Vision Research 183 (2021) 1–15

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Color perception and compensation in color deficiencies assessed with hue scaling

Kara J. Emery, Mohana Kuppuswamy Parthasarathy, Daniel S. Joyce, Michael A. Webster *

Graduate Program in Integrative Neuroscience and Department of Psychology, University of Nevada, Reno, Reno, NV 89557, United States

ARTICLE INFO

Keywords:
Color vision
Color deficiency
Hue scaling
Adaptation
Compensation

ABSTRACT

Anomalous trichromats have three classes of cone receptors but with smaller separation in the spectral sensitivities of their longer-wave (L or M) cones compared to normal trichromats. As a result, the differences in the responses of the longer-wave cones are smaller, resulting in a weaker input to opponent mechanisms that compare the LvsM responses. Despite this, previous studies have found that their color percepts are more similar to normal trichromats than the smaller LvsM differences predict, suggesting that post-receptoral processes might amplify their responses to compensate for the weaker opponent inputs. We evaluated the degree and form of compensation using a hue-scaling task, in which the appearance of different hues is described by the perceived proportions of red-green or blue-yellow primary colors. The scaling functions were modeled to estimate the relative salience of the red-green to blue-yellow components. The red-green amplitudes of the 10 anomalous observers were 1.5 times weaker than for a group of 26 normal controls. However, their relative sensitivity at threshold for detecting LvsM chromatic contrast was on average 6 times higher, consistent with a 4-fold gain in the suprathreshold hue-scaling responses. Within-observer variability in the settings was similar for the two groups, suggesting that the suprathreshold gain did not similarly amplify the noise, at least for the dimension of hue. While the compensation was pronounced it was nevertheless partial, and anomalous observers differed systematically from the controls in the shapes of the hue-scaling functions and the corresponding loci of their color categories. Factor analyses further revealed different patterns of individual differences between the groups. We discuss the implications of these results for understanding both the processes of compensation for a color deficiency and the limits of these processes.

1. Introduction

Normal human color vision is trichromatic because it is based on sampling the light spectrum with three classes of cone receptors with photopigments maximally sensitive to short, medium, or long wavelengths (S, M, L). However, inherited forms of color deficiencies are common and in most cases result from alterations of the genes affecting the L and M photopigment opsins (Neitz & Neitz, 2011). In some individuals the opsin is not expressed leading to a loss of one cone class, or dichromacy. More commonly, the opsin is expressed but the absorption spectrum is shifted (anomalous trichromacy) so that it is closer to the normal M cone (protanomalous) or L cone (deuteranomalous). As an X-linked recessive trait, these forms of color deficiency are rare in females but affect about 8% of Caucasian males, with roughly 2% dichromats and 6% anomalous trichromats.

In anomalous trichromats (AT), the color losses are a consequence of the weaker difference signal between their longer wave pigments. In the normal trichromat (NT) the peaks of the L and M cones are separated by ~30 nm (Stockman & Sharpe, 2000). Their relative outputs are compared in post-receptoral channels that receive opposing inputs (excitation or inhibition) from the two cone classes, and this LvsM comparison is one of the cardinal directions along which color is encoded by neurons of the parvocellular pathway (Derrington, Krauskopf & Lennie, 1984). A second comparison involves signals in the S cones opposed by the L and M cones (SvsLM), and represents a second cardinal dimension of precortical color coding carried by the koniocellular pathway (Martin, White, Goodchild, Wilder & Sefton, 1997). In anomalous trichromats the differences between their longer wave pigments can be as little as 2 to 12 nm (Neitz & Neitz, 2011). This leaves sensitivity in the SvsLM pathway relatively intact, but reduces the chromatic signals available to the LvsM channel, roughly in inverse proportion to the spectral separation (Boehm, MacLeod, & Bosten, 2014). This altered

* Corresponding author.
E-mail address: mwebster@unr.edu (M.A. Webster).

https://doi.org/10.1016/j.visres.2021.01.006
Received 10 September 2020; Received in revised form 7 December 2020; Accepted 14 January 2021
Available online 23 February 2021

input is manifest as altered color matching and could result in weaker threshold color discrimination. However, the relationship between chromatic sensitivity and the observer's cone spectral sensitivities can be highly variable (Bosten, 2019), and the peak separation can be influenced by a number of other factors including potential differences in the optical densities of the different photopigments (Neitz, Neitz, He & Shevell, 1999; Thomas, Formankiewicz & Mollon, 2011).

Moreover, several lines of evidence suggest that the color experience of anomalous observers is not as impoverished as their altered threshold sensitivities predict. When tested with larger chromatic differences, the salience (Regan & Mollon, 1997), apparent contrast (Knoblauch, Marsh-Armstrong & Werner, 2020), and phenomenal differences (Boehm, MacLeod, & Bosten, 2014) in LvsM stimuli are stronger relative to normal trichromats or to other stimulus dimensions (e.g SvsLM or luminance differences) than their relative sensitivity at threshold. These effects are again variable across individuals, but suggest that in general post-receptoral mechanisms can compensate for the reduced spectral separation of the longer-wave cones. One potential form of this compensation would be to increase the gain of the neural responses within the LvsM mechanisms (MacLeod, 2003; Regan & Mollon, 1997; Webster, Juricevic & McDermott, 2010b). That is, the weaker cone-opponent inputs could be discounted simply by amplifying the neuron's output. This would preserve coding efficiency by matching the neuron's responses to the range of its inputs, and in principle could undo the weakened input to the opponent mechanism. However this could only correct for threshold sensitivity losses if the amplification occurred before the site of the noise limiting the thresholds.

Compensation for a color loss could also reflect post-perceptual or cognitive strategies. For example, color deficient observers could learn to estimate the magnitude of color contrasts relative to the range they habitually experience. One indication of these cognitive strategies is in color naming. Even dichromats use a wide range of color terms to describe colors including red and green, and the way they apply these can often again be much more similar to normal trichromats than expected from their color losses (Bonnardel, 2006; Broackes, 2010; Jameson & Hurvich, 1978; Uchikawa, 2008; Wachtler, Dohrmann & Hertel, 2004). A number of sensory cues could contribute to this behavior, including exploiting information from rod receptors, retinal inhomogeneities, neural nonlinearities, or lightness differences. However, in such cases the compensation arguably involves learning to associate the color terms with sensory signals that differ from the signals utilized by trichromatic observers.

The potential for multiple forms of compensation make it important and intriguing to ask how and to what extent compensation for a color deficiency is manifest in different contexts and tasks. In the present study, we explored color responses in a hue-scaling task, in which observers report their color experiences by decomposing the stimulus into the perceived proportions of primary colors red versus green and blue versus yellow. This is a common technique for measuring color appearance, and has been used by a number of previous studies to examine color perception in both normal and color- deficient observers (Abramov & Gordon, 1994; Jameson & Hurvich, 1959). For the latter the responses point to greater variability both within and between individuals for judging the hues and also systematic shifts in the categories or perceived strength of different hues (Hemminger & Georgi, 1982; Nagy, Nemeth, Samu & Abraham, 2014; Smith, Pokorny & Swartley, 1973). For the most part, these effects have not been evaluated within the specific context of sensory compensation. However, some authors have noted that hue percepts in these observers could be at least partially normalized to color signals in the environment in the same way as normal trichromats, and thus could partially discount the sensitivity differences (Neitz, Carroll, Yamauchi, Neitz & Williams, 2002; Pokorny & Smith, 1977).

Typically hue-scaling responses are assumed to directly characterize the red-green and blue-yellow color-opponent processes mediating color appearance (Abramov & Gordon, 1994). However, as a subjective task involving breaking down a hue into its component sensations, the judgments potentially reflect high level and abstract stages of color perception; and some aspects of the responses – such as how categorical the observer is – can depend on how the task is structured (Matera, Emery, Volbrecht, Vemuri, Kay & Webster, 2020). Hue scaling could therefore allow for both perceptual and conceptual modes of compensation. Our primary aim was to assess to what extent compensation can undo the reduced difference signals in the receptors in suprathreshold color appearance judgments like hue scaling.

We were also interested in testing hue scaling to further understand the bases for individual differences in color appearance. Normal variations in color appearance tend to show little association with inter-observer variations in spectral sensitivity (Emery & Webster, 2019; Webster, 2020). In previous work, we examined hue-scaling differences for normal trichromatic observers to explore the sources of variation underlying these differences (Emery, Volbrecht, Peterzell & Webster, 2017a, 2017b). Even though observers were explicitly instructed to judge the colors in terms of red-green and blue-yellow dimensions, the inter-observer differences provided little evidence that these dimensions were unique or yoked as opponent mechanisms. Instead, ratings for both the primary hues and binary hues (putative mixtures of the primaries like orange or purple) varied independently, suggesting they were controlled by separate factors, such as how observers learned to classify the stimuli corresponding to different color categories. A second aim of the present study was to extend these measurements to anomalous observers to explore whether similar patterns of variation occurred for their color appearance judgments.

2. Methods

Participants. Ten anomalous observers participated in the experiment. All were young-adult college students and all were male. Color vision was assessed with the Cambridge Colour Test (Cambridge Research Systems) and settings on a clinical anomaloscope (OCULUS Inc.). Participants were diagnosed as anomalous trichromats based on standardized criteria and scores on the anomaloscope and included 7 deuteranomalous (one extreme) and 3 protanomalous individuals. Comparison data for normal trichromats were taken from a previous study of 26 observers (also young-adult students, 16 female) (Emery et al., 2017a). These observers were also screened with the Cambridge Colour Test, and were tested using the same stimuli, procedures, and equipment as in the current study. All observers were recruited from the University of Nevada, Reno and participated with informed consent. All procedures were conducted following protocols approved by the university's IRB and consistent with the Declaration of Helsinki.

Stimuli. Stimuli were based on a version of the MacLeod-Boynton chromaticity diagram (MacLeod & Boynton, 1979) scaled so that nominal units along the LvsM and SvsLM axes corresponded approximately to multiples of detection threshold based on a prior study (Webster, Miyahara, Malkoc & Raker, 2000a), and were centered on a white point equivalent to Illuminant C. Units in the scaled space are related to the MacLeod-Boynton coordinates by:

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

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

Where LvsM and SvsLM are coordinates used in the study; l_{mb} and s_{mb} are the corresponding MB coordinates; 0.6568, 0.01825 are the MB coordinates of the white point, and 2754, 4099 are the scaling factors for stimulus contrasts. We note that with this nominal scaling, thresholds measured for the conditions of the current study were roughly 2 times higher for the LvsM axis (see below).

The set of tested stimuli corresponded to 36 chromaticities lying on a circle centered on the white point and with a radius of 60 contrast units. The 36 colors spanned the circle in 10° steps. We refer to these stimuli throughout as the "*stimulus angles*" (Fig. 1a). The stimuli had a constant luminance of 20 cd/m^2 as defined photometrically and were shown as a uniform 2° square centered on an 11.3° by 8.5° uniform gray

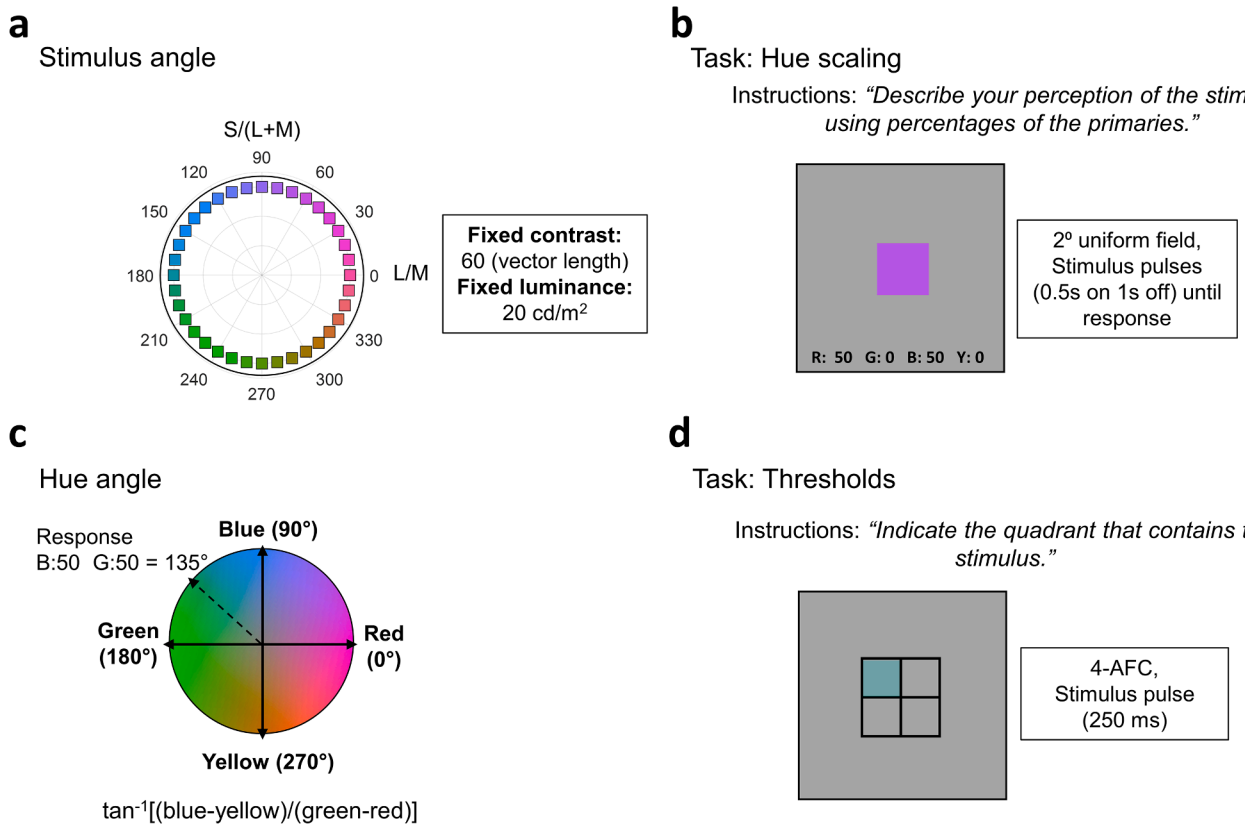

Fig. 1. a) Stimuli in the LvsM and SvsLM space; b) example of the hue-scaling task for a stimulus rated as 50% red and 50% blue; c) specification of hue angles in the perceptual red-green and blue-yellow plane; d) display arrangement for the 4AFC detection threshold task. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

background with the same luminance and mean chromaticity. The display was a SONY Multiscan 500PS Trinitron CRT monitor controlled by a ViSaGe Stimulus Generator (Cambridge Research Systems). The monitor was calibrated and gamma-corrected based on measurements with a PR 655 spectroradiometer (Photo Research). The display was viewed binocularly from a distance of 200 cm in an otherwise dark room.

3. Procedure

Hue scaling. In hue scaling, the observer is asked to judge the proportions of red, green, blue, or yellow in the stimulus. For example, an "orange" hue that appeared as a balanced mixture of red and yellow would be scaled 50% red and 50% yellow. Note these relative measurements provide information about the "hue" of the stimulus (i.e. the relative amounts of red-green or blue-yellow in the color) but do not provide information about the "saturation" (i.e. the absolute strength of the color). In our case, responses were made by varying the percentages displayed along the bottom of the screen for the four primaries (Fig. 1b), following procedures we used in a previous study (Emery et al., 2017a). At the start of each trial the values for the four primaries were set to zero and the participant then adjusted the values with a handheld keypad in 5% steps. Participants were told that the responses were constrained so that the total had to equal 100% and so that proportions could not be assigned to both red and green or to both blue and yellow. While making these adjustments, a stimulus angle was shown on the screen as a repeating pulse of 500 ms alternated with 1000 ms of the gray field. There was no time restriction on making the settings. Once the subject selected the proportions, they pressed a key to record the response and the next stimulus angle was displayed.

In a single session, responses were made for all 36 stimulus angles shown in random order, and then repeated after an intermission. The first session also included a brief set of practice settings to familiarize them with the task. Each observer participated in two sessions on different days, and thus made 4 settings for each stimulus. As a test of overall consistency, observers with a standard deviation in their repeated settings that averaged more than three times the mean standard deviation for the group were to be excluded from the analysis. However, no observers met this criterion for the NT nor the AT group and thus all were included. As a test for consistency in the settings for each stimulus, we also evaluated deviant trials by stimulus angle for each observer. We used a jackknife technique to detect outlying trials by computing the mean and standard deviation for each possible group of three trials to determine whether, for the cluster with the smallest standard deviation, the remaining trial was more than three standard deviations away from the group mean. In this case, this fourth trial would be excluded from subsequent analyses. This criterion resulted in omitting 13% of all trials for the anomalous trichromats and 14% for the normal trichromats. Results shown are based on the mean of the resulting settings.

To represent the responses, the proportions were converted into an angle in a perceptual red-green vs. blue-yellow opponent space (Fig. 1c) based on the following equation

$$\text{hue angle} = tan^{-1}[(\text{blue} - \text{yellow})/(\text{red} - \text{green})]$$

We use the term "*hue angle*" to refer to the perceived color of the stimulus angles. Within this perceptual space the red-green axis corresponds to hue angles of 0° and 180°, while the blue-yellow axis corresponds to 90° and 270°, and binary hues would lie between these axes. For example, a purple that was judged to have an equal proportion of red and blue (50:50) would have a hue angle of 45°.

Thresholds. In a separate set of measurements, chromatic contrast thresholds were also collected for 9 of the 10 observers. (One participant (DA-06) was unavailable to complete these settings.) In this case the 2°

field was divided into four 1° quadrants with narrow black lines (Fig. 1d). A 250 ms chromatic pulse was shown in one of the quadrants, and the observer used the keypad to indicate the location. The contrast was varied in 0.1-log unit steps in a staircase with 23 reversals, and the threshold was taken as the mean of the final 20 reversals. Thresholds were measured in counterbalanced order for stimuli varying along the two poles of the LvsM and SvsLM axes, with two settings made for each. Values shown are based on the average of the 4 settings for each axis.

4. Results

Thresholds. Fig. 2a compares the ratio of the contrast thresholds along the LvsM/SvsLM axes for the 9 anomalous trichromat (AT) observers tested in this task, along with representative threshold ratios for 9 of the normal trichromat (NT) observers. The AT observers were on average 12.0 times less sensitive to LvsM contrast compared to their thresholds for SvsLM contrast. For the NT observers the LvsM/SvsLM ratio was instead 1.94. Thus the relative sensitivity to the LvsM contrast for the AT group was 6.15 times lower than the NT group, a difference that was highly significant ($t(8) = 5.29$, $p = 0.0007$). However, there were large differences across the AT observers, with relative sensitivity losses ranging from 2.6 to 12.2 (sd = 2.95). Fig. 2b shows the anomaloscope matching ranges for the 10 AT observers, including one observer identified as an extreme deuteranomalous (DA-06) based on his luminance match. The anomaloscope matching ranges for the observers were not correlated with their threshold ratios.

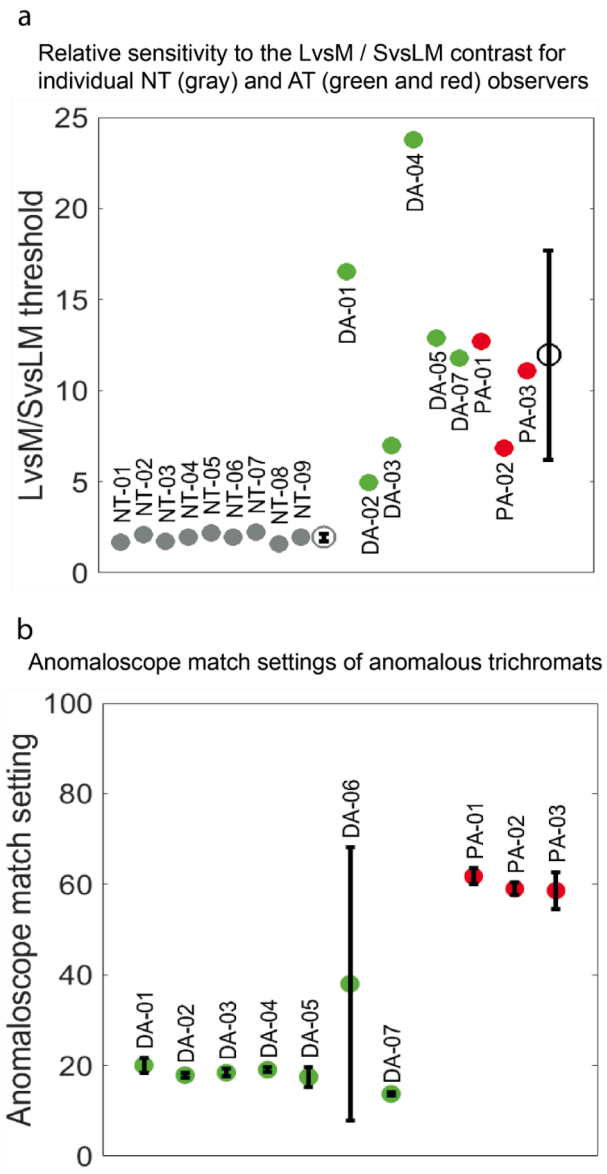

Fig. 2. a) Relative sensitivity to the LvsM / SvsLM contrast for individual color normal (NT, gray) and AT observers (DA green symbols and PA red symbols). Points show the ratio of the LvsM to SvsLM thresholds. Large symbols show the mean of the settings for each group and error bars indicate ± 1 standard deviation. b) Anomaloscope match ranges for DA (green) and PA (red) observers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Average hue-scaling functions. Fig. 3 shows the average hue-scaling functions for the anomalous and normal groups. In the two upper panels, these are shown as the mean values for the red and green (Fig. 3a), and blue and yellow (Fig. 3b) response functions. Specifically, the values show the average proportion of each of the four primaries reported for each hue angle. Fig. 3c instead plots the mean scaling functions as represented by the hue angles. In this case, the ordinate values represent the "hue angle" in the red-green vs. blue-yellow space, with the angles corresponding to pure red, blue, green, and yellow indicated by the horizontal lines (at 0°, 90°, 180°, and 270° respectively). Values for the abscissa instead correspond to the "stimulus angle" in the LvsM and SvsLM space. As noted by many previous studies e.g. (Krauskopf, Williams & Heeley, 1982; Webster, Miyahara, Malkoc & Raker, 2000b; Wuerger, Atkinson & Cropper, 2005), the unique hues are not aligned with the LvsM and SvsLM axes, and this is reflected in the warped relationship between the hue angles and the stimulus angles seen in the curves (Emery et al., 2017a).

These average curves belie large individual differences within both groups, which we consider in detail below. Nevertheless, there are also overall differences between the AT and NT observers. Because some of the hue angles were not normally distributed, to examine these we used Mann-Whitney U tests to assess the group differences in the hue-scaling functions at each stimulus angle (with Bonferroni correction for multiple comparisons). This revealed significant differences in the hue responses at a range of stimulus angles in the purple (30, 60, 70, 90 deg) and yellow-green (250–290 deg) regions, and these differences are indicated by the asterisks in Fig. 3c. These overall effects reflect a tendency for the range of blue and yellow responses to be expanded in the AT observers with a concomitant contraction in the range for red and green responses. This change in the relative weighting of the blue-yellow and red-green responses is consistent with previous measures of chromatic response functions based on a hue cancellation task (Romeskie, 1978). The weaker red-green responses biased the corresponding hue angles for the AT observers toward the pure blue (90°) and pure yellow (270°) hue axis (Fig. 3, bottom left), so that their average function appears more step-like.

Finally, Fig. 3d indicates the reliability of the observers' settings, based on the standard deviation of the settings averaged across the observers in each group. These values are shown after applying an arcsine transform to correct for differences in variance for proportional measures (Gordon, Abramov & Chan, 1994). For the majority of stimulus angles the within-observer variability is similar for the two groups. This was confirmed by Mann-Whitney comparisons of the relative standard deviations for the AT and NT observers (again corrected for multiple comparisons). In this case no significant differences were found.

Individual hue-scaling functions. The individual hue angle functions for the 10 AT observers are shown in Fig. 4, with each compared to the average NT response. These illustrate the wide variety of responses among the AT observers, including one participant (PA-03) who did not use the red primary for any of the stimulus angles. His hue angles thus ranged only from unique blue through green to unique yellow and back. The remaining observers had largely monotonic functions (with the exception of the extreme AT DA-06), and varied in how strongly biased the percepts were toward the blue and yellow hue axes. Some observers showed strong categorical biases (e.g. DA-04, DA-07, PA-01) and thus

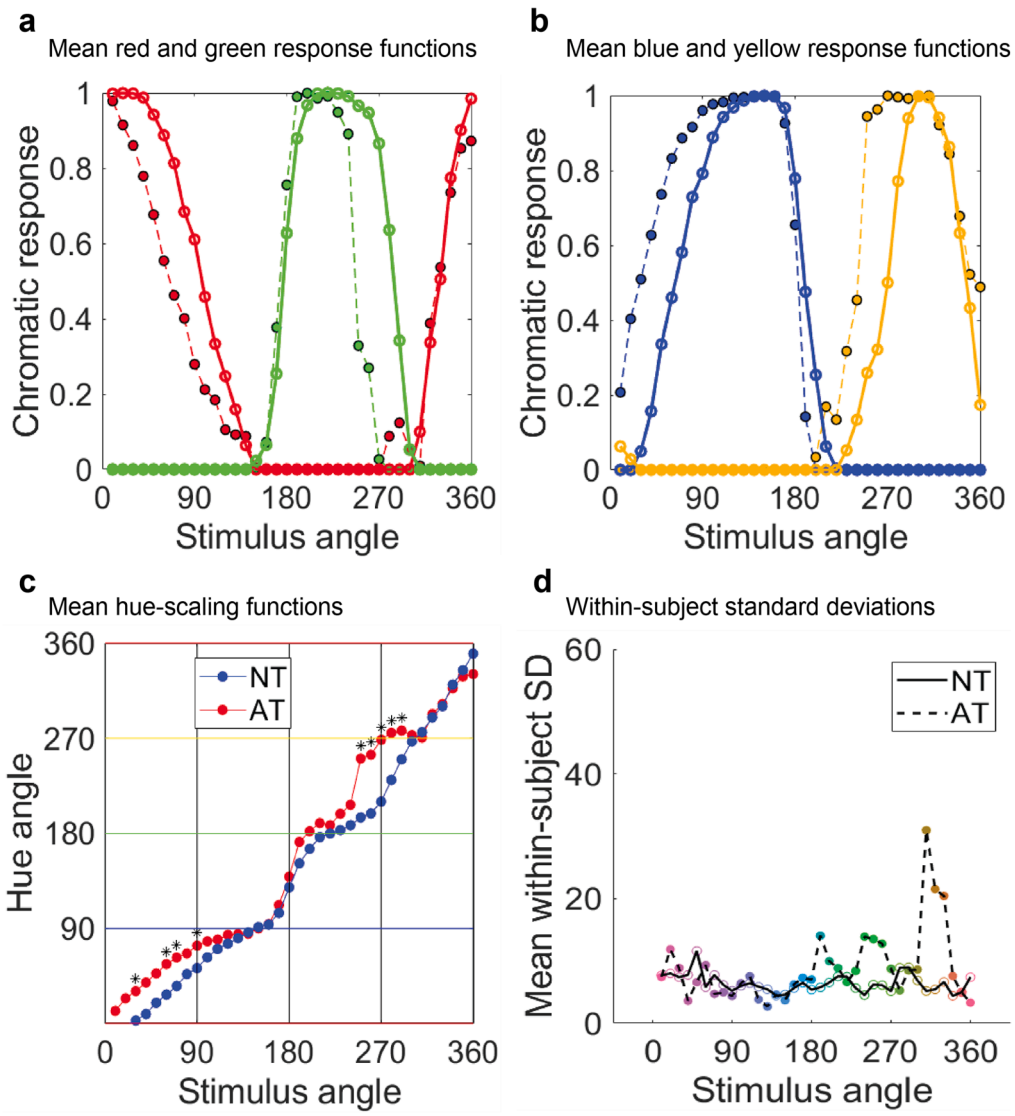

Fig. 3. a and b) Mean hue-scaling functions for the NT (solid) and AT (dashed) observers for the red-green (a) and blue-yellow (b) responses. Curves show the proportion of each primary hue in the setting as a function of stimulus angle; c) corresponding mean hue-scaling functions for NT (blue) and AT (red) observers. Asterisks indicate stimulus angles at which there were significant differences between the AT and NT observers; d) average within-observer standard deviation in the 4 repeated settings for NT (solid) and AT (dashed) as a function of stimulus angle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

more step-like functions, while for others the curves came close to approximating the average curve for NTs (DA-03, DA-05, PA-02).

Unique and binary hues. From the hue-scaling functions for both groups we estimated the stimulus angles that corresponded to the unique hues (red, green, blue, and yellow) or balanced binary hues (orange, purple, yellow-green or blue green). Note these correspond to hue angles at 45° intervals in the plots of Fig. 4. For each individual, these were estimated from linear interpolation between the points straddling each hue line, and thus are affected by variability in these local estimates. (We also tried smoothing the curves with an 8th order polynomial and then estimating the hue loci from the function (Emery et al., 2017a). This produced more compact estimates of the loci, but they deviated more from the specific individual settings. However, the category loci estimated by this procedure remained broadly similar to those estimated by the linear interpolation). The loci for individual observers are shown in Fig. 5. Despite the pronounced inter-observer variability, the mean hue loci were significantly different ($p < 0.002$) for the two groups for all unique and binary hues except orange and purple ($p = 0.272$ and 0.165 respectively; based on Bonferroni corrected values for multiple *t*-test comparisons).

Modelled differences in the hue-scaling functions. To further quantify the differences in the hue-scaling functions, we fit the functions for individual observers with a model that approximated the proportional responses for the four primaries as half-rectified cosine functions. Importantly, this characterization is not intended as a mechanistic account of hue scaling, because as noted individual differences in the scaling function cannot be explained by underlying red-green and blue-yellow opponent responses (Emery et al., 2017a), and the predicted functions do not incorporate known nonlinearities in color appearance e.g. (Knoblauch & Shevell, 2001; Larimer, Krantz & Cicerone, 1975). Instead, the model is merely descriptive and was used to approximate the relative salience of red-green and blue-yellow dimensions in the observers' reported percepts, so that we could compare these values across observers.

To fit the functions, the width or period of the cosine responses were allowed to vary with the constraint that the blue vs. yellow (and the red vs. green) functions together spanned the full 360° of the stimulus plane. For example, if the blue function subtended (180 + x)° then the yellow function was narrowed to (180–x)°. This fixed the blue and yellow (or red and green) peaks to be 180° apart, with the absolute phase also varied to provide a least-squares fit to the observed responses. The resulting curves for the mean NT and AT responses are shown in the upper four panels of Fig. 6, with 3 parameters each for fitting the red-green or blue-yellow responses. Previous studies have specifically addressed the fits of cosine functions to hue scaling and found that these are too broad to characterize the responses, and that better fits were

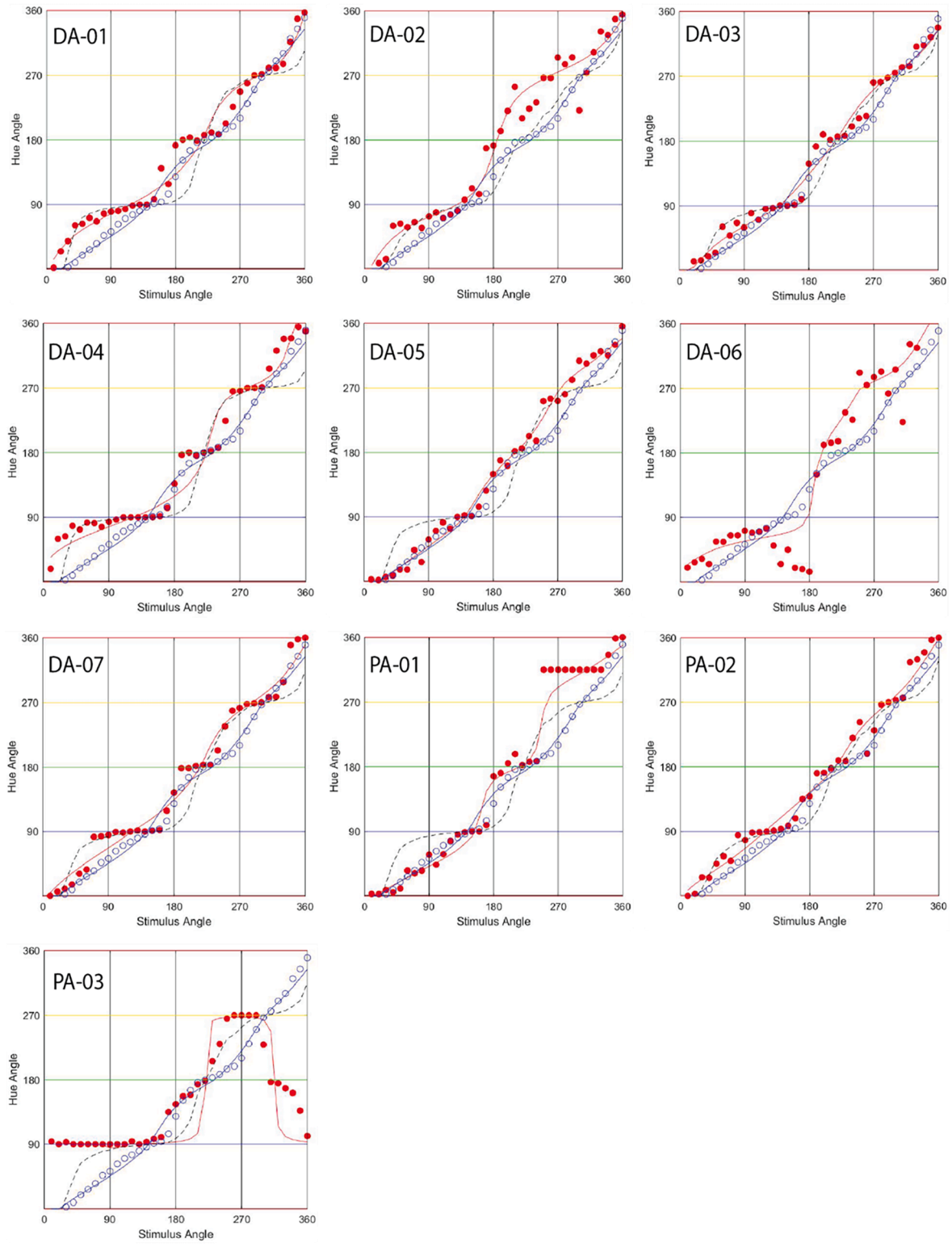

Fig. 4. Individual hue-scaling functions for the 10 AT observers. red filled circles: measured hue angles; blue circles: mean NT hue angles; red and blue lines: model fits to the AT and NT functions; dashed line: hue-scaling function predicted from relative threshold sensitivities. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

instead obtained when the function was raised to a power greater than one, which was interpreted as an expansive nonlinearity in the contrast response (De Valois, De Valois, Switkes & Mahon, 1997). However, this narrowing was not readily apparent in the responses for our hue-scaling functions for the NT observers, and our fits were therefore restricted to linear cosines.

To further fit the hue-scaling angles, we additionally varied the heights of the primary responses. Since the proportions only give the relative heights, the peak values for blue and yellow were fixed at 1.0, and then the amplitude of red and green values were independently varied as two further free parameters. The resulting estimates for the mean functions are given in the lower left panel of Fig. 6. Finally, the

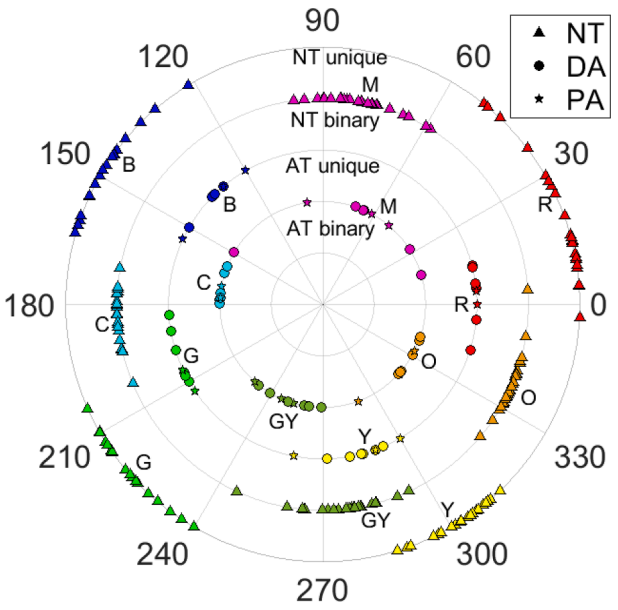

Fig. 5. Loci of the unique and binary hues for the individual NT (triangles) and AT (DA – circles, PA - stars) observers, based on linear interpolation of the hue-scaling functions. (R-Red, G- Green, B- Blue, Y- Yellow, M-Magenta, GY – Green-yellow, C-Cyan, O-Orange). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

best-fitting values for the parameters are illustrated in the polar plot in the lower right panel of Fig. 6. For the NT responses, the blue-yellow responses had a mean cosine phase of 127.5° with a broader period for blue (1.19). The average red-green responses peaked at an angle of 43.2° and were slightly more symmetric (1.11). Finally, to fit the hue angles, the red and green amplitudes had to be increased by 1.11 and 1.29 respectively relative to the unit blue-yellow amplitudes. For the average of the AT observers, the response phases were shifted to 106.1° for the blue-yellow axis and 32.8° for the red-green axis, and the amplitude of the red and green responses was reduced to 0.77 and 0.83, respectively.

The fits to each AT's hue-scaling function are shown by the solid lines in Fig. 4. The predicted curves provided a reasonable approximation to the individual hue-scaling functions, with an rms error of 13.2° for the AT observers and 16.0° for the NT observers. This allowed us to compare the parameter values for the fits across the two groups, which are summarized in Fig. 7. T-tests of the parameters (corrected for multiple comparisons) confirmed that the AT observers had significantly lower amplitudes for both red ($t(34) = 3.37$, $p = 0.002$) and green ($t(34) = 4.11$, $p < 0.001$), and these amplitudes were correlated with each other within both groups (AT: $r = 0.71$, $p < 0.001$; NT: $r = 0.56$, $p = 0.003$). On the other hand, the modeled red and green values were not correlated with the observers' anomaloscope match ranges. A further difference was that the phases for the blue-yellow responses were rotated 21.4° closer to the SvsLM axis for the AT observers and thus differed from the NT phases ($t(34) = 5.95$, $p < 0.001$). In turn, the AT red-green phases were also rotated toward the LvsM axis, though this difference was only significant when observer PA-03 – who again did not use "red" responses – was excluded from the comparison (17.4° rotation, ($t(33) = 8.74$, $p < 0.001$)).

As a second test for these group differences, we used a *k*-means cluster analysis to classify each hue-scaling function into one of two groups to determine whether the two groups tended to separate into ATs and NTs. This analysis resulted in two clusters of sizes 11 and 25, successfully classifying all of the AT observers into one group along with one misclassified NT, and 25 of the 26 NT observers in the second group. Notably, based on the model fits, the misclassified NT observer also had the lowest red and green amplitudes (0.56 and 0.48) among the NT observers, with values closer to the range of the anomalous observers (mean amplitude ± sd: red = 0.77 ± 0.39; green = 0.83 ± 0.31). Given this, it is possible that this individual was misclassified as a normal trichromat in the original study (Emery et al., 2017a), and was instead correctly identified as an anomalous trichromat based on their hue scaling. However, this observer's CCT results did not indicate a color deficiency, as this was a required screening for participation in the previous study (Emery et al., 2017a).

Hue scaling vs. *threshold sensitivity*. While the red-green amplitudes were conspicuously lower for the AT observers, they were nevertheless substantially higher than the amplitudes predicted by their threshold sensitivities. The predictions for this loss are shown by the dashed lines in Fig. 4. For these predictions we assumed that losses in LvsM sensitivity would primarily impact the red-green responses. We therefore recalculated the average NT scaling function after reducing the red and green responses by the observer's relative threshold elevation. In general, the threshold losses again predict stronger biases toward the blue-yellow axes than implied by the observed hue-scaling functions. Recall that these were on average 6.2 times higher than the mean for NT observers, equivalent to an effective chromatic contrast of 1/6.2 or 16% of the contrast for the NT group. Alternatively, the mean of the hue-scaling amplitudes was reduced by only 0.80/1.20 or 66% of the NT value, suggesting an average multiplicative gain of 4.1 in the suprathreshold responses. Notably, however, neither the red nor green response amplitudes were correlated with the individual AT's threshold losses (red: $r = -0.18$, NS; green: $r = 0.16$, NS).

Factor analyses of the hue-scaling functions. As noted, we previously examined the underlying patterns of individual variation in the hue-scaling functions for the NT group by using factor analysis, which extracts latent sources of variation in the observed measurements from their covariance structure (Emery et al., 2017a). The general assumption is that if two variables are correlated (e.g. the scaling responses for two adjacent stimulus angles) then this is because they measure the influence of a common underlying factor; while settings that are uncorrelated (e.g. the scaling responses for very different stimulus angles) reflect the separate influences of different factors. Factor analysis thus partitions the variance across the variables into factors that may reflect the actual sources of the individual differences in the observed measurements (Mollon, Bosten, Peterzell & Webster, 2017). In our previous studies, we found that this required 7–8 systematic factors to account for the individual differences among the normal trichromat observers, and that these factors were each narrowly tuned for different stimulus angles (i.e. each accounted for variability in only a limited range of nearby angles).

Fig. 8a shows the hue-scaling functions for the AT and NT observers and the individual differences present in these functions for each group. To validate that these individual differences were significant (Mollon et al., 2017), we assessed test–retest reliability and compared the magnitude of between- and within- observer variability. Reliability for each group was assessed separately by comparing responses between the two sessions of hue scaling completed by each observer on different days. We first de-trended the scaling functions by representing each observer's response relative to the mean response for each session. Spearman's rank correlations indicated responses between sessions were significantly correlated for the AT observers ($\rho(358) = 0.69$, $p < 0.001$) and for the NT observers ($\rho(934) = 0.41$, $p < 0.001$), confirming test–retest reliability. To compare the magnitude of the between- and within- observer variability, we first arcsine-transformed the scaling functions to normalize the variance. One-sided Mann-Whitney U tests confirmed that standard deviations across mean settings between observers were larger than standard deviations across trials within an observer for both the AT ($Z(10) = 6.75$, $p < 0.001$) and NT ($Z(10) = 7.19$, $p < 0.001$) groups (Fig. 8 middle panels). Additionally, the average relative magnitude of between/within variability was 3.40 and 2.34 for AT and NT observers, respectively. These results suggest that the individual differences represent true variability between the observers and thus are not simply due to an artifact such as measurement noise.

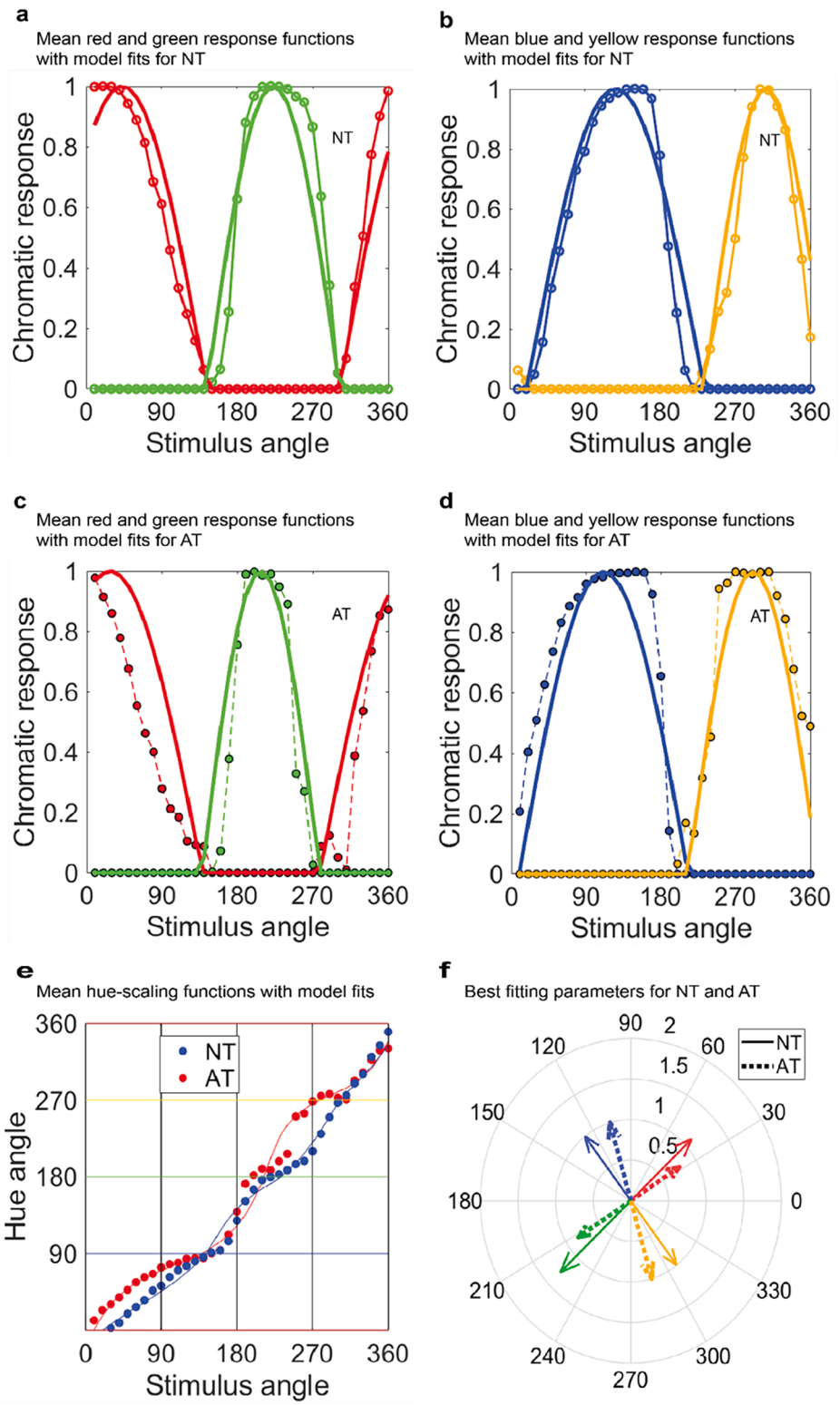

Fig. 6. Modeled hue-scaling functions. a and b: fits of half-rectified cosine functions (solid lines) to the mean proportions of the red-green (a) and blue yellow (b) NT functions (solid lines / circles); c and d: corresponding fits to the mean AT functions; e) fits to the hue angles based on scaling the amplitude of the red and green responses for the mean AT (red) and NT (blue) hue-scaling functions; f) representation of the phase (vector direction) and amplitude (vector length) of the best fitting cosine functions for the average NT (solid) and AT (dashed) hue-scaling functions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We then explored whether the pattern of variation was different between the AT and NT observers. Some suggestion of this is evident simply from visual inspection of the correlation matrices (Fig. 8 right panels). The bottom right panel shows the correlations across the 36 stimulus angles for the 10 AT observers. Strong positive correlations (blue) were primarily limited to nearby angles, along the negative diagonal. However, there is also a clear band of negative correlations (red) along the positive diagonal, which in some cases corresponds to widely separated stimulus angles. This contrasts with the pattern we found previously for the NT observers (Emery et al., 2017a) which was instead indicative of multiple, narrowly-tuned factors, with no clear association to cardinal or diagonal axes. This is shown in the upper right panel for the correlations based on a subset of 10 pseudo-randomly chosen NT participants (sub-sampled so that the sample size was the same between the AT and NT groups). In this case the positive correlations are restricted largely to nearest neighbors and the negative correlations are

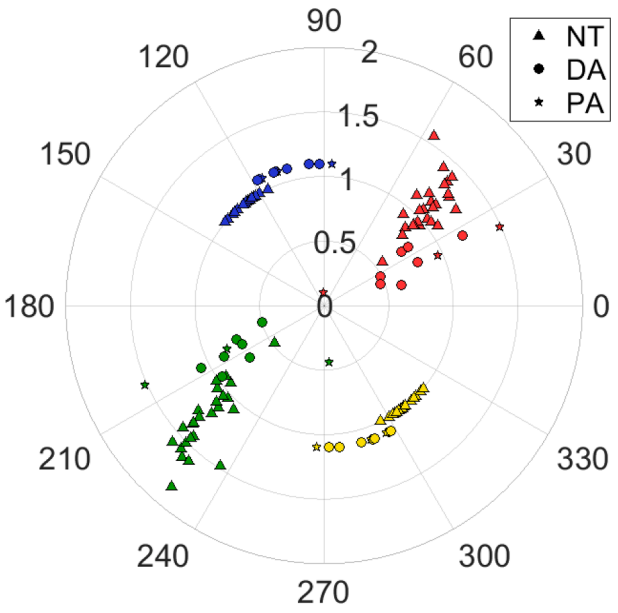

Fig. 7. Polar plot of the best fitting phases and amplitudes to the individual NT (triangles) and AT (DA – circles, PA - stars) observers.

less pronounced or systematic. We verified that a similar pattern emerged for different groups of 10 NT observers.

For the factor analyses, we followed procedures detailed in our previous studies (Emery et al., 2017a; Matera et al., 2020). Briefly, initial factors were extracted using principal components because the datasets deviated from multivariate normality (i.e. the statistical distances were not chi-square distributed according to a one-sample Kolmogorov-Smirnov test (AT and NT: D = 1, $p < 0.001$)). Maximum likelihood solutions could not be calculated due to singular correlation matrices for both datasets. The factors were rotated using the Varimax criterion, which forces the factors to be orthogonal and favors a solution with factors that have high loadings (squared correlations) for few variables and low loadings for the remaining variables. We determined orthogonal solutions as appropriate for each dataset given that the oblique and orthogonal solutions were correlated (AT: $\rho(178) = 0.56$, $p < 0.001$; NT: $\rho(142) = 0.98$, $p < 0.001$). To determine the number of factors to extract, we used the criterion that each factor was 'systematic' such that it had high loadings on two or more consecutive variables, indicative of systematic tuning rather than random variability in the measurements (Webster & MacLeod, 1988). The criteria for systematic loadings were determined by a bootstrapping procedure that assessed when the values for adjacent loadings were unlikely to emerge by chance by determining thresholds for both the minimum loading and number of consecutive loadings. Factors with at least 2 consecutive loadings higher than 0.81 were determined as systematic for the AT dataset ($p = 0.024$). By these criteria, we found a solution of five systematic factors for the AT dataset that accounted for 92% of the total variance. A six-factor solution was calculated and visualized to confirm that this additional factor was not systematic. For the NT observers, factors with at least 3 consecutive loadings higher than 0.62 were determined as systematic ($p = 0.03$). By these criteria, we found a solution of four systematic factors for the NT dataset that accounted for 72% of the total variance. A five-factor solution was calculated and visualized to confirm that this additional factor was not systematic.

Fig. 9 shows the loadings for these factors for each of the datasets, and illustrates a number of differences in the factors for the two groups. First as noted, a smaller amount of the total variance was accounted for in the NT dataset, and this is true even for the five-factor NT solution (matching the number of factors for the AT solution) which accounts for 80% of the total variance (compared to 92% for the AT dataset). Notably, this was similar to the result for the full NT set, where we previously found that the first 7 factors accounted for 77% of the variance (Emery et al., 2017a). Second, in the present analysis the factors for the NT group are noisy because of the small number of observers (n = 10) and do not reveal a readily interpretable pattern. In particular, the loadings for these observers do not reveal the multiple, narrowly-tuned factors that are well-defined for the full sample (n = 26) and which we have confirmed for independent samples and testing procedures for similar larger populations (Emery et al., 2017a; Matera et al., 2020). In contrast, the factors derived from the 10 AT observers appear more regular or broadly systematic. This difference is likely because the ratio of between- to within- observer variance was substantially higher for this group, and thus the inter-observer differences are more readily captured by the factor analysis. Four of the 5 factors have a single dominant peak. We therefore fit these with Gaussian functions to estimate the peak and bandwidth of each of these factors (Emery et al., 2017a). The peaks for 3 of the factors tended to lie at angles intermediate to both the LvsM and SvsLM cardinal axes and to the AT observers' unique hue loci (Fig. 5), and may correspond to variations in the binary hues or categorical boundaries between the unique hues. However, this correspondence is very rough and we do not have firm evidence for identifying the basis or robustness for these factors.

For the anomalous set, the final systematic factor (factor 4) exhibits loadings near zero for the LvsM stimulus axis (stimulus angles of 0° and 180°) and peaking near the S-axis (angles of 90° and 270°; note that the polarity of factor loadings is arbitrary). Again a comparable factor is not evident in the factors for the NT dataset. This factor has approximately the pattern of loadings that would be expected if the anomalous observers' scaling functions differed because of mean shifts along the LvsM stimulus axis. Predictions for this variation are shown in Fig. 10 and were calculated from:

$$\alpha_{ji} = \sigma_x(d\rho_s/dx_i)/\sigma_j$$

Here α_{ji} is the predicted loading on stimulus angle j for the i-th factor, σ_x is the standard deviation in the factor, $d\rho_s/dx_i$ is the change in the setting for a unit change in the factor, and σ_i is the observed standard deviation in the observed values for the j-th stimulus (Webster & MacLeod, 1988). Again, we modeled the factor as a shift in the stimuli along the LvsM-axis, which would produce corresponding distortions in the angles of the stimuli. For example, a shift toward the +L pole would cause a counterclockwise rotation in the +S axis and a clockwise rotation in the –S axis. The standard deviation of this shift (σ_x) was varied to minimize the error between the predicted and observed loadings and corresponded to a value of 7.5 units (Fig. 10, which also shows the correlation structure reconstructed for this factor). As this shows, this factor largely underlies the large negative correlations observed in the AT settings but not the NT settings. This factor could potentially suggest that the anomalous observers differed in how much weight they gave to red vs. green in their hue-scaling functions. For the AT group, there was not an obvious similar factor corresponding to biases along the S or other axes. A second source of variation we might have expected was a change in the relative strength or contrast along the LvsM vs SvsLM axis, which, like the threshold predictions for the hue-scaling functions, might produce variability related to the relative strength of the red-green and blue-yellow responses. However, none of the other extracted factors were consistent with this pattern.

Of course, we do not know whether a very different factor pattern would emerge for the AT observers if we had a larger sample size, or if the sample of AT observers had individual differences of comparable magnitude to the NT sample. However, in general these analyses suggest the individual differences between our 10 AT observers were due to different and more robust sources than the differences for a comparable set of NT observers. One possibility is that the factors for the AT observers may be more closely related to the nature and manifestation of their color deficiencies, while for the NT observers the factors may instead reflect differences in how colors are categorized (Emery et al., 2017a).

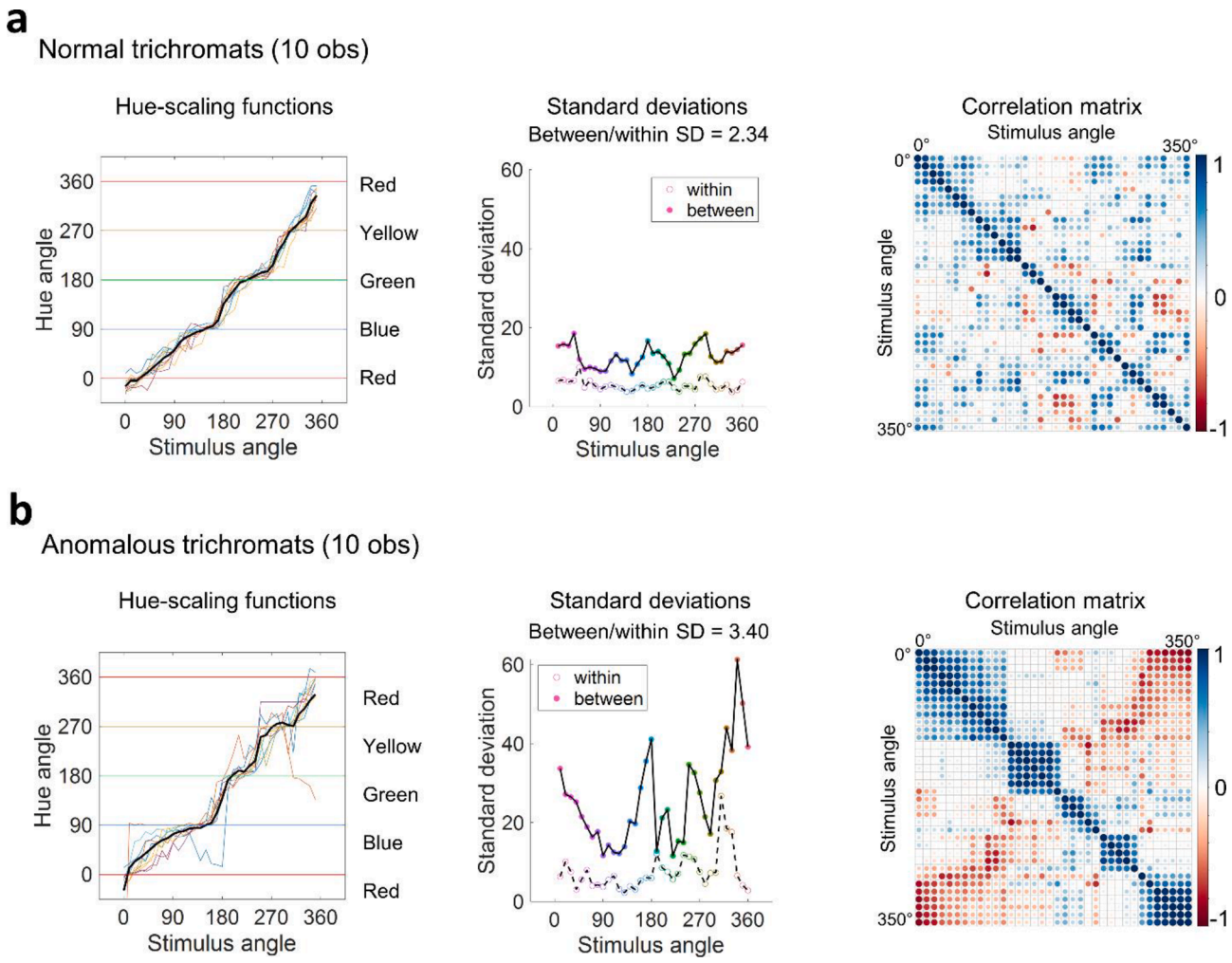

Fig. 8. Analysis of individual differences in the hue-scaling functions. a) a sample of 10 NT observers; b) 10 AT observers. Left panels: hue-scaling functions for the individual observers; middle panels: comparison of the standard deviation of the settings within (dashed) and between observers (solid); right panels: correlations between the stimulus angles (blue = positive correlations; red = negative). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Discussion

To summarize, in this study we compared hue-scaling functions in anomalous trichromats and normal trichromats. Hue scaling is widely used as a measure of color appearance, and requires observers to decompose their color percepts into the underlying primary dimensions of red-green and blue-yellow. As such, it provides a more graded and nuanced measure of appearance than categorical responses like color naming, and one which could potentially tap more directly into perceptual stages of judging colors rather than decision stages of classifying them. Supporting this, in our prior work we found that naming vs. scaling the same hues led to distinct patterns of individual differences (Emery et al., 2017b). The present results are consistent with previous studies in showing that anomalous observers report these percepts to be different from normal trichromats (Hemminger & Georgi, 1982; Nagy et al., 2014; Romeskie, 1978; Smith et al., 1973). However, the differences are substantially smaller than predicted by their threshold sensitivity for chromatic contrast. These results thus add to a growing literature pointing to compensation for the losses in chromatic sensitivity (Bosten, 2019). On the other hand, the fact that clear differences remain between the groups suggest that this compensation is only partial, and thus that there may be fundamental constraints on the processes of compensation. We consider these two sides to the problem in turn.

Compensation for color deficiency. Several previous studies have found that the suprathreshold color percepts of anomalous and normal trichromats differ less than their threshold sensitivities for color (Boehm, MacLeod, & Bosten, 2014; Knoblauch, Marsh-Armstrong, & Werner, 2020; Regan and Mollon, 1997). For example, Boehm et al. used multidimensional scaling to compare the relative perceived magnitude of signals along the LvsM and SvsLM chromatic axes. In the reconstructed spaces, AT's LvsM sensitivity was within ~80% of NTs even though their thresholds were on average only 38% of NTs, suggesting an ~2-fold gain in the strength of the visible color differences. Our AT participants had higher threshold losses, potentially because we relied on self-identified anomalous observers rather than a general screening of the population (Bosten, 2019). However, we similarly estimated a boost of the LvsM-mediated percepts that was ~4 times higher than the LvsM thresholds.

Several distinct processes could contribute to the heightened performance of the anomalous observers. Unfortunately, these remain difficult to disentangle because phenomenal measures like hue scaling do not directly measure the observer's percepts but only their descriptions of those percepts. One possibility is that they experience the colors as impoverished but have learned to label them in similar ways. For example, the red-green dimension could be given comparable weight to the blue-yellow dimension, simply because they have scaled the proportions based on the range of contrasts they are used to seeing along each dimension. A recent study suggests that these judgments

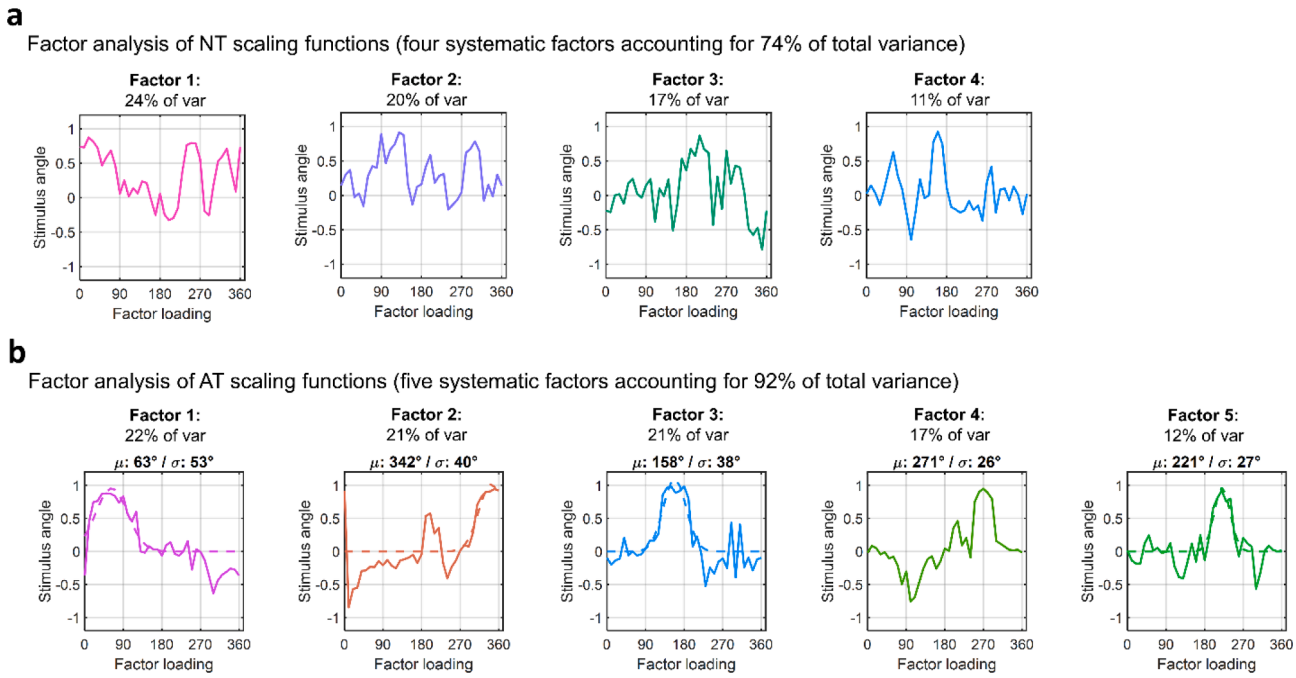

Fig. 9. Loadings for systematic factors extracted from factor analyses of the correlation matrices, and percentage of variance accounted for by each. a) factors for the NT hue-scaling functions; b) factors for the AT hue-scaling functions; dashed lines in the AT plots show Gaussians fit to the loadings with the best-fitting peak and bandwidth indicated above the panel.

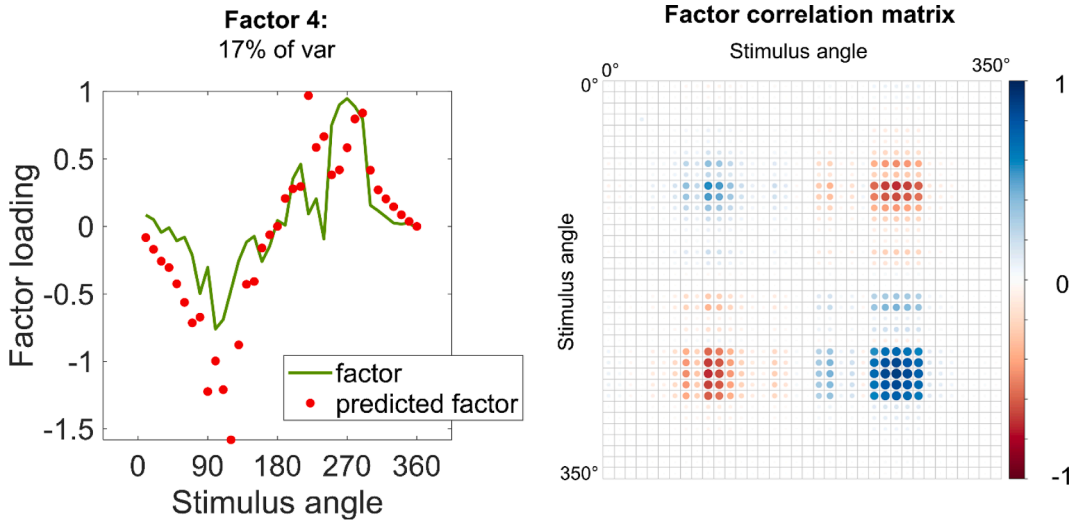

Fig. 10. Predicted factor loadings for factor 4 of the AT factor pattern, based on variations in biases in the mean of the stimuli along the LvsM axis. Left panel: comparison of predicted (red circles) to observed (green line) loadings; right panel: correlations between the stimulus angles reconstructed from the predicted factor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

could be highly malleable (Werner, Marsh-Armstrong & Knoblauch, 2020). After wearing filters designed to enhance reddish-greenish colors for a few days, anomalous trichromats experienced these hues as more salient even after the glasses were removed. Exposure to the world through the glasses may have led to a latent form of perceptual learning, where observers develop expertise in distinguishing subtle differences in sensory cues. This learning may be more closely tied to how well signals are utilized or decoded rather than encoded (Dosher & Lu, 2017), and in this sense could again reflect an amplification of the salience of the color signals rather than an actual boost in the response to color contrast.

A related issue is that color deficient observers could learn to base their judgments – or indeed their percepts – on different cues or sensory signals. As we noted in the Introduction, the unexpected performance of even dichromats has been attributed to the recruitment of additional sensory signals. Potential candidates include reliance on lightness differences in addition to chromatic signals; rod-cone interactions (or contributions from intrinsically photosensitive ganglion cells, for which a role has yet to be explored); retinal inhomogeneities in spectral sensitivity, for example, from macular screening or variations in photopigment optical density; and differences in how the cone signals are processed within post-receptoral mechanisms that might have different response nonlinearities (Bonnardel, 2006; Smith & Pokorny, 1977;

Smith et al., 1973; Wachtler et al., 2004).

A further challenge in interpreting compensatory processes is that they may depend crucially on the task used to measure them. For example, as we have noted, tasks like color naming may be more amenable to cognitive strategies, like learning to label the same stimulus in the same way as others even if the sensory impressions may be very different. Other tasks such as detection thresholds may instead be limited by very different constraints or levels of the visual system. With regard to the large color differences probed by studies of color compensation, an important example is measurements of reaction times. Compared to judgments of perceived contrast or salience, these show little evidence for an amplification of suprathreshold contrasts in anomalous observers, and are instead consistent with the threshold losses (Muller, Cavonius & Mollon, 1991; Vanston, Tregillus, Webster & Crognale, in press). Such differences emphasize the importance of probing compensation across a variety of tasks to decipher when and how it occurs.

Despite these caveats, there are strong theoretical reasons for expecting that part of the compensation for a loss in receptor sensitivity does reflect an actual amplification of post-receptoral responses. One generic mechanism for this is adaptation, which adjusts neural responses to match the strength of their inputs (Webster, 2015). For example, neurons receiving weaker signals may increase their gain to preserve the full dynamic range of their outputs. Adaptation is thought to play a critical role in calibrating color vision and perception more generally (Webster, 1996). For example, the stimulus that appears white depends on balanced activity across the cones and post-receptoral pathways, but this balance must be matched for something like the average spectral distribution in the observer's environment (Bosten, Beer, & MacLeod, 2015; Walraven & Werner, 1991). These adjustments must also correct for spectral biases introduced by the observer, for example because of preretinal screening pigments or the absorption spectra or relative numbers of the cones (e.g. (Brainard et al., 2000; Hardy, Frederick, Kay & Werner, 2005; Webster & Leonard, 2008). Anomalous observers would be expected to experience white very differently if they were not adapted to adjust for their own spectral sensitivity. Given that this adaptation is adjusting both normal and anomalous observers to the same environment, then their color percepts should be normalized in similar ways (Webster et al., 2010b). In particular, previous studies have noted that the normalization of the relative L and M cone weights for a stimulus norm like average daylight should drive similar calibrations of color coding and percepts for anomalous observers (Neitz et al., 2002; Pokorny & Smith, 1977).

Color vision adapts not only to the average stimulus but also to the range of color signals or contrast (Webster, 1996). Previous studies have noted that a process of this kind acting at post-receptoral levels could in principle undo the sensitivity losses from the reduced separation of the longer wavelength cones (MacLeod, 2003; Webster et al., 2010b). A similar process has also been invoked to explain the relative balance of neural signals for chromatic and luminance contrast in spite of the inherent differences in the magnitude of the cone contrasts supporting these dimensions (MacLeod, 2003). That is, the gamut of colors appears similar for observers despite pronounced differences in the magnitude of the cone signals mediating different directions within the gamut (McDermott & Webster, 2012). Notably however, in the case of normal trichromacy, this balance is also manifest at threshold. The cone contrasts for detecting a difference between the cones (color) are much lower than the thresholds for detecting their sums (luminance) (Chaparro, Stromeyer, Huang, Kronauer & Eskew, 1993), and suprathreshold perceived contrasts scale roughly as multiples of the threshold differences (Switkes, 2008; Switkes & Crognale, 1999). The "compensation" at threshold could reflect differences not only in signal strength but also in noise. In particular, thresholds for discriminating changes in luminance or saturation (measures of overall activity) should be higher than for detecting a change in hue (measure of differences in activity) if noise within the mechanisms is correlated (Danilova & Mollon, 2016).

In most (but not all) anomalous observers the signal to noise ratio is evidently weaker since their detection thresholds are higher. However, the relationship between their cone fundamentals (e.g. as inferred from Rayleigh matches) and threshold sensitivity (e.g. as inferred by matching range) may be much more variable (Bosten, 2019). Similarly, there may be a much weaker coupling between thresholds and suprathreshold perception. Consistent with this, the strength of red-green responses we estimated from the scaling functions were not correlated with the observers' LvsM thresholds. This disconnect between threshold vs. surpathreshold chromatic responses may be more akin to the phenomenon of contrast constancy, in which the large differences in threshold sensitivity for spatial contrast, as characterized by the bandpass contrast sensitivity function, are largely absent at suprathreshold contrasts (Georgeson & Sullivan, 1975). Here the differences are thought to arise because thresholds reflect the S/N ratio while perceived contrast primarily reflects only the signal (Brady & Field, 1995). Similarly, individuals vary widely in contrast sensitivity and acuity because of differences in the optical quality of their eyes. Yet suprathreshold judgments of image focus are largely compensated for the blur in the retinal image (Sawides, de Gracia, Dorronsoro, Webster & Marcos, 2011).

The adaptations to spatial contrast are thought to have a cortical locus, because for example the compensations for blur are strongly binocular (completely normalized for both eyes by the eye with better optical quality) (Radhakrishnan, Dorronsoro, Sawides, Webster & Marcos, 2015) and selective for the habitual differences in orientation sensitivity in astigmatism (Vinas, Sawides, de Gracia & Marcos, 2012). At shorter timescales, adaptation to just a few hours of a loss in contrast can lead to improvements in contrast sensitivity and to an amplification of contrast responses in early visual cortex as measured by fMRI BOLD signals (Kwon, Legge, Fang, Cheong, & He, 2009). Similar processes could plausibly underlie the compensation for habitual losses in chromatic contrast sensitivity, yet the neural bases for this compensation are only beginning to be investigated. The parvocellular pathway carrying the LvsM cone responses shows relatively little adaptation to chromatic contrast, suggesting gain changes mediating compensation might also have a cortical locus (Solomon, Peirce, Dhruv & Lennie, 2004). This is also supported by studies showing that chromatic contrast adaptation can be selective for a multitude of chromatic directions (Webster & Mollon, 1994), consistent with the wider range of chromatic preferences for cells in early visual cortex (Lennie, Krauskopf & Sclar, 1990). A recent study by Tregillus et al. directly tested for color compensation in the cortex and found that BOLD responses to chromatic contrast in anomalous observers were weaker than normal trichromats in cortical area V1 but similar in V2 and V3, suggesting V1 is the earliest site of the neural amplification (Tregillus et al., 2020). A study of color responses based on visual evoked potentials also found that amplification of the weaker AT cone signals only occurred for binocular viewing, which would also be consistent with a cortical site (Rabin, Kryder & Lam, 2018).

Limits to compensation. Thus far studies of compensation for a color deficiency have tended to focus on whether it occurs, with the null hypothesis that the receptoral losses lead to a simple reduced form of normal color vision. However, given the theoretical rationale for expecting compensation, as well as the many studies that have now demonstrated it, a more intriguing question may be why it rarely seems to be complete. What is it about visual coding that prevents post-receptoral processing to fully undo the signal losses from the receptors? Fewer studies have approached the nature of compensation from this direction, and the answers remain unclear. One important factor may again be that the cone losses impact the S/N ratio in the system. A gain change that operated after the levels which interject noise would not change this ratio (MacLeod, 2003; Rieke & Rudd, 2009), with the consequence that anomalous observers might experience colors to have similar strengths yet with greater inherent uncertainty (Webster et al., 2010b). A study of magnitude estimation for color differences in

fact found that the primary difference between normal and anomalous observers was greater noise in the responses of the latter (Hemminger & Georgi, 1982). However, this difference was not evident in our study. We instead found that the reliability of the settings was comparable for the AT and NT observers, and that the primary difference was instead in the apparent strength of the color responses. On the other hand, because we measured only hue and not saturation, we cannot rule out the possibility that saturation or perceived contrast settings are more variable in the anomalous observers. For example as noted above, discriminating a saturation change is harder than a hue change even for color-normal observers, potentially because the noise within different chromatic mechanisms is correlated (Danilova & Mollon, 2016).

A second potential limit could again relate to the cues anomalous observers are using to judge the hues. As noted these might involve different pathways or signals from the primary color-opponent mechanisms (Bonnardel, 2006; Smith et al., 1973; Wachtler et al., 2004). If the strength of these secondary cues is inherently weaker, then this could limit the dynamic range of color responses in the color deficient observers.

Compensation could also be incomplete if there are limits to the degree to which visual responses can adjust to changes in the observer. For example, color vision in the periphery can remain far more similar to central vision than expected from the large retinal variations with eccentricity (Webster, Halen, Meyers, Winkler & Werner, 2010a), yet does not maintain complete constancy (Bompas, Powell, & Sumner, 2013; McKeefry, Murray, & Parry, 2007; Parry, McKeefry, & Murray, 2006). As noted above, retinal and geniculate cells in the parvocellular (P) pathway also do not show significant adaptation to chromatic contrast (Solomon et al., 2004) (though see (Chang, Hess & Mullen, 2016)). Lutze et al. also pointed out that the contrast gain in this pathway is similar for normal and color-deficient observers and also for dichromatic and trichromatic species of primates (Lutze, Pokorny & Smith, 2006). This suggests that the response functions – at least at this early level of the visual system – were not calibrated for the diet of cone responses experienced over life for different individuals. However, P cells also respond to achromatic contrast and may also be the principal pathway carrying lightness (as opposed to luminance) information (Lennie, Pokorny & Smith, 1993). The gain could therefore also reflect long-term adaptation to the achromatic contrasts. More generally, this raises the problem that the adjustments for a color deficiency do not occur in isolation, but instead within the context of a system that must be optimized for many different parameters and goals. It is possible that partial compensation reflects a compromise among many potentially competing calibrations.

A final example of a potential limiting factor could be the ability of the system to distinguish the signals from the separate cone classes. The L and M cones appear to differ only in their photopigment opsins but are otherwise morphologically and physiologically similar. This has suggested that the brain must learn from visual experience how to identify the type of cone from which the signal is coming. Simulations of this learning have shown that it is possible to classify the normal trichromatic cone mosaic with high accuracy from the correlations of the cone responses to natural scene statistics (Benson, Manning, Brainard, & Bethge, 2014; Wachtler, Doi, Lee, & Sejnowski, 2007). However, this classification breaks down as the spectral separation between the cones is reduced (Benson et al., 2014). Misclassifications would tend to dilute the LvsM signals in post-receptoral mechanisms, and this could impose a further limit on how fully the visual responses to color could be renormalized. It would be instructive to model whether the LvsM contrast losses expected from misclassifying the cones could account for some of the residual losses anomalous trichromats exhibit in color appearance.

Implications for sensory plasticity. While our study is focused on the specific context of color deficiencies, the results have broader implications for understanding general processes of plasticity and adaptation. In fact, inherited color losses are an ideal natural model for exploring plasticity, because the photopigment changes alter only the first step of visual processing, are present from birth, and remain stable throughout the individual's life (Isherwood, Joyce, & Webster, 2020). Thus individuals have a lifetime of opportunity to adjust to the change, providing a window into compensatory processes over the longest possible ontogenetic timescales. The pronounced yet partial perceptual discounting that color anomalous observers make for their receptor losses thus provide strong clues about general principles and limits of compensation and calibration in sensory systems.

Implications for individual differences in color perception. A striking feature of normal variations in color perception is that they are both large and also difficult to predict from the individual differences in spectral sensitivity (Emery & Webster, 2019; Webster, 2020). Thus what appears white or a unique hue does not strongly depend on how sensitive the observer is to different wavelengths. This again points to compensatory processes in color vision that tend to correct for the particular sensitivity limits of the observer. The dissociation between sensitivity and appearance has also raised the prospect that the perceptual structure of color is determined more by salient properties of the environment than by the specific dimensions underlying the neural coding of color (Mollon, 2006). For example, unique blue and yellow may look unique because they correspond to a prominent stimulus variation – the daylight locus – and not because they isolate signals in a putative blue-yellow opponent mechanism (Mollon, 1982; Pokorny & Smith, 1977). By this account, one might expect the hue loci of anomalous observers to be very similar to normal trichromats – as long as the environmental signals are discriminable, they could learn like others to associate them with specific perceptual responses. The extent to which the color percepts in anomalous trichromats – with markedly different sensor properties – are nevertheless concordant with those of normal trichromats supports this idea, and the general principle that percepts are adapted to match the environment. However, the fact that these compensations also tend to be only partial suggests that an understanding of color appearance lies somewhere in between. That is, there also seem to be important biological limits on the loci of color categories, and these are evidenced by the systematic differences in the hue loci for the anomalous observers as seen in the present and previous studies. This could be because the processes of compensation themselves have biological constraints, so that the varieties of perceptual plasticity are limited. Physiological constraints on hue loci have also been hinted at in the normal population. For example, unique green settings are correlated with the density of macular pigment (Welbourne, Thompson, Wade & Morland, 2013) and the lightness of the iris (Jordan & Mollon, 1995), and have also been found to vary with the L to M cone ratio (Schmidt, Neitz & Neitz, 2014). In the latter case these have been successfully modeled as an interplay between normalization to environmental signals that adjust the relative weights of the L and M cones, and consequences of this weighting for the presumed spectral sensitivities of color-opponent channels. How our understanding of the basis of color perception and how it is calibrated unfolds will no doubt take many more twists. However, the similarities and differences between observers viewing the same world through very different receptors are likely to provide powerful insights.

Acknowledgement

Supported by EY-010834.

References

Abramov, I., & Gordon, J. (1994). Color appearance: On seeing red–or yellow, or green, or blue. *Annual Review of Psychology, 45*(1), 451–485.

Benson, N. C., Manning, J. R., Brainard, D. H., & Bethge, M. (2014). Unsupervised learning of cone spectral classes from natural images. *PLoS Computational Biology, 10* (6), e1003652.

Boehm, A. E., MacLeod, D. I. A., & Bosten, J. M. (2014). Compensation for red-green contrast loss in anomalous trichromats. *Journal of Vision, 14*(13), Article 19-19.

Bompas, A., Powell, G., & Sumner, P. (2013). Systematic biases in adult color perception persist despite lifelong information sufficient to calibrate them. *Journal of Vision, 13*(1), 19.
Bonnardel, V. (2006). Color naming and categorization in inherited color vision deficiencies. *Visual Neuroscience, 23*(3-4), 637–643.
Bosten, J. (2019). The known unknowns of anomalous trichromacy. *Current Opinion in Behavioral Sciences, 30*, 228–237.
Bosten, J. M., Beer, R. D., & MacLeod, D. I. A. (2015). What is white? *Journal of Vision, 15*(16), 5.
Brady, N., & Field, D. J. (1995). What's constant in contrast constancy? The effects of scaling on the perceived contrast of bandpass patterns. *Vision Research, 35*(6), 739–756.
Brainard, D. H., Roorda, A., Yamauchi, Y., Calderone, J. B., Metha, A., Neitz, M., et al. (2000). Functional consequences of the relative numbers of L and M cones. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 17*(3), 607–614.
Broackes, J. (2010). 12 What Do the Color-Blind See? In: Color ontology and color science (p. 291).
Chang, D. H. F., Hess, R. F., & Mullen, K. T. (2016). Color responses and their adaptation in human superior colliculus and lateral geniculate nucleus. *Neuroimage, 138*, 211–220.
Chaparro, A., Stromeyer, C. F., 3rd, Huang, E. P., Kronauer, R. E., & Eskew, R. T., Jr. (1993). Colour is what the eye sees best. *Nature, 361*(6410), 348–350.
Danilova, M. V., & Mollon, J. D. (2016). Superior discrimination for hue than for saturation and an explanation in terms of correlated neural noise. *Proceedings of the Royal Society B: Biological Sciences, 283*(1831), 20160164.
De Valois, R.L., De Valois, K.K., Switkes, E., & Mahon, L. (1997). Hue scaling of isoluminant and cone-specific lights. Vision Research, 37 (7), 885-897.
Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of Physiology, 357*, 241–265.
Dosher, B., & Lu, Z.-L. (2017). Visual perceptual learning and models. *Annual Review of Vision Science, 3*(1), 343–363.
Emery, K. J., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017a). 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.
Emery, K. J., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017b). Variations in normal color vision. VII. Relationships between color naming and hue scaling. *Vision Research, 141*, 66–75.
Emery, K. J., & Webster, M. A. (2019). Individual differences and their implications for color perception. *Current Opinion in Behavioral Sciences, 30*, 28–33.
Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physiology, 252*(3), 627–656.
Gordon, J., Abramov, I., & Chan, H. (1994). Describing color appearance: Hue and saturation scaling. *Perception & Psychophysics, 56*(1), 27–41.
Hardy, J. L., Frederick, C. M., Kay, P., & Werner, J. S. (2005). Color naming, lens aging, and grue: What the optics of the aging eye can teach us about color language. *Psychological Science, 16*(4), 321–327.
Hemminger, H., & Georgi, W. (1982). Color sensation of normal and anomalous trichromats measured by magnitude estimation. *Psychological Research, 44*(2), 147–163.
Isherwood, Z.J., Joyce, D.S., M.K., P., & Webster, M.A. (2020). Pasticity in perception: Insights from color deficiencies. Faculty Reviews, 9:8.
Jameson, D., & Hurvich, L. M. (1959). Perceived color and its dependence on focal, surrounding, and preceding stimulus variables. *Journal of the Optical Society of America, 49*(9), 890–898.
Jameson, D., & Hurvich, L. M. (1978). Dichromatic color language: "reds" and "greens" don't look alike but their colors do. *Sensory Processes, 2*(2), 146–155.
Jordan, G., & Mollon, J. D. (1995). Rayleigh matches and unique green. *Vision Research, 35*(5), 613–620.
Knoblauch, K., Marsh-Armstrong, B., & Werner, J. S. (2020). Suprathreshold contrast response in normal and anomalous trichromats. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 37*(4), A133–A144.
Knoblauch, K., & Shevell, S. K. (2001). Relating cone signals to color appearance: Failure of monotonicity in yellow/blue. *Visual Neuroscience, 18*(6), 901–906.
Krauskopf, J., Williams, D. R., & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research, 22*(9), 1123–1131.
Kwon, M., Legge, G. E., Fang, F., Cheong, A. M. Y., & He, S. (2009). Adaptive changes in visual cortex following prolonged contrast reduction. *Journal of Vision, 9*(2), 20.
Larimer, J., Krantz, D. H., & Cicerone, C. M. (1975). Opponent process additivity. II. Yellow/blue equilibria and nonlinear models. *Vision Research, 15*(6), 723–731.
Lennie, P., Krauskopf, J., & Sclar, G. (1990). Chromatic mechanisms in striate cortex of macaque. *Journal of Neuroscience, 10*(2), 649–669.
Lennie, P., Pokorny, J., & Smith, V. C. (1993). Luminance. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 10*(6), 1283–1293.
Lutze, M., Pokorny, J., & Smith, V. C. (2006). Achromatic parvocellular contrast gain in normal and color defective observers: Implications for the evolution of color vision. *Visual Neuroscience, 23*(3-4), 611–616.
MacLeod, D. I., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America, 69*(8), 1183–1186.
MacLeod, D. I. A. (2003). Colour discrimination, colour constancy, and natural scene statistics (The Verriest Lecture). In J. D. Mollon, J. Pokorny, & K. Knoblauch (Eds.), *Normal and Defective Colour Vision*. London: Oxford University Press.
Martin, P. R., White, A. J., Goodchild, A. K., Wilder, H. D., & Sefton, A. E. (1997). Evidence that blue-on cells are part of the third geniculocortical pathway in primates. *European Journal of Neuroscience, 9*(7), 1536–1541.
Matera, C. N., Emery, K. J., Volbrecht, V. J., Vemuri, K., Kay, P., & Webster, M. A. (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.
McDermott, K. C., & Webster, M. A. (2012). The perceptual balance of color. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 29*(2), A108–A117.
McKeefry, D. J., Murray, I. J., & Parry, N. R. (2007). Perceived shifts in saturation and hue of chromatic stimuli in the near peripheral retina. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 24*(10), 3168–3179.
Mollon, J. D. (1982). Color vision. *Annual Review of Psychology, 33*(1), 41–85.
Mollon, J. D. (2006). Monge (The Verriest Lecture). *Visual Neuroscience, 23*, 297–309.
Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision Research, 141*, 4–15.
Muller, M., Cavonius, C. R., & Mollon, J. D. (1991). Constructing the color space of the deuteranomalous observer. In B. Drum, & A. L. Serra (Eds.), *Colour Vision Deficiencies X* (pp. 377–387). Kluwer.
Nagy, B. V., Nemeth, Z., Samu, K., & Abraham, G. (2014). Variability and systematic differences in normal, protan, and deutan color naming. *Frontiers in Psychology, 5*, 1416.
Neitz, J., Carroll, J., Yamauchi, Y., Neitz, M., & Williams, D. R. (2002). Color perception is mediated by a plastic neural mechanism that is adjustable in adults. *Neuron, 35*(4), 783–792.
Neitz, J., & Neitz, M. (2011). The genetics of normal and defective color vision. *Vision Research, 51*(7), 633–651.
Neitz, J., Neitz, M., He, J. C., & Shevell, S. K. (1999). Trichromatic color vision with only two spectrally distinct photopigments. *Nature Neuroscience, 2*(10), 884–888.
Parry, N. R., McKeefry, D. J., & Murray, I. J. (2006). Variant and invariant color perception in the near peripheral retina. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 23*(7), 1586–1597.
Pokorny, J., & Smith, V. C. (1977). Evaluation of single-pigment shift model of anomalous trichromacy. *Journal of the Optical Society of America, 67*(9), 1196–1209.
Rabin, J., Kryder, A., & Lam, D. (2018). Binocular facilitation of cone-specific visual evoked potentials in colour deficiency. *Clinical and Experimental Optometry, 101*(1), 69–72.
Radhakrishnan, A., Dorronsoro, C., Sawides, L., Webster, M. A., & Marcos, S. (2015). A cyclopean neural mechanism compensating for optical differences between the eyes. *Current Biology, 25*(5), R188–R189.
Regan, B.C., & Mollon, J.D. (1997). The relative salience of the cardinal axes of colour space in normal and anomalous trichromatsIn: C.R. Cavonius (Ed.) Colour Vision Deficiencies XIII (pp. 261-270). Kluwer: Dordrecht.
Rieke, F., & Rudd, M. E. (2009). The challenges natural images pose for visual adaptation. *Neuron, 64*(5), 605–616.
Romeskie, M. (1978). Chromatic opponent-response functions of anomalous trichromats. *Vision Research, 18*(11), 1521–1532.
Sawides, L., de Gracia, P., Dorronsoro, C., Webster, M. A., & Marcos, S. (2011). Vision is adapted to the natural level of blur present in the retinal image. *PLoS ONE, 6*(11), e27031.
Schmidt, B. P., Neitz, M., & Neitz, J. (2014). Neurobiological hypothesis of color appearance and hue perception. *Journal of the Optical Society of America A, 31*(4), A195. https://doi.org/10.1364/JOSAA.31.00A195.
Smith, V. C., & Pokorny, J. (1977). Large-field trichromacy in protanopes and deuteranopes. *Journal of the Optical Society of America, 67*(2), 213–220.
Smith, V. C., Pokorny, J., & Swartley, R. (1973). Continuous hue estimation of brief flashes by deuteranomalous observers. *American Journal of Psychology, 86*(1), 115–131.
Solomon, S. G., Peirce, J. W., Dhruv, N. T., & Lennie, P. (2004). Profound contrast adaptation early in the visual pathway. *Neuron, 42*(1), 155–162.
Stockman, A., & Sharpe, L. T. (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research, 40*(13), 1711–1737.
Switkes, E. (2008). Contrast salience across three-dimensional chromoluminance space. *Vision Research, 48*(17), 1812–1819.
Switkes, E., & Crognale, M. A. (1999). Comparison of color and luminance contrast: Apples versus oranges? *Vision Research, 39*(10), 1823–1831.
Thomas, P. B. M., Formankiewicz, M. A., & Mollon, J. D. (2011). The effect of photopigment optical density on the color vision of the anomalous trichromat. *Vision Research, 51*(20), 2224–2233.
Tregillus, K.E.M., Isherwood, Z.J., Vanston, J.E., Engel, S.A., MacLeod, D.I.A., Kuriki, I., & Webster, M.A. (2020). Color compensation in anomalous trichromats assessed with fMRI. Current Biology.
Uchikawa, K. (2008). Trichromat-like categorical color naming in dichromats. *Vision, 20*(2), 62–66.
Vanston, J.E., Tregillus, K.E.M., Webster, M.A., & Crognale, M.A. (in press). Task-dependent contrast gain in anomalous trichromats.
Vinas, M., Sawides, L., de Gracia, P., & Marcos, S. (2012). Perceptual adaptation to the correction of natural astigmatism. *PLoS One, 7*(9), e46361.
Wachtler, T., Dohrmann, U., & Hertel, R. (2004). Modeling color percepts of dichromats. *Vision Research, 44*(24), 2843–2855.
Wachtler, T., Doi, E., Lee, T., & Sejnowski, T. J. (2007). Cone selectivity derived from the responses of the retinal cone mosaic to natural scenes. *Journal of Vision, 7*(8), 6.
Walraven, J., & Werner, J. S. (1991). The invariance of unique white; a possible implication for normalizing cone action spectra. *Vision Research, 31*(12), 2185–2193.
Webster, M. A. (1996). Human colour perception and its adaptation. *Network: Computation in Neural Systems, 7*(4), 587–634.
Webster, M. A. (2015). Visual adaptation. *Annual Review of Vision Science, 1*(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., Halen, K., Meyers, A. J., Winkler, P., & Werner, J. S. (2010). Colour appearance and compensation in the near periphery. *Proceedings of the Royal Society B-Biological Sciences, 277*(1689), 1817–1825.
Webster, M. A., Juricevic, I., & McDermott, K. C. (2010). Simulations of adaptation and color appearance in observers with varying spectral sensitivity. *Ophthalmic and Physiological Optics, 30*(5), 602–610.
Webster, M. A., & Leonard, D. (2008). Adaptation and perceptual norms in color vision. *Journal of the Optical Society of America A, 25*(11), 2817–2825.
Webster, M. A., & MacLeod, D. I. (1988). Factors underlying individual differences in the color matches of normal observers. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 5*(10), 1722–1735.
Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000a). Variations in normal color vision. I. Cone-opponent axes. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 17*(9), 1535–1544.
Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000b). 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.
Webster, M. A., & Mollon, J. D. (1994). The influence of contrast adaptation on color appearance. *Vision Research, 34*(15), 1993–2020.
Welbourne, L. E., Thompson, P. G., Wade, A. R., & Morland, A. B. (2013). The distribution of unique green wavelengths and its relationship to macular pigment density. *Journal of Vision, 13*(8), 15 15.
Werner, J. S., Marsh-Armstrong, B., & Knoblauch, K. (2020). Adaptive changes in color vision from long-term filter usage in anomalous but not normal trichromacy. *Current Biology, 30*(15), 3011–3015 e3014.
Wuerger, S. M., Atkinson, P., & Cropper, S. (2005). The cone inputs to the unique-hue mechanisms. *Vision Research, 45*(25-26), 3210–3223.