ensemble C2 we also compared the false alarms for the gray target to those for the nearest-neighbor chromatic non-member (+30 or -30). This is shown in the final row of Table 4, and in both cases the gray was reported significantly less often than the chromatic non-members.

**Biased color ensembles (non-gray average)** The next color sets included biased ensembles (C4–C6) with higher contrasts for one of the hues and lower contrasts for the complementary hue. If observers coded hue and contrast independently, then they might be expected to mistake a high contrast of both hues as a member of the set. For example, when shown a high-contrast red and low-contrast green, they might separately encode the saturation (high and low) and hue (red and green), and thus misreport the presence of a high-contrast green. This would be equivalent to an “illusory conjunction” of hue and saturation, in which two features are correctly perceived but how they are related are not (Treisman & Schmidt, 1982). However, the responses instead mirrored the asymmetry of the distribution, with lower false alarms for the outlying contrast (Fig. 5). We assessed this by comparing the hit rate for a contrast shown as a member hue to the false-alarm rate for the same contrast shown in the non-member complementary hue (Table 5). In each case the false alarms were lower, suggesting again that the hue and saturation were not encoded independently.

**Ensembles with one hue category** The remaining ensembles (C7–C9) were included to examine potential categorical effects at the achromatic boundary. These sets included only one hue category, either with (C7) or without (C8 and C9) the zero-contrast gray as a member. There was a strong drop in the false-alarm rates for the probes on the opposite side of gray (Fig. 6). However, the rate of fall-off was similar for the three sets. In particular, the change from the lowest-contrast member to the nearest non-member was not significantly different whether that step was to a lower contrast of the same hue (C9), to gray (C8), or to the complementary hue (C7) ( $F(2,24) = 1.02, p=0.38$ ). Thus, for these conditions – where only one hue was displayed – the achromatic point did not emerge as special. Nevertheless, these results remain consistent with an averaging process that occurs primarily within rather than between the complementary hue categories.

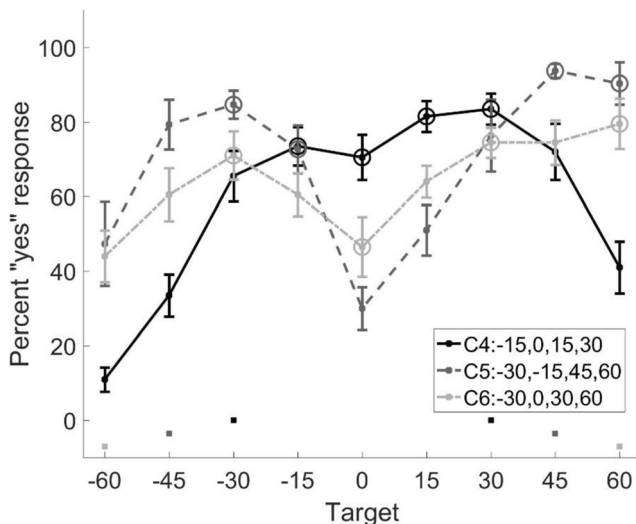
**Table 4** The first three rows show comparisons of the false-alarm rates for the non-member achromatic target with the hit rate for the highest contrast members of each ensemble. The final row compares the false

alarms for the gray target with the false alarms for the nearest non-member targets for ensemble C2

Ensemble	Contrast levels compared (mean membership rate $\pm$ SD)			Omnibus RMANOVA test	Bonferroni-adjusted post hoc paired t-test
C1: -30, -15, 15, 30	-30 (88 $\pm$ 12.6)	0 (50.6 $\pm$ 15.3)	30 (85.7 $\pm$ 7.2)	F (2,12) = 23.1; p < 0.001	-30 vs. 0: p = 0.004 0 vs. 30: p = 0.008
C2: -45, -15, 15, 45	-45 (72.5 $\pm$ 19.7)	0 (41 $\pm$ 20.7)	45 (77.5 $\pm$ 8.63)	F (2,18) = 4.23; p = 0.031	-45 vs. 0: p = 0.4 0 vs. 45: p = 0.04
C3: -60, -30, 30, 60	-60 (75.5 $\pm$ 16.9)	0 (19.5 $\pm$ 21.9)	60 (80.5 $\pm$ 18.5)	F (2,18) = 35.5; p < 0.001	-60 vs. 0: p = 0.001 0 vs. 60: p < 0.001
C2: -45, -15, 15, 45	-30 (76.5 $\pm$ 18.3)	0 (41 $\pm$ 20.7)	30 (77.5 $\pm$ 11.8)	F (2,18) = 15.6; p < 0.001	-30 vs. 0: p = 0.009 0 vs. 30: p = 0.004

**Experiment 2: Color ensembles in cone-opponent space**

The rejection rate for gray relative to the other intermediate non-member targets suggests that gray is not readily perceived as the average of an ensemble made of equal contrasts of two complementary colors. That is, observers were less likely to mistake a zero-contrast stimulus for a member of an ensemble composed of visible color contrasts, even though it corresponded to the mean chromaticity of the set. However, this effect might also be due to the “perceptual” distance of gray from the sample contrasts. For example, if there was a saturating nonlinearity in the contrast response, then higher contrasts would appear more similar to each other and gray might appear farther removed from the ensemble members.



**Fig. 5** Membership rates for asymmetric ensembles with a higher contrast for one hue than the complementary hue. Plots show the percent of times each target contrast was reported as a member of the set for the three different ensembles. Data points are the mean across observers  $\pm$ 1 standard error. The circled data points show targets that were members of the set

That is, the intermediate non-members might be perceptually more similar to the displayed set, and thus more likely to be misclassified. We used equal spacing of contrasts within a perceptually uniform space to control for this potential confound. However, since such spaces are known to only approximate uniformity, in the next set of experiments we repeated key conditions after empirically evaluating the contrast scaling as described in the *Methods*.

**Unbiased color ensembles (gray average)** Results for the first two ensembles for the new chromatic contrasts are shown in Fig. 7. This replicates the pattern found previously (Fig. 2) with the gray targets showing lower false alarms even though they represented the mean level of the set. We again assessed this by comparing the membership rates for the gray versus the highest contrast members, or for the gray versus the nearest chromatic non-members (Table 6). For each ensemble the gray versus member differences were significant for one of the contrasts but not for the second; and similarly, for the gray versus chromatic non-member, the gray false alarms were lower in two of the four comparisons. Thus, these conditions produced mixed evidence but still suggest that the mean grays tend to be less likely to be perceived as part of the ensemble.

**Biased color ensembles (non-gray average)** Responses for the two biased ensembles are shown in Fig. 8. In one case the set had only a single hue (E4), and false alarms fell precipitously for targets with the complementary hue. Thus, not surprisingly, observers were sensitive to both the hue and contrast of the elements. However, the drop is notably stronger when crossing the gray boundary compared to the higher contrast (120) foil. This could reflect a categorical effect for the gray, such that the averaging and perceived membership was again largely confined to the displayed hue. However as noted above, we also found steep drops for lower contrast stimuli that fell outside the gamut of the ensemble (Fig. 6). Finally, the remaining ensemble had a high and low contrast for one hue but only a

**Table 5** Comparison of membership rates for complementary non-member vs. member contrasts

Ensemble	Contrast levels compared (mean membership rate $\pm$ SD)		Paired t-test
C4: -15,0,15,30	-30	30	$t(9) = 3.85$
	(65.5 $\pm$ 20.2)	(83.5 $\pm$ 12.4)	$p = 0.004$
C5: -30,-15,45,60	-45	45	$t(9) = -2.88$
	(79.3 $\pm$ 14.9)	(93.6 $\pm$ 4.15)	$p = 0.28$
C6: -30,0,30,60	-60	60	$t(9) = -6.6$
	(44 $\pm$ 20.6)	(79.5 $\pm$ 20.3)	$p < 0.001$

low contrast for the complementary hue. The responses again paralleled this asymmetry, with higher reports for the 120-contrast member than the -120-contrast non-member ( $t(8) = 5.25, p < .001$ ). This again suggests that the contrast and hue were not encoded independently in ways that led them to be strongly confounded. On the other hand, the fall-off for the outside non-members is relatively gradual compared to the single-hue ensemble. This could reflect a partial confound of hue and contrast or potentially an effect of the overall variance of the ensembles.

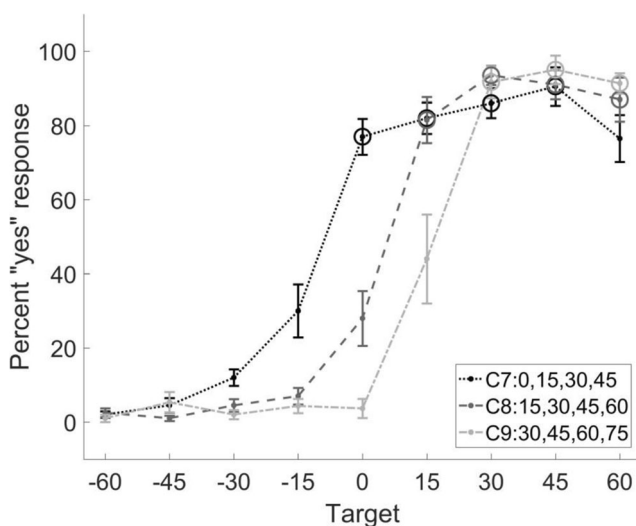
In sum then, the settings confirmed the primary findings of the first experiment, suggesting that these findings were not simply due to how the contrasts of the elements were defined. Both experiments suggest that (a) ensemble coding of contrast does not reflect a simple metrical averaging of the contrasts; (b) hue and saturation appear to be represented conjointly in ensemble coding, so that the average is not computed independently for the two attributes; and (c) the falloff in false

alarms is strong across the category boundary, suggesting this boundary may delineate how the colors are averaged.

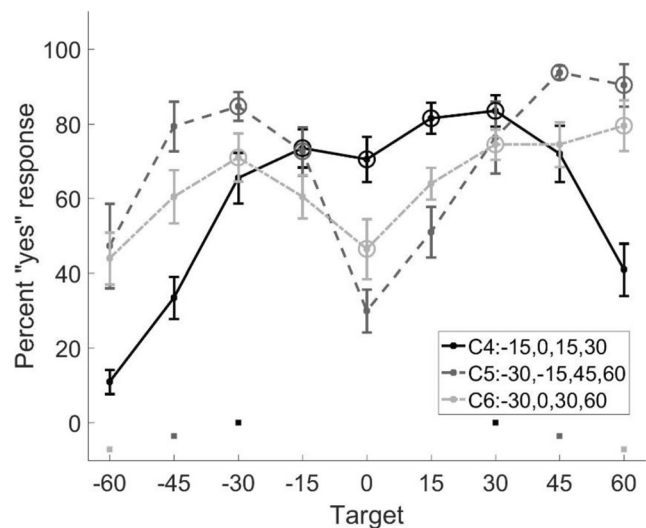
### Experiment 3: Luminance contrast ensembles

In the final set of experiments our aim was to compare the properties of ensemble coding for lightness variations versus saturation variations. While light and dark are again complementary pairs, they may not share the same degree of categorical separation as complementary colors. For example, lightness in some cases may behave more like a single continuum. We therefore asked whether ensembles varying in luminance contrast might be encoded differently from those defined by chromatic contrast. To test this, we conducted the same measurements but now for stimuli varying only in lightness.

**Unbiased lightness ensembles (gray average)** In the first case we again examined ensembles where the mean luminance



**Fig. 6** Membership rates for ensembles with a single hue category with (C7) or without (C8, C9) an achromatic member. Plots show the percent of times each target contrast was reported as a member of the set for the three different ensembles. Data points are the mean across observers  $\pm 1$  standard error. The circled data points show targets that were members of the set



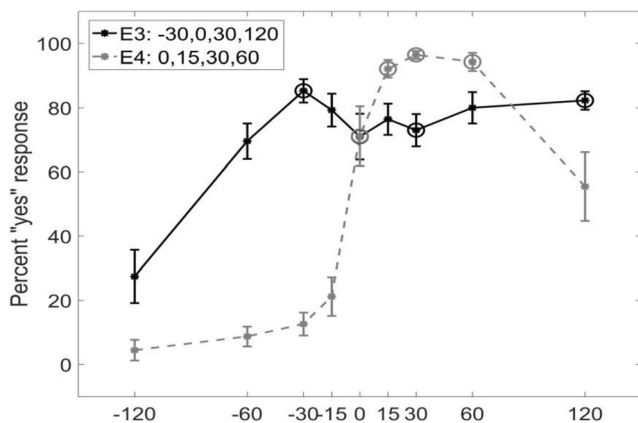
**Fig. 7** Contrast ensembles defined by the cone-opponent axes with a gray average. Plots show the percent of times each target contrast was reported as a member of the set for the two different ensembles. Data points are the mean across observers  $\pm 1$  standard error. The circled data points show targets that were members of the set. Symbols at the bottom of the plot show the targets that were compared in statistical analyses of the effects (see text)



**Table 6** The top two rows compare false-alarm rates for the achromatic target with the hit rate for the highest contrast members of each set. The bottom two rows compare false-alarm rates for the achromatic target with the false-alarm rate for the nearest chromatic non-member for each set

Ensemble	Contrast levels compared (mean membership rate $\pm$ SD)			Omnibus RMANOVA test	Bonferroni-adjusted post hoc paired t-test
E1: -60, -15, 15, 60	-60 (92.7 $\pm$ 9.1)	0 (64.5 $\pm$ 31.2)	60 (89.4 $\pm$ 8.8)	F (2,16) = 8.33; p = 0.003	-60 vs. 0: p = 0.028 0 vs. 60: p = 0.09
E2: -120, -30, 30, 120	-120 (90.5 $\pm$ 4.6)	0 (46.3 $\pm$ 35.2)	120 (72.3 $\pm$ 20.6)	F (2,11) = 10.42; p = 0.001	-120 vs. 0: p = 0.013 0 vs. 120: p = 0.12
E1: -60, -15, 15, 60	-30 (89.5 $\pm$ 6.3)	0 (64.5 $\pm$ 31.2)	30 (85.2 $\pm$ 13.8)	F(2,16) = 4.996; p = 0.02	-30 vs. 0: p = 0.09 0 vs. 30: p = 0.24
E2: -120, -30, 30, 120	-60 (81.8 $\pm$ 16.5)	0 (46.3 $\pm$ 35.2)	60 (84 $\pm$ 11.9)	F(2,16) = 10.88; p = 0.001	-60 vs. 0: p = 0.03 0 vs. 60: p = 0.02

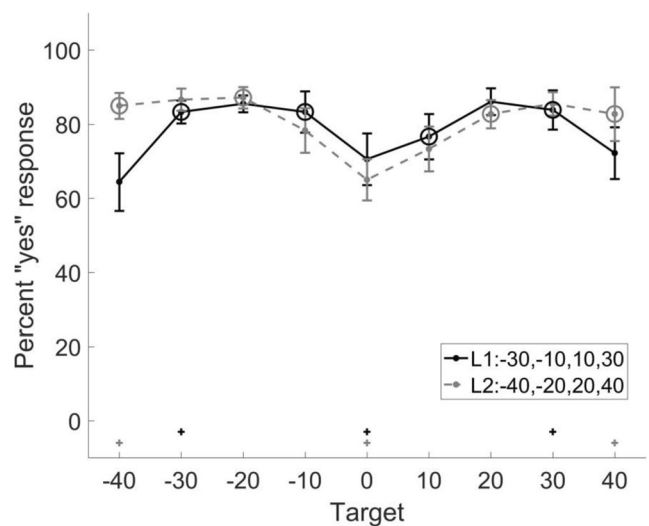
corresponded to the zero-contrast background. Responses for these conditions are shown in Fig. 9. There is again some hint of a trough in the membership responses at the achromatic point. This was assessed as before by comparing the false alarms for the gray to the hits for the high-contrast members. However, the difference was significant only for the higher-variance ensemble (Table 7). The proportion of false alarms for gray also appeared markedly higher than for the comparable saturation ensembles (C1–C3), a difference that was highly significant ( $t(43) = -4.76, p < 0.001$ ; mean false alarm for gray: saturation ensembles -35.6 (SD -24.4) and for lightness ensembles -67.8 (SD 18.6). Thus, observers were more likely to misperceive a gray when it was part of the lightness set than the saturation set. This was further confirmed in a 2 (ensemble types: L vs. C)  $\times$  2 (gray non-member vs. non-gray non-member) mixed ANOVA. The gray/non-gray was taken as the repeated measure and the ensemble type was taken as the between-group factor. The mixed-model comparison showed a significant difference between the type of non-membership ( $F(1, 17) = 21.8; p < 0.001$ ) with a significant interaction



**Fig. 8** Asymmetric contrast ensembles in the opponent color space. Plots show the percent of times each target contrast was reported as a member of the set for the two different ensembles. Data points are the mean across observers  $\pm$ 1 standard error. The circled data points show targets that were members of the set

between the two factors ( $F(1, 17) = 7.9; p = 0.01$ ) due to the higher false-alarm rates for gray in the luminance ensembles.

**Biased lightness ensembles** We similarly examined the percepts for the asymmetric ensembles. This again exhibited responses that paralleled the stimulus set (Fig. 10). As in the color ensembles, we compared the false alarms versus hits for the same absolute contrast when it was a non-member (e.g., decrement) or member (e.g., increment) of the set. Values for members of the ensembles were always greater than the false alarms for non-members, again suggesting observers were sensitive to how the magnitude and sign of the contrast were combined within the set (Table 8), or alternatively in this case, sensitive to the actual gamut of the luminance contrasts.



**Fig. 9** Luminance contrast ensembles with a gray (background level) average. Plots show the percent of times each target contrast was reported as a member of the set for the two different ensembles. Data points are the mean across observers  $\pm$ 1 standard error. The circled data points show targets that were members of the set. Symbols at the bottom of the plot show the targets that were compared in statistical analyses of the effects (see text)

**Table 7** Comparison of false-alarm rate for the non-member gray vs. the hit rate for the high-contrast member

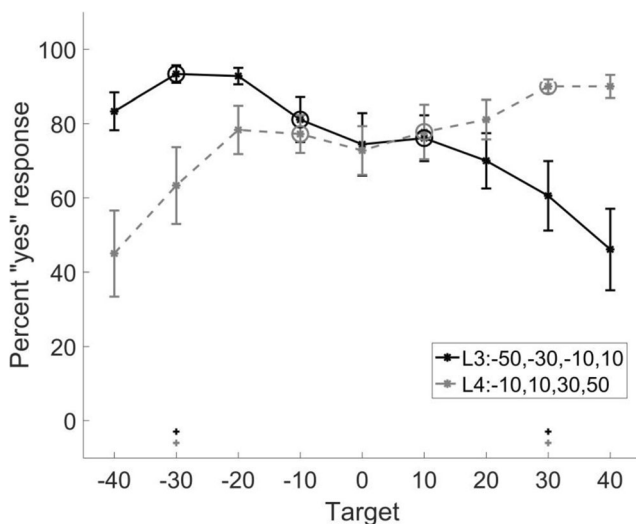
Ensemble	Contrast levels compared (mean membership rate ± SD)	Omnibus RMANOVA test	Bonferroni-adjusted post hoc paired t-test		
L1: -30, -10, 10, 30	-30 (83.3 ± 9.4)	0 (70.6 ± 20.9)	30 (83.9 ± 15.9)	F (2,16) = 3.35; p = 0.061	NA
L2: -40, -20, 20, 40	-40 (85.0 ± 10.6)	0 (65.0 ± 16.7)	40 (82.7 ± 21.8)	F (2,16) = 4.47; p = 0.029	-40 vs. 0: p = 0.036 0 vs. 40: p = 0.14

**Decrement-only or increment-only ensembles** The last set of conditions again probed the rate of fall off in false alarms around the gray category boundary. Figure 11 shows that there is a relatively gradual drop for luminance contrasts outside the ensemble. Again, as for color, there was no evidence for a steeper drop at the gray boundary. False alarms for the first contrast level outside the ensemble (L5:L10) were similar irrespective of where the ensemble ended, or whether it was a dark/ light ensemble. This was assessed with a two-way RMANOVA (3 ensemble ranges x contrast sign (increments vs. decrements)), which did not result in main effects for the range ( $F(2,42)=2.66, p=0.082$ ) or contrast sign ( $F(1,42)=3.62, p=.064$ ). However, the falloff in the false-alarm rate appeared more gradual for luminance than for color. Comparisons showed that for the first contrast level outside the ensemble, the errors were significantly higher ( $F(1, 45) = 50.88; p\text{-value} < 0.001$ ) for the lightness ensembles (mean =  $70.83 \pm 3.7$ ) than for the saturation ensembles ( $33.62 \pm 3.6$ ) for all of the sets. To control for baseline differences in the membership rates, we also compared the magnitude of the fall-off in membership reports between ensembles C4:C6 and L5:L10 using a 2 (ensemble type - saturation vs. lightness) X 3 (Gray-

1, Gray, Gray+1) ANOVA. There was a main effect of the ensemble type ( $F(1,43) = 16.9; p < 0.001$ ), but no main effect of the level of fall off relative to gray ( $F(2,43) = 0.13; p = 0.8$ ) and no interaction ( $F(2,43) = 0.525; p = 0.59$ ). Thus, the lightness dimension appeared to show substantially less sensitivity to the range of levels characterizing the sets, and, importantly, showed substantially more integration across the complementary light and dark categories.

### Discussion

An important aspect of ensemble perception is the ability to estimate the average value of the parameter of interest. In most cases this is assumed to represent the metric average of the ensemble, though previous studies have shown that the averaging can for example exclude outliers in the ensembles (Haberma & Whitney, 2010) and may give more weight to more salient elements (Kanaya, Hayashi, & Whitney, 2018). Based on the responses for contrast values within an ensemble, our findings with color show that color contrasts within a hue show evidence for such ensemble representation. These results are similar to other studies in ensemble color perception (Chetverikov et al., 2017; Maule et al., 2014; Maule & Franklin, 2015; Maule & Franklin, 2016; Maule, Stanworth, Pellicano, & Franklin, 2018; Virtanen et al., 2020; Webster et al., 2014). However, in our experiments, we aimed to study the extent to which this process could generalize across different stimulus categories, by focusing on averaging across the gray boundary. As noted in the *Introduction*, gray represents a unique categorical boundary in color perception, and thus might pose the greatest challenge to pooling signals across qualitatively different stimulus categories. Our results suggest that ensemble coding for color fails to strongly generalize across complementary color categories. In particular, grays are less likely to be perceived as part of the ensemble, even when they represent the average stimulus and even though the gray is matched for the perceptual distance from the ensemble members. Note that this is unlikely to be simply because the gray target equaled the background, because similar effects were observed when there was a large luminance difference between the elements and the background. Thus at least in the extreme our results are inconsistent with a simple



**Fig. 10** Luminance contrast ensembles with a biased mean contrast. Plots show the percent of times each target contrast was reported as a member of the set for the two different ensembles. Data points are the mean across observers  $\pm 1$  standard error. The circled data points show targets that were members of the set. Symbols at the bottom of the plot show the targets that were compared in statistical analyses of the effects (see text)

**Table 8** Tests for false-alarm vs. hit rates for the same absolute contrast as non-member (e.g., decrement) or member (e.g., increment)

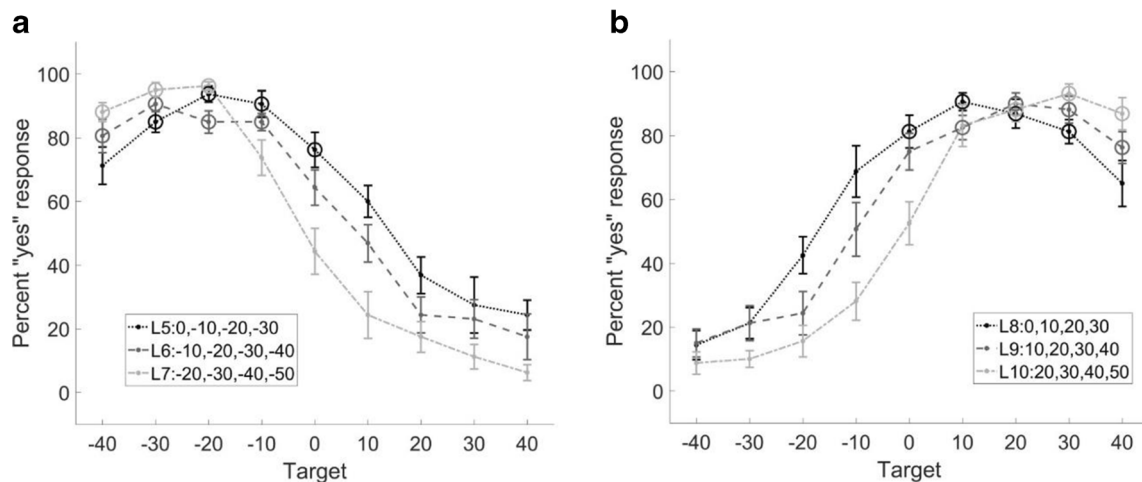
Ensemble	Contrast levels compared (mean membership rate $\pm$ SD)		Paired t-test
L3: -50, -30, -10, 10	-30 (93.3 $\pm$ 7.7)	30 (60.6 $\pm$ 28.1)	t(8) = 3.71 p = 0.006
L4: -10, 10, 30 50	-30 (63.3 $\pm$ 31.0)	30 (90.0 $\pm$ 5.6)	t(8) = 2.74 p = 0.025

metrical averaging process underlying ensemble coding for color.

A related result was reported by Rajendran and Webster (Rajendran & Webster, 2020), who examined achromatic adjustments for multi-colored arrays. They had observers adjust the mean chromaticity of the arrays so that it appeared neutral, and found that the adjustments along one chromatic axis (e.g., LvsM) were affected by the variance along an orthogonal axis (e.g., SvsLM). This is in contrast to the selectivity of masking effects for color, for example (Sankeralli & Mullen, 1997), and suggests instead that adding any variance in the distribution of hues made it more difficult to infer the average color. Importantly, some observers also reported making the achromatic adjustments by matching the relative contrasts of different hues, so that the mean was only estimated indirectly. Such results suggest that while color can be characterized and quantified in a three-dimensional space (for the trichromatic observer), the visual system may not necessarily encode color in terms of a metrical spatial representation. Consistent with this, for individuals naïve to color theory, identifying the complement of a given hue is non-intuitive and prone to large error (Webster, 2020). Moreover, studies of individual differences in color appearance suggest that hue categories vary independently across observers in ways that are inconsistent with an underlying metrical scaffolding (Emery, Volbrecht, Peterzell,

& Webster, 2017). Such results suggest that different hues may be represented more like qualitatively different “objects” than quantitatively different vectors. As such the summary percepts for hues may not involve or allow for an actual averaging in the perceptual representation, and may instead depend on indirect inferences that may be more “post-perceptual,” for example an implicit weighting of the relative “amounts” within different categories.

It is also not clear to what extent these considerations are unique to color. Many visual attributes do have a clear metrical sense (e.g., size or direction of motion), and for these, notions of the relative values of the stimuli and their summary statistics do seem intuitive and readily computable. However, for other attributes the basis for ensemble percepts are less certain. For example, observers can accurately estimate the mean expression or gender of a crowd of faces (Haberman & Whitney, 2007). Yet like color, even though expressions can be conceptualized in a low-dimensional space (Young et al., 1997), the perceptual relationships between different facial expressions are not readily accessible, and for example what constitutes a visually complementary expression may be a difficult inference (Juricevic & Webster, 2012; Skinner & Benton, 2010). Moreover, just as gray may not be directly perceived as the mean of two complementary hues, the mean of two opposite expressions may not be directly encoded as a



**Fig. 11** Responses for (a) decrement-only or (b) increment-only luminance ensembles. Plots show the percent of times each target contrast was reported as a member of the set for the different ensembles. Data points

are the mean across observers  $\pm$ 1 standard error. The circled data points show targets that were members of the set

neutral face. This raises the general question of whether ensemble perception may operate in fundamentally different ways for different perceptual attributes, and what these operations may indicate about the nature of the representations for these attributes. In particular there may be a general distinction between metrical versus non-metrical codes, with very different routes to ensemble percepts for each. We suggest that color – at least for very large hue differences – is among the latter.

A related set of work has examined how ensemble coding operates across categories or objects to understand when elements should be averaged together versus segmented into different sets (Cha & Chong, 2018; Khayat & Hochstein, 2019; Utochkin, 2015). For example, Elias and Sweeny (Elias & Sweeny, 2020) tested ensemble percepts for ellipses which were tall or flat and for which a uniform circle was thus a category boundary. They found poorer integration across than within the categories, and argued that this is because of the competing need to differentiate the categories. Our results are again consistent with the idea that very different hues behave as qualitatively distinct categories, rather than as points in a metrical space, and thus that the integration occurs largely within rather than across the hue categories.

An alternative we explored for color coding was that the visual system might independently represent a color by its perceptual attributes of hue and saturation, and then average within each of these attributes. This scheme might lead to cross-attribute errors in the false alarms. However, we also did not find evidence for this representation. This suggests that even though hue and saturation are perceptually distinct attributes, they are not processed separately as ensemble percepts. This is consistent with the finding that hue and saturation behave as integral dimensions in similarity judgments (Burns & Shepp, 1988). With regard to ensemble coding, our results are again consistent with forming separate representations of summary statistics *within* each hue category, and that the mean across categories is estimated indirectly.

Categorical effects are typically evidenced by poor discrimination within the category while heightened discrimination between categories (Harnad, 1987; Witzel & Gegenfurtner, 2018). In this regard our evidence for a categorical effect at the gray boundary was mixed. On the one hand, when ensembles included both complementary hues, for non-member targets within the ensemble range the achromatic level was unique in showing reduced false alarms (Figs. 4 and 7). However, when the ensemble included only a single hue, then gray was no more likely to be rejected than targets that were lower than the ensemble contrast range whether they were the same or different hue (Fig. 6). Thus, these low contrasts were not categorically perceived as part of the ensemble. Notably however, targets at a higher contrast than the ensemble range were more likely to be classified as part of the ensemble. In any case, our results do not point to a strong

categorical representation of contrast in the ensemble percepts. It should be emphasized again that these conditions represent the strongest categorical differences for color, since the two categories represent complementary hues. They also correspond to categorical boundaries in the responses of early color-opponent mechanisms (i.e., the LvsM and SvsLM opponent axes), which have been found to determine categorical discriminations in pre-verbal infants (Skelton, Catchpole, Abbott, Bosten, & Franklin, 2017). The lack of a strong categorical effect across gray thus suggests that categorical effects in color ensemble coding (e.g., between adjacent hues) are likely to be weak in general. They also tend to be weak in other measures of color appearance, and are strongly dependent on the task and on the potential stages influencing performance. For example, judgments that reflect basic discrimination or similarity ratings (Matera et al., 2020; Webster & Kay, 2012; Witzel & Gegenfurtner, 2013) may be less susceptible to categorical effects than tasks that require a speeded response (Gilbert, Regier, Kay, & Ivry, 2006; Winawer et al., 2007). The latter have been attributed to post-perceptual influences (Roberson, Pak, & Hanley, 2008) and have also been difficult to replicate (Brown, Lindsey, & Guckes, 2011; Martinovic, Paramei, & MacInnes, 2020; Witzel & Gegenfurtner, 2011). Thus the prevalence and nature of categorical color coding as well as the processing stages at which it might arise remains uncertain (Forder, He, & Franklin, 2017; Siuda-Krzywicka, Boros, Bartolomeo, & Witzel, 2019). In any case, our results suggest that there are in fact categorical effects for color in ensemble coding, but in the sense that the nature of the representation of color impedes explicit averaging across very different hue categories (but not in the sense that ensemble membership strongly generalizes across different contrasts within the same category).

Importantly, we observed different trends for variations in lightness levels. In this case, observers were much more likely to experience the neutral stimulus as a member of the set, even though the luminance and chromatic stimuli were matched for perceptual differences. Moreover, the false alarms for outliers faded more gradually with distance for lightness levels than for color. This raises the possibility that something like a metrical average is more likely to be computed for lightness than hue. Moreover, it points to intriguing asymmetries between luminance and chromatic processing. Opponency is considered a hallmark of color appearance and there are clear opposing differences between both complementary hues and complementary lightness levels as well as clear physiological substrates identified for lightness increments and decrements (Komban et al., 2014). Yet subjectively, Hering originally considered that light and dark sensations are not mutually exclusive in the way that red versus green or blue versus yellow are (Werner, Cicerone, Kliegl, & DellaRosa, 1984), suggesting that increments and decrements are more likely to be perceived as part of a uniform continuum than two



qualitatively different sensations. The ensemble coding differences between luminance and chromatic contrast are consistent with these subjective impressions, and suggest that the significance and signature of gray – a singularity at the center of color space – may depend importantly on whether the path through it varies in chromaticity or luminance.

## Conclusions

Ensemble percepts of color contrast appear to reflect averaging within rather than across complementary hue categories, suggesting that the mean of ensembles containing very different hues may be inferred only indirectly rather than computed explicitly. This may reflect a representation of color in which different hues are coded as qualitative rather than quantitative differences, for which summary percepts like the mean may depend more on the relative weights of the elements than the actual mean of these weights. In contrast to these results for chromatic contrast, luminance increments and decrements did appear more like quantitative variations that could support direct summary estimates. This suggests that the neutral point of color vision has a very different status for luminance and color. The differences between these dimensions may reflect fundamental differences in the degree and mechanisms of ensemble perception that depend on the nature of the visual representations for different stimulus attributes, and whether these representations allow an explicit averaging or only an indirect estimate of the mean of an ensemble. Similar effects in ensemble coding may occur for many visual attributes beyond color, depending on whether the visual representation is metrical or non-metrical.

**Acknowledgements** Supported by NIH EY-010834 (MW). JM and AF were funded by the European Research Council (COLOURMIND: 772193).

**Open practices statement** None of the data or materials for the experiments reported here are available, and none of the experiments was preregistered.

## References

- Alvarez, G. A. (2011). "Representing multiple objects as an ensemble enhances visual cognition." *Trends in Cognitive Sciences* **15**(3): 122-131.
- Ariely, D. (2001). "Seeing sets: representation by statistical properties." *Psychological Science* **12**(2): 157-162.
- Brown, A. M., D. T. Lindsey and K. M. Guckes (2011). "Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical." *Journal of Vision* **11**: 12-12:11-20.
- Burns, B. and B. E. Shepp (1988). "Dimensional interactions and the structure of psychological space: the representation of hue, saturation, and brightness." *Perception & Psychophysics* **43**(5): 494-507.
- Cha, O. and S. C. Chong (2018). "Perceived Average Orientation Reflects Effective Gist of the Surface." *Psychological Science* **29**(3): 319-327.
- Chetverikov, A., G. Campana and A. Kristjansson (2017). "Representing Color Ensembles." *Psychological Science* **28**(10): 1510-1517.
- Elias, E., M. Dyer and T. D. Sweeny (2017). "Ensemble Perception of Dynamic Emotional Groups." *Psychological Science* **28**(2): 193-203.
- Elias, E. and T. D. Sweeny (2020). "Integration and segmentation conflict during ensemble coding of shape." *Journal of Experimental Psychology. Human Perception and Performance* **46**(6): 593-609.
- Elliott, S. L., M. A. Georgeson and M. A. Webster (2011). "Response normalization and blur adaptation: Data and multi-scale model." *Journal of Vision* **11**(2).
- Emery, K. J., V. J. Volbrecht, D. H. Peterzell and M. A. Webster (2017). "Variations in normal color vision. VI. Factors underlying individual differences in hue scaling and their implications for models of color appearance." *Vision Research* **141**: 51-65.
- Forder, L., X. He and A. Franklin (2017). "Colour categories are reflected in sensory stages of colour perception when stimulus issues are resolved." *PLoS One* **12**(5): e0178097.
- Gilbert, A. L., T. Regier, P. Kay and R. B. Ivry (2006). "Whorf hypothesis is supported in the right visual field but not the left." *Proceedings of the National Academy of Sciences of the United States of America* **103**(2): 489-494.
- Haberman, J., T. Harp and D. Whitney (2009). "Averaging facial expression over time." *Journal of Vision* **9**(11): 1.1-13.
- Haberman, J. and D. Whitney (2007). "Rapid extraction of mean emotion and gender from sets of faces." *Current Biology* **17**(17): R751-753.
- Haberman, J. and D. Whitney (2010). "The visual system discounts emotional deviants when extracting average expression." *Attention, Perception, & Psychophysics* **72**(7): 1825-1838.
- Harnad, S., Ed. (1987). *Psychophysical and cognitive aspects of categorical perception: A critical overview*. Cambridge, Cambridge University Press.
- Juricevic, I. and M. A. Webster (2012). "Selectivity of face aftereffects for expressions and anti-expressions." *Frontiers in Psychology* **3**: 4.
- Kanaya, S., M. J. Hayashi and D. Whitney (2018). "Exaggerated groups: amplification in ensemble coding of temporal and spatial features." *Proceedings of the Biological Sciences* **285**(1879).
- Khayat, N. and S. Hochstein (2019). "Relating categorization to set summary statistics perception." *Attention, Perception, & Psychophysics* **81**(8): 2850-2872.
- Komban, S. J., et al. (2014). "Neuronal and perceptual differences in the temporal processing of darks and lights." *Neuron* **82**(1): 224-234.
- Leib, A. Y., A. Kosovicheva and D. Whitney (2016). "Fast ensemble representations for abstract visual impressions." *Nature Communications* **7**: 13186.
- MacLeod, D. I. and R. M. Boynton (1979). "Chromaticity diagram showing cone excitation by stimuli of equal luminance." *Journal of the Optical Society of America* **69**(8): 1183-1186.
- Martinovic, J., G. V. Paramei and W. J. MacInnes (2020). "Russian blues reveal the limits of language influencing colour discrimination." *Cognition* **201**: 104281.
- Matera, C. N., K. J. Emery, V. J. Volbrecht, K. Vemuri, P. Kay and M. A. Webster (2020). "Comparison of two methods of hue scaling." *Journal of the Optical Society of America. A, Optics, Image Science, and Vision* **37**(4): A44-A54.
- Maule, J. and A. Franklin (2015). "Effects of ensemble complexity and perceptual similarity on rapid averaging of hue." *Journal of Vision* **15**(4): 6.
- Maule, J. and A. Franklin (2016). "Accurate rapid averaging of multihue ensembles is due to a limited capacity subsampling mechanism." *Journal of the Optical Society of America. A, Optics, Image Science, and Vision* **33**(3): A22-29.

- Maule, J., K. Stanworth, E. Pellicano and A. Franklin (2018). "Color Afterimages in Autistic Adults." Journal of Autism and Developmental Disorders **48**(4): 1409-1421.
- Maule, J., C. Witzel and A. Franklin (2014). "Getting the gist of multiple hues: metric and categorical effects on ensemble perception of hue." Journal of the Optical Society of America. A, Optics, Image Science, and Vision **31**(4): A93-102.
- Parkes, L., J. Lund, A. Angelucci, J. A. Solomon and M. Morgan (2001). "Compulsory averaging of crowded orientation signals in human vision." Nature Neuroscience **4**(7): 739-744.
- Rajendran, S. S. and M. A. Webster (2020). "Color variance and achromatic settings." Journal of the Optical Society of America. A, Optics, Image Science, and Vision **37**(4): A89-A96.
- Roberson, D., H. Pak and J. R. Hanley (2008). "Categorical perception of colour in the left and right visual field is verbally mediated: evidence from Korean." Cognition **107**(2): 752-762.
- Sankeralli, M. J. and K. T. Mullen (1997). "Postreceptoral chromatic detection mechanisms revealed by noise masking in three-dimensional cone contrast space." Journal of the Optical Society of America. A, Optics, Image Science, and Vision **14**(10): 2633-2646.
- Siuda-Krzywicka, K., M. Boros, P. Bartolomeo and C. Witzel (2019). "The biological bases of colour categorisation: From goldfish to the human brain." Cortex **118**: 82-106.
- Skelton, A. E., G. Catchpole, J. T. Abbott, J. M. Bosten and A. Franklin (2017). "Biological origins of color categorization." Proceedings of the National Academy of Sciences of the United States of America **114**(21): 5545-5550.
- Skinner, A. L. and C. P. Benton (2010). "Anti-expression aftereffects reveal prototype-referenced coding of facial expressions." Psychological Science **21**(9): 1248-1253.
- Sweeny, T. D., S. Haroz and D. Whitney (2013). "Perceiving group behavior: sensitive ensemble coding mechanisms for biological motion of human crowds." Journal of Experimental Psychology. Human Perception and Performance **39**(2): 329-337.
- Switkes, E. and M. A. Crognale (1999). "Comparison of color and luminance contrast: apples versus oranges?" Vision Research **39**(10): 1823-1831.
- Treisman, A. and H. Schmidt (1982). "Illusory conjunctions in the perception of objects." Cognitive Psychology **14**(1): 107-141.
- Utochkin, I. S. (2015). "Ensemble summary statistics as a basis for rapid visual categorization." Journal of Vision **15**(4): 8.
- Valentine, T., M. B. Lewis and P. J. Hills (2015). Face-space: A unifying concept in face recognition research Quarterly Journal of Experimental Psychology (Hove): 1-24.
- Virtanen, L. S., M. Olkkonen and T. P. Saarela (2020). "Color ensembles: Sampling and averaging spatial hue distributions." Journal of Vision **20**(5): 1.
- Watamaniuk, S. N. and A. Duchon (1992). "The human visual system averages speed information." Vision Research **32**(5): 931-941.
- Webster, J., P. Kay and M. A. Webster (2014). "Perceiving the average hue of color arrays." Journal of the Optical Society of America. A, Optics, Image Science, and Vision **31**(4): A283-292.
- Webster, M. A. (2015). "Visual adaptation." Annual Review of Vision Science **1**: 547-567.
- Webster, M. A. (2020). "The Verriest Lecture: Adventures in blue and yellow." Journal of the Optical Society of America. A, Optics, Image Science, and Vision **37**(4): V1-V14.
- Webster, M. A. and P. Kay (2012). "Color categories and color appearance." Cognition **122**(3): 375-392.
- Webster, M. A. and D. Leonard (2008). "Adaptation and perceptual norms in color vision." Journal of the Optical Society of America A **25**(11): 2817-2825.
- Webster, M. A., E. Miyahara, G. Malkoc and V. E. Raker (2000). "Variations in normal color vision. II. Unique hues." Journal of the Optical Society of America. A, Optics, Image Science, and Vision **17**(9): 1545-1555.
- Werner, J. S., C. M. Cicerone, R. Kliegl and D. DellaRosa (1984). "Spectral efficiency of blackness induction." Journal of the Optical Society of America. A **1**(9): 981-986.
- Whitney, D. and A. Yamanashi Leib (2018). "Ensemble Perception." Annual Review of Psychology **69**: 105-129.
- Winawer, J., N. Witthoft, M. C. Frank, L. Wu, A. R. Wade and L. Boroditsky (2007). "Russian blues reveal effects of language on color discrimination." Proceedings of the National Academy of Sciences of the United States of America **104**(19): 7780-7785.
- Witzel, C. and K. R. Gegenfurtner (2011). "Is there a lateralized category effect for color?" Journal of Vision **11**: 12-16:11-24.
- Witzel, C. and K. R. Gegenfurtner (2013). "Categorical sensitivity to color differences." Journal of Vision **13**(7): 1.
- Witzel, C. and K. R. Gegenfurtner (2018). "Color Perception: Objects, Constancy, and Categories." Annual Review of Vision Science **4**: 475-499.
- Young, A. W., D. Rowland, A. J. Calder, N. L. Etcoff, A. Seth and D. I. Perrett (1997). "Facial expression megamix: tests of dimensional and category accounts of emotion recognition." Cognition **63**(3): 271-313.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.