

Context-dependent judgments of color that might allow color constancy in scenes with multiple regions of illumination

R. J. Lee* and H. E. Smithson

Department of Experimental Psychology, South Parks Road, Oxford, OX1 3UD, UK

**Corresponding author: robert.lee@psy.ox.ac.uk*

Received September 2, 2011; revised November 18, 2011; accepted December 2, 2011;
posted December 12, 2011 (Doc. ID 153983); published January 27, 2012

For a color-constant observer, a change in the spectral composition of the illumination is accompanied by a corresponding change in the chromaticity associated with an achromatic percept. However, maintaining color constancy for different regions of illumination within a scene implies the maintenance of multiple perceptual references. We investigated the features of a scene that enable the maintenance of separate perceptual references for two displaced but overlapping chromaticity distributions. The time-averaged, retinotopically localized stimulus was the primary determinant of color appearance judgments. However, spatial separation of test samples additionally served as a symbolic cue that allowed observers to maintain two separate perceptual references. © 2012 Optical Society of America

OCIS codes: 330.1690, 330.1720, 330.5020, 330.5510, 330.7320.

1. INTRODUCTION

Our perceptual judgments—for example, of size, orientation, or color—are often dependent upon the context in which a stimulus is encountered. Such judgments are typically relative rather than absolute, in that the properties of what is being judged are compared with a reference that might change from situation to situation. The reference might be concurrently available in the scene—for example, when the size of an object is judged relative to that of an adjacent one. Alternatively, the reference might be maintained internally, either explicitly defined in memory or specified implicitly, perhaps as the average of recent observations. It has been shown that observers can simultaneously maintain more than one internal reference for one perceptual quality, and can make judgments against the correct reference depending on the context [1]. In this paper, we specifically investigate our ability to make this kind of context-dependent judgement about color.

The stimuli we used in this study were artificial, and were defined to exhibit particular chromatic statistics across time and space, but the conditions we chose have parallels in natural scenes. For a scene composed of illuminated objects, the chromatic signals reaching the eye depend on the reflectance properties of the objects and on the spectral composition of the illumination. Different illumination produces different chromatic statistics and, to maintain constancy of object color perception, the visual system must compensate for this difference. Outside the laboratory, it is rare that a scene is uniformly illuminated, so a scene may comprise multiple regions, each with different chromatic statistics. To produce color-constant judgments in such scenes, observers would have to maintain multiple perceptual references. Given that it is desirable that constancy should be maintained during natural viewing of complex scenes, we sought cues that are sufficient to support the maintenance of separate perceptual references

against which to make color judgments that are appropriate to different illuminants.

Effects of simultaneous and successive color induction imply that the color of a target region can be encoded relative to colors that are nearby in space or time. Both distal and central mechanisms have been proposed to account for such effects. We considered the influence of three factors in setting color judgments of target stimuli. These were the chromatic statistics of a spatial surround and the history of samples presented at the retinotopic and spatiotopic coordinates of the target. These conditions parallel natural viewing of illuminated objects in which target objects appear in differently illuminated spatial regions, and eye movements are made within or across illumination boundaries.

A. Perceptual References and Color

In an influential set of papers, Helson [2,3] proposed that judgments are made relative to an internal standard, and that this standard is set by the geometric mean of previous stimuli to which the observer has been exposed. Helson's "adaptation level theory" proposes that the internal standard is held centrally, and that the effect of adaptation is to adjust the value of the standard, or null point, on the decision axis. In color perception, it is clear that an observer's achromatic point can move in color space following exposure to chromatically biased stimuli. The mechanisms underlying such perceptual adjustments have been studied extensively (e.g., see [4] for review). Importantly, they reflect adjustments that can occur at different processing stages. For example, adaptation occurs within the cones themselves [5,6] at early postreceptoral sites [7,8] and at cortical sites [9]. These adjustments can be shown to exhibit additive and subtractive and divisive and multiplicative components [10]. For the current experiments, it is important to note that distal adaptation is not always complete [11,12] so stimulus properties that drive differential

adaptation need not be discarded early in processing, nor is it desirable that they should be [13]. Adjustments can be made over very different time scales, ranging from milliseconds [14] to several minutes [15]. Adaptation over very long periods of time has also been shown to have predictable effects on chromatic appearance, and these effects are long lasting [16,17].

Several characteristics of visual input have been identified as potential determinants of the adaptation level of these mechanisms. Adaptation that achieves independent multiplicative adjustments of the L-, M- and S-cone signals (where L, M and S are the long-, middle-, and short-wavelength sensitive photoreceptors) is generally referred to as von Kries adaptation [5]. Ives argued that setting the coefficients of a diagonal transform in inverse proportion to the cone signals elicited by the illuminant would achieve color constancy. We now know that, to first approximation, Ives' scheme would indeed discount the illuminant since, for a wide range of illuminants and reflectances encountered in the environment, an illuminant change imposes an approximately multiplicative transformation of the cone signals elicited by a collection of reflectances [18]. The cone signals elicited by the illuminant might be obtained directly (e.g., from a view of the source or from specular highlights), or from cone signals of the brightest surface (assuming this is white). Alternatively, the same transformation could be achieved by normalization to the signals from a known object of diagnostic color, or from a spatial mean of cone signals across the scene (assuming the world is on average achromatic). These potential referents are reviewed in more detail in [19].

To achieve color constancy by using the mean chromaticity as a reference requires that the mean is taken over a sufficiently large number of reflectance samples. In the limit of only a single sample, spectral reflectance and the spectral content of the illuminant are completely confounded. However, several mechanisms have been proposed to combine multiple samples. Adaptation in the cones themselves has been shown to be very local, on the same scale of individual photoreceptors [20–22]. Cells with large receptive fields (e.g., those found in V4 [23,24]) might integrate signals from a large area of the visual field. Scattered light within the eye carries a spectral composition that depends on the mean chromaticity of the retinal image, and it is possible that the cone-rich ora serrata might detect this signal [25]. If the time constants of temporal adaptation were sufficiently long, spatially localized mechanisms might be made to extract a spatial average if eye movements sampled sufficient spatial locations [26]. Indeed, we have shown that, with limited eye movements, the temporal sequence of chromaticities is equally as effective in determining color appearance as the spatial array of chromaticities [27,28]. To first approximation, temporal adaptation is set by the mean chromaticity of the sequence of chromatic samples to which the observer is exposed, at least over a time course of tens of seconds [12]. However, it is also clear that there are nonlinearities that feed into temporal summation such that the arithmetic mean is not the best predictor of adaptation state [29]. The geometric mean [2] might be a better predictor.

B. Multiple References

We have discussed processes by which a perceptual reference might be specified, according to signals that are available in

the scene at any moment or to signals derived from the recent or prolonged stimulus history. However, any process that achieves color constancy through comparison with (or normalization to) a reference—whether it be the mean over time or space, the brightest signal, or a signal of known identity, like the illuminant itself or an object of diagnostic color—must ensure that the reference is appropriate for the illumination falling on the target surface. As a general scheme for color constancy, comparison with a reference stimulus will work only if it is possible to maintain as many references as there are regions of illumination and to update this reference, or switch to a new reference, if the illumination changes. Many striking failures of color constancy arise because cues to the illumination on the target are impoverished while cues to a second illuminant dominate. Scenes with multiple regions of illumination have received considerable attention in the lightness constancy literature—where illumination regions are often described as frameworks [30]—and indeed the Gelb effect [31] is a striking misattribution of high luminance to the lightness of a surface rather than interpreting the bright region as a concealed source of illumination. A global spatial average, derived, for example, from the scattered light in the eye, cannot support constancy for different illuminants that are concurrently present in distinct regions of the visual field [25]. Similarly, any spatial integration—for example, from cells that integrate over large receptive fields or from temporal integration across eye movements—will provide a useful estimate of the illuminant only if the spatial parameter is matched to a region of single illumination.

The issue of maintaining multiple perceptual references has been tested in the case of size and orientation judgments. Morgan [1] performed an experiment in which observers were required to judge the separation of lines against a standard. Different standards could be maintained for lines appearing in different orientations or in different locations over the course of the experiment, and judgments could be made against the appropriate standard on a given trial. In a similar experiment [32], they showed that these standards could be generated from the mean of the set of values observed over the course of an experimental session, with no increase in threshold compared to conditions in which the reference was explicitly given. Further, with multiple sets, the cue indicating which reference to use in each judgement could be symbolic (a briefly presented digit). We use the term symbolic cue to indicate a marker for set membership that resides in a different stimulus dimension from the one being judged. The location cues that we use could additionally be considered as grouping factors, and previous work shows that grouping can indeed influence color appearance when all other factors are held constant [33].

C. Our Experiment

In our experiment, we tested the extent to which observers maintained different criteria for the two chromatic contexts that we defined. In some conditions, the only difference between target stimuli was the chromatic distribution from which they are drawn; in other conditions, additional cues were introduced such that samples from different stimulus sets were presented in separate spatial or retinotopic locations or against backgrounds comprised of samples from one or another distribution.

For simplicity, we used stimuli that differed along only one dimension in color space, and the two contexts were distinguished by translating the sample chromaticities along this dimension, to obtain separate but overlapping distributions. We chose equiluminant stimuli that varied in their $L/(L + M)$ coordinates, so one set appeared green-biased and the other red-biased. For equiluminant stimuli, a translation along the $L/(L + M)$ axis is consistent with von Kries scaling. Although we used an impoverished distribution of chromaticities compared to distributions encountered in the environment, we note that a change in the spectral composition of illumination on a set of reflectances is well described as a translation of the $L/(L + M)$ coordinates of the stimuli (see [28], their Fig. 2, top right).

Samples to be judged were presented one after another, on a background of chromatic samples. Instead of using a matching task, as is often used in color-constancy experiments, we required observers to make binary decisions about stimuli, and we controlled the time course of the stimulus presentation. In this way, we had more control over the adaptation state of the observer. To illustrate the logic of our experiment, it is simplest to consider two extreme conditions: (a) with no cues to separate the two contexts and (b) with several available cues. In the no-cue condition, the chromatic statistics of the spatial surround were defined by the combination of the two sets, and test samples were selected at random from the two sets and presented at a location that was fixed in retinotopic and spatiotopic coordinates. In this case, there is nothing to distinguish the two sets, so observers must use the same perceptual reference for every stimulus presentation. This is contrasted with a condition in which multiple cues are available: Stimuli from one set are presented in one location, while stimuli from the other set are presented in a different location, and each test location is surrounded by a background of chromatic samples drawn from the corresponding set. This is analogous to a scene with two regions of illumination, one green-biased and the other red-biased. In this case, we can identify several cues to the different chromatic contexts, which might cause observers to judge chromaticities from the two sets against different references. At any moment, the mean chromaticity of the consistent surround, and the local contrast between the test patch and the surround, allows different judgments to be made to stimuli from the two contexts. Over time, the two regions of the retina become differently adapted, and the test-patch locations are exposed to either green-biased or red-biased samples. Furthermore, there is a symbolic cue that one location is associated with greenish and the other with reddish samples.

In the full set of stimulus conditions, we manipulated the availability of these different cues. One issue of interest is the relationship between spatial and temporal adaptation and eye movements. We included conditions in which the spatiotopic separation of stimuli was maintained, but in which we required observers to make eye movements that interleaved on the retina samples from the two chromatic distributions. With this manipulation we tested whether, when retinotopic adaptation is matched for the two chromatic contexts, identical samples might still receive different judgments based on the chromatic context to which they belong. If color constancy is achieved via localized retinotopic adaptation to the time-averaged stimulation (perhaps with eye movements

to obtain a spatially extended average), constancy would result only if adaptation were confined to a region of single illumination. The constancy mechanism might be rescued, however, if a symbolic cue allowed the maintenance of separate references (or central “adaptation levels”) for different spatial regions.

We find that the strongest evidence for observers maintaining separate references occurs in conditions with differential retinotopic adaptation. The cues of spatial location and context from the spatial surround do have effects, and these are most pronounced when the cues are combined. Although we do not restrict our interpretation to constancy, we do point out that the maintenance of multiple references is vital for constancy in real-world environments.

2. METHODS

A. Observers

Four observers (HES, LKY, RB, and RJL, two female, two male) collected the full set of data. A further four observers (CCP, KF, KLB, and LJN, three female, one male) collected a subset of the data. All observers made Rayleigh matches in the normal range on the anomaloscope, and were aged between 22 and 40 years. All observers were experienced in visual psychophysics experiments, but only HES and RJL were highly experienced in making color appearance judgments and aware of the experimental design.

B. Apparatus

Stimuli were presented on a CRT monitor (Mitsubishi DiamondPro 2070SB). The monitor display had size $0.386\text{ m} \times 0.295\text{ m}$, a spatial resolution of 800×600 pixels (each pixel measured approximately 0.5 mm), and a refresh rate of 100 Hz . The display was driven by a CRS ViSaGe system (Cambridge Research Systems Ltd., Rochester, U.K.), which has 14-bit-per-channel chromatic resolution. The system was gamma-corrected with a CRS OptiCAL and spectrally calibrated with readings taken with a CRS SpectroCAL. The distance between the observer and the face of the monitor was 0.6 m . The horizontal dimension of the monitor subtended a visual angle of 36° , and the vertical dimension an angle of 27° .

C. Stimuli

1. Scenes

Each scene comprised a background and a test patch. The background was made from 300 ellipses, each randomly oriented and positioned across the display. The size of each ellipse was also randomized, so that one axis of the ellipse was fixed at 50 pixels ($\sim 2.3^\circ$) and the other varied between 25 and 100 pixels ($\sim 1.15^\circ$ to 4.6°). The ellipses were placed so that many overlapped. A new spatial arrangement of ellipses was used on every trial, and the chromaticities of the ellipses were also chosen from trial to trial, as specified below. The test patch was a square of side length 100 pixels ($\sim 4.6^\circ$), placed in one of five positions, depending on the experimental condition and trial (explained below). The test patch was not overlapped by any of the background ellipses. A small black dot, used as a fixation target, was placed at the horizontal center of the display, and could appear in one of three vertical positions, again depending on the condition and trial. A smaller black dot appeared in the center of the test patch to indicate its location.

2. Chromaticities

Each ellipse and the test patch was assigned a chromaticity from one of two overlapping sets. The chromaticities all lay on a horizontal line in the MacLeod–Boynton [34] equiluminant chromaticity diagram, constructed from the Stockman and Sharpe [35] cone fundamentals. Differences between test patches therefore isolated the $L/(L + M)$ chromatic mechanism of the standard observer, while the $S/(L + M)$ excitation was matched across patches, as was the $(L + M)$ excitation. A translation along the $L/(L + M)$ axis is equivalent to a “von Kries transform,” i.e., a linear scaling of photoreceptor excitations, with the constraint that luminance ($L + M$) is held constant. We did not simulate real reflecting surfaces or illuminants but, while we unnaturally limit chromatic variation to a single dimension and therefore eliminate some cues that might be used to disambiguate surface and illuminant changes (e.g., [36]), the transformation we impose in this dimension is similar in form to those obtained with environmental surfaces and illuminants. The distribution of chromaticities was symmetric about the mean, which passed through the chromaticity of equal-energy white. The extremes of the distribution were fixed at $L/(L + M)$ coordinates of 0.6395 and 0.6695. Fourteen chromatic samples were used, evenly distributed along this line (in steps of 0.0023 on the chromaticity diagram). The appearances of these colors ranged from green (the color with the lowest coordinate) through gray to pink (the color with the highest $L/(L + M)$ coordinate). The 14 chromaticities contributed to two sets. Set A comprised the ten chromaticities with the lowest $L/(L + M)$ coordinates and so had a green bias and a mean $L/(L + M)$ coordinate of 0.6449. Set B comprised the ten chromaticities with the highest $L/(L + M)$ coordinates and so had a red bias and a mean $L/(L + M)$ coordinate of 0.6592. The chromatic difference between the means of our two sets (0.0092 MacLeod–Boynton units) is small in comparison to that of real illuminants. As an example, sunlight and skylight differ by 0.0269 $L/(L + M)$ units and 0.0150 $S/(L + M)$ units in the same color space. The separation between the two chromatic distributions is four units on our scale of 14 chromaticities, such that six samples are common to two sets, and the four greenest samples are contained only in Set A and the four reddest samples are contained only in Set B. The overall mean of the 14 samples used in the experiment lies midway between the chromaticities of the seventh and eighth samples, at position 7.5. The mean of Set A plots at position 5.5, and the mean of Set B plots at position 9.5. In subsequent discussions of the stimuli, we express all chromatic differences relative to the magnitude of the offset between distributions. These relationships are summarized in the top panel of Fig. 2.

3. Experimental Conditions

There were three variables manipulated experimentally, each with two possible values, making eight conditions in total. This combination of variables is complex, but the schematic representation of the different conditions, shown in Fig. 1 (top), can be used as a guide, and the layout of results follows this pattern. Each condition is labeled with a pair of letters, and we use * as a wildcard. The first independent variable was the position of the test patches relative to the fixation target. The test could either be always to one side of the fixation target (conditions C* and D*), or to one or the other side,

depending on the set from which the target chromaticity was drawn (conditions A* and B*). The second variable was the position of the fixation target, which either remained in the center of the display throughout a session (conditions A* and C*), or appeared either above or below the center on each trial, depending on the set from which the target chromaticity was drawn (conditions B* and D*). The intention of these manipulations was to have control over whether the chromaticities from the two sets were retinotopically interleaved (C*, D*) or separated (A*, B*), while also controlling whether the sets were spatiotopically interleaved (C*) or separated (A* separated horizontally, B* separated horizontally and vertically, and D* separated vertically). Test patches were presented at the same distance from fixation in all conditions. Finally, we manipulated the chromaticities of the background ellipse pattern. The test patch could appear either on a background composed only of ellipses with chromaticities drawn from the same set as the test patch (we refer to this as “consistent background” conditions *C), or on a background composed of ellipses with chromaticities drawn from both sets, covering the whole range of 14 chromaticities (we refer to this as the “inconsistent background,” conditions *I). For the conditions in which the test-patch locations were separated spatiotopically, the display was divided in two, horizontally or vertically as appropriate, so that all the chromaticities on one side of the division were drawn from Set A and all the chromaticities on the other side were drawn from Set B. In the one condition where the test patches were spatially and retinotopically interleaved but still required to be embedded in a consistent background, the set of chromaticities used for the background was updated as appropriate from trial to trial. When the background was split, the spatial pattern of ellipses remained continuous across the split. Some examples of the backgrounds to the stimuli are shown in Fig. 2.

We also devised a further condition, EC, identical to DC, in which the test patches from both sets were spatially separated but retinotopically interleaved, except that, instead of moving the fixation target between presentations of test patches from different sets, the image was moved very quickly up or down to align the test patch with the fixation target, which remained in the center. The objective was to simulate the motion of the stimulus over the retina, and so to present a stimulus that was retinotopically similar to that in condition DC over the course of the session, while not requiring the observer to make eye movements. The animation consisted of two frames only, which was approximately the duration of the saccades made by observers in the other conditions.

D. Procedure and Task

In each session, only one combination of the three experimental variables was used. Each of the ten test chromaticities from both of sets A and B were used to color the test patch twice each session, so there were 40 experimental trials per session. The order of presentation of test samples was randomized within each session. Each stimulus (test and background) was displayed for 2 s, and was followed by the next immediately. The observer's task was to classify the test patch as either “red” or “green,” and to indicate their response by pressing one of two buttons. To prevent observers from developing an association between the spatial location of the test patches and the spatial position of the response buttons,

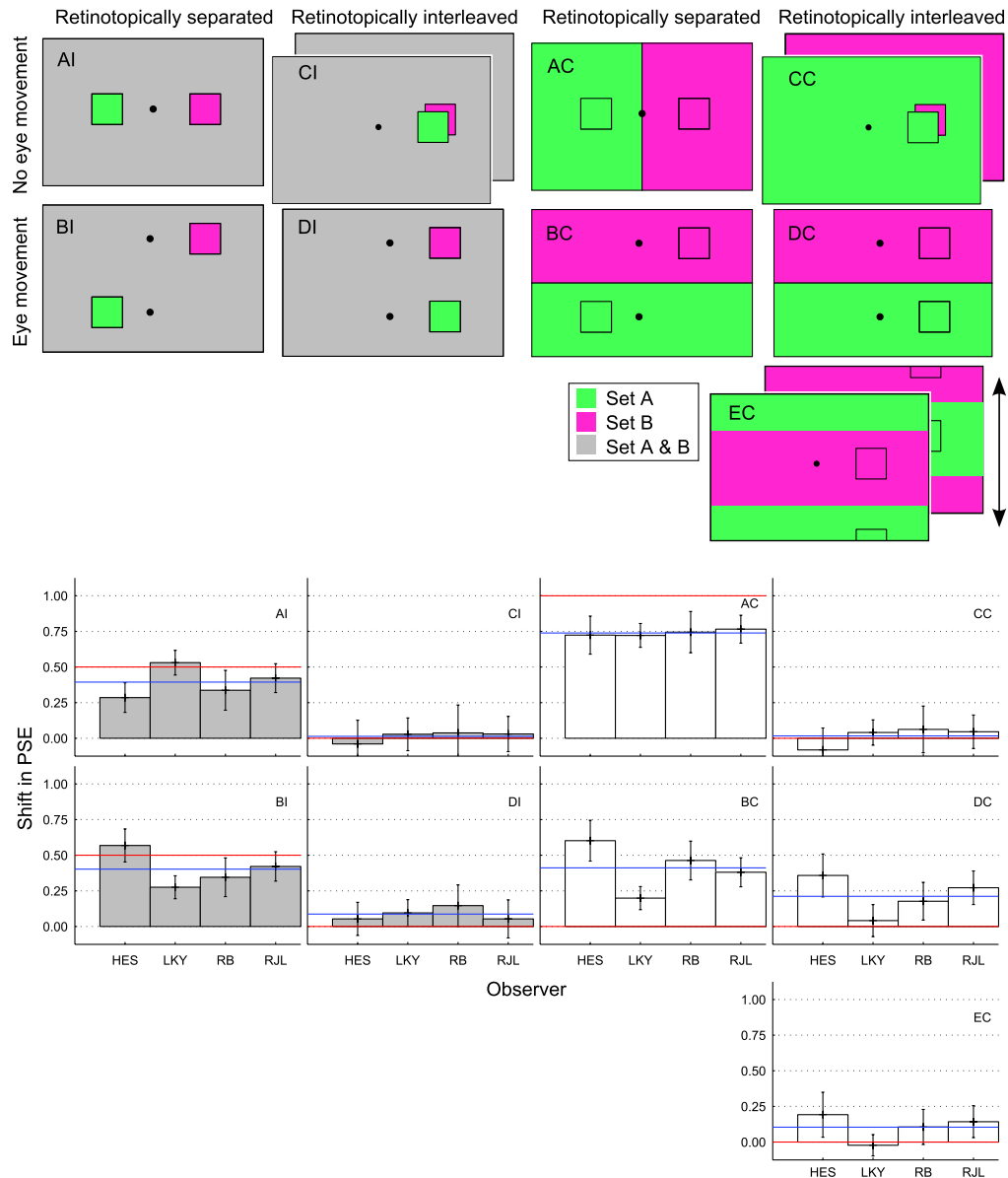


Fig. 1. Top: schematic representation of the stimulus layouts for the nine experimental conditions. The shading indicates the set of chromaticities that were used in each location. The backgrounds were covered with ellipses whose chromaticities were drawn from the set indicated, and examples are shown in Fig. 2. The small squares show test-patch locations. The small black dots show fixation targets. On each trial only a single test patch and the adjacent fixation target were used. Bottom: measured shifts in the point of subjective equality (PSE) for four observers, expressed as a proportion of the difference between the mean chromaticities of the two chromatic sets. Each panel shows data from one condition, as above. Error bars show 95% confidence intervals on the difference between the PSEs derived from fitted psychometric functions. The horizontal blue lines (upper line in all but AI, BI, and AC) indicate the mean shifts across all observers in each condition. The horizontal red lines indicate the shifts predicted if the PSE is determined by the time-averaged chromaticity displayed at the retinotopic location of the test patch.

the sets associated with the two locations were reversed for half of the occasions that each session was run. Observers had the entire length of the 2 s stimulus presentation to respond, but the next trial was presented whether or not a response was recorded. Ten additional stimuli, randomly chosen from those in the session, were presented at the start of the session. In a previous study, we showed that, after a change in illumination conditions, color classification reaches steady-state levels within approximately ten 2 s trials [27]. The responses to these ten additional presentations were discarded. The order in which sessions from the eight conditions were run was counterbalanced. In general, data were collected in blocks of four consecutive sessions.

The observers were instructed to fixate the target dot throughout the session, and were told that this would require an eye movement in the experimental conditions in which the fixation target could appear in one of two locations. It should be noted that, because the order of stimulus presentation within a session was randomized, the observer could not predict the location of the target on any given trial. To ensure that the observers were fixating as instructed, we monitored and recorded their gaze direction with an Eyelink 1000 (SR Research, Ontario, Canada). For subsequent analyses, we discarded responses to stimuli for any trials in which the observer was not looking within 20 pixels ($\sim 1^\circ$) of the fixation target for at least 75% of the presentation duration.

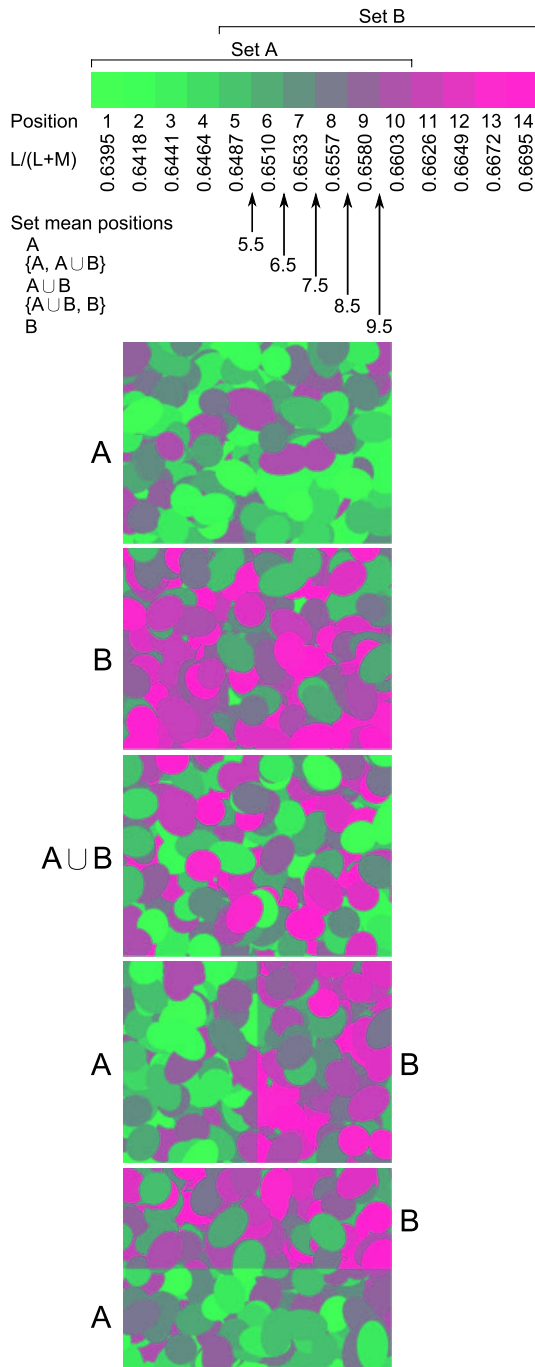


Fig. 2. Top plot shows the chromaticities used in the experiment. Means of important sets of chromaticities are indicated, and are used to make predictions based on adaptation to the time-averaged chromaticity in different conditions. The images below are examples of the spatial layout of our stimulus surrounds. The top image represents a surround composed of chromaticities from the green-biased set, and the one below is composed of chromaticities from the red-biased set. The center image is made of chromaticities from both sets. The bottom two images show how the surround pattern was divided so that one-half of the display was composed of green-biased chromaticities and the chromaticities in the other half were offset to make them red-biased.

The four observers who obtained complete datasets each completed 144 sessions, giving, in total, the opportunity for 32 classifications of each test chromaticity from each set in

each condition. However, some of these classifications were discarded because the observer was not correctly fixating, and in some trials the observer failed to make a response within the time allowed. The additional observers obtained the same number of classifications per test chromaticity, but only in two conditions: retinotopically interleaved but spatially separated test patches on consistent background (DC), and the extra, animated, saccade-simulation condition (EC).

3. RESULTS

A. Psychometric Functions

Our aim is to determine whether judgments are systematically different for samples drawn from Set A compared to those drawn from Set B, even though six of those samples share the same chromaticity. From our data we extract, for each test sample, the proportion of valid responses in which the sample was classified as “red.” The proportion of classifications that were rejected because we were unable to confirm that the observer’s gaze direction was within the defined limits varied from observer to observer, but, on average, was approximately 10%. The proportion of trials on which a response was not made was less variable across observers, and was, on average, 1.9%.

Psychometric functions (of a logistic form) were fitted to the proportions of “red” classifications of each chromaticity using the `psignifit` toolbox version 2.5.6 for MATLAB (see <http://bootstrap-software.org/psignifit/>), which implements the maximum-likelihood method described by Wichmann and Hill [37]. Psychometric functions are characterized by their position on the decision axis, and their slope. The most robust differences in our data were in position, so the first stage in all analyses was to determine the chromaticity that was equally likely to be judged as red or green—the point of subjective equality (PSE). We determined the chromaticity at which the fitted function crossed 0.5, representing an equal probability of “red” or “green” classification. We use the difference between the PSE obtained for stimuli drawn from Set A and the PSE for stimuli drawn from Set B as our metric. If these values are the same, we have no evidence that observers maintain separate references for the two chromatic contexts; if these values show a positive difference, we have evidence that observers maintain a redder reference for the red-biased set than for the green-biased set. In the following analyses we use the chromaticity difference in PSE for the green- and red-biased sets [$r_B - r_A$ when $f(r_A) = f(r_B) = 0.5$, where $f(r_A)$ and $f(r_B)$ are the psychometric functions fitted to the classification probabilities for sets A and B, respectively].

B. Shift in PSE

The chromatic shifts between PSEs derived for the two stimulus sets are shown in the bottom half of Fig. 1, with a separate plot for each condition, arranged to match the layout of the top half of the figure. Each panel shows data from the four observers who completed all conditions. The shift in the PSE is expressed as proportion of the chromatic difference between the means of the green- and red-biased sets and so can be interpreted in a similar way to a color-constancy index or Brunswick ratio (where 1.0 indicates perfect constancy; see [38] for a definition of constancy indices and Brunswick ratios). There are clear trends across observers in the size of the chromatic shifts in the PSEs in different conditions and the

blue horizontal lines on each panel indicate the means across observers.

We now consider specific conditions, and comparisons between pairs of conditions. Condition CI is effectively a control condition, for this arrangement offered no cues to separate the two sets of chromaticity samples. The test patches are retinotopically and spatiotopically interleaved, so neither a spatiotopically nor retinotopically localized temporal average serves to distinguish the sets, and samples are presented on inconsistent backgrounds, so there is additionally no spatial context to offer a cue. While it is possible that an observer's perceptual reference of "neither red nor green" might drift over the course of a session, there is no reason to expect that it should differ systematically for samples that the experimenter had labeled as belonging to Set A compared to samples belonging to Set B. The differences in PSEs for the two sets show that this is indeed the case. Differences are very close to zero, and the error bars on our estimates confirm that the PSEs for the two psychometric functions for each observer are indistinguishable within the 95% confidence intervals.

It is possible that the differences in the means of the two sets are too small to reliably influence the PSE. However, data from condition AC, in which the test patches are retinotopically and spatiotopically separated and presented on consistent backgrounds, indicate that this is not the case. All observers show similarly sized shifts in this condition, of a magnitude of approximately 0.75. A shift in the PSE of 1.0 would indicate that an observer had completely compensated for the difference between the two sets, making judgments that were internally consistent within one or another context. As mentioned above, the values we plot can be interpreted as a constancy index or Brunswick ratio, and the values observed in this condition are comparable to those seen in constancy experiments using judgments of similar stimuli (see "Color naming and related methods" section of Table 1 in [38] for examples).

To determine the effect of placing test patches on backgrounds composed of chromaticities from the corresponding set, we can compare conditions CI and CC, which differ only in that regard. We have just noted the absence of any significant shift in condition CI, and this is still true in CC. This result indicates that adding a consistent background is insufficient by itself to promote the use of different references for judgments of stimuli from the two sets.

To determine the effect of placing the test patches from different sets in different retinotopic locations, we make four comparisons: AI compared to CI, BI compared to DI, AC compared to CC, and BC compared to DC. In conditions C* and D*, stimuli from Set A and from Set B are presented at the same retinal location. In conditions A* and B*, stimuli from the two sets appear in retinotopically different locations, to the left and right of fixation (and additionally in different locations in space). All four observers show smaller shifts when stimuli are interleaved than when they are retinotopically separated.

In conditions C* and D*, stimuli from the two sets are retinotopically interleaved, and the time-averaged chromaticity at this location is the overall mean of the two sets. Although conditions A* and B* all maintain retinotopic separation of the two stimulus sets, they are not matched in the time-averaged chromaticity they present. Our decision to present only one stimulus per trial (so that the other stimulus stream cannot

itself be used as a reference), means that the time-averaged chromaticity at the test locations in conditions AI and BI will include some of the inconsistent background. So, for the Set A location, the mean will be 6.5 (rather than 5.5 for Set A alone) on our scale of 14 chromaticities, and for the Set B location, the mean will be 8.5 (compared to 9.5 for Set B alone). For condition AC, the time-averaged chromaticities at the two test locations are 5.5 and 9.5, and for condition BC, the time-averaged chromaticities at the two test locations are identical, corresponding to the overall mean of 7.5. The red horizontal lines on the plots in Fig. 1 indicate the predicted separation of the PSEs if this shift were determined only by retinotopic differences in the time-averaged chromaticity. The reduction in separation of the PSEs for condition AI compared to AC is between 30% and 60%. A reduction of 50% can be attributed to the reduced difference in time-averaged chromaticity in the two locations (2 units difference compared to 4 units difference). A reduction of more than 50% indicates a contribution of the spatial surround in condition AC, and a reduction of less than 50% indicates some preference for adaptation to test samples rather than background in condition AI. Conditions BC and DC both present identical time-averaged chromaticities in the two test locations, and are matched in the availability of cues from the spatial surround. The increased separation in the PSEs seen in condition BC must, therefore, be attributed to the additional separation of the two test streams.

The purpose of stimulus configurations in which we required the observer to saccade between two locations was to allow us to present the test patches from different sets in spatially separated locations on the display but to interleave them retinotopically. To consider the effect of spatiotopic separation independently from retinotopic adaptation, we compare conditions that are matched in spatial context and in temporal adaptation at the test locations. These are the pairs of conditions: AI and BI, CI and DI, and CC and DC (but not BC and AC, since in BC the saccade causes the time average to mix Set A and Set B at each location). Conditions AI and BI support a similar difference in the PSEs between sets. As reported above, conditions CI and CC show no difference in the PSEs between sets. Data are inconclusive regarding the shift in the PSE in condition DI, since confidence intervals include zero for two out of the four observers. Condition DC elicits a shift in PSE for three of our four original observers. We have already noted that, when temporal adaptation is matched at the two test locations, the spatial separation and consistent background cues individually do not produce reliable shifts in PSE. However, it appears that the combination of these cues (condition DC) may be sufficient.

We specifically investigated the effect of executing saccades by comparing conditions DC and EC, which are retinotopically equivalent. In condition DC the observer made vertical saccades, whereas in EC the display was animated to simulate the motion of the image across the retina while the observer maintained fixation in the center of the display. We see a slight reduction in the size of the measured shifts in PSE with simulated compared to real saccades, but the shifts remain significantly different from zero for two observers.

Because the shifts in PSEs that we obtained in conditions DC and EC were significantly different from zero for only a subset of our observers, we repeated these conditions with

four new observers. Data for the full set of eight observers who participated in these conditions are shown in Fig. 3. All four of the new observers showed significant separation between judgments of the two chromatic sets, for both conditions.

C. Sensitivity Adjustments

An estimate of sensitivity to differences in chromaticity can be obtained from the slope of the psychometric function. It is possible that the variance of the set of stimuli against which stimuli are judged might influence sensitivity to chromatic differences. We compare slope estimates derived from psychometric functions obtained in condition CI (where there is no difference in PSE, and the two sets are effectively combined) and in condition AC (where there is a large difference in PSE, and chromaticities from the two sets are judged separately). Slope estimates for four observers for green-biased and red-biased stimulus sets are plotted in Fig. 4. There is some tendency for red-biased sets to elicit steeper psychometric functions (indicating higher discrimination sensitivity) than green-biased sets, and for the condition AC to generate steeper functions than condition CI, which is consistent with an improvement in discrimination when judgments are made within a consistent context of relatively low variance. However, the errors on these estimates are high, and any differences are at the limit of what we can reliably measure. It is not possible to determine differences between other conditions.

D. Temporal Context

As part of our investigation into the degree to which temporal consistency improves the maintenance of different references for the two sets, we calculated the shifts in PSE after separating the classifications into those made when the previous test chromaticity was from the same set and those made when the previous test chromaticity was from the other set. We did this for all eight observers, for the conditions in which all eight obtained data. The shifts are shown in Fig. 5. A two-way repeated-measures ANOVA on the data indicate no significant effect of the condition ($F_{(1,7)} = 2.91$, $p = 0.13$), but a significant difference between the data obtained from trials in which the previous test chromaticities were the same or different

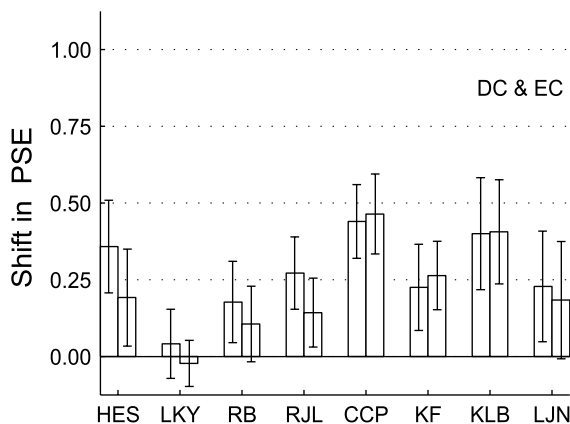


Fig. 3. Measured shifts in the PSE, as in Fig. 1, for all eight observers in the two conditions in which all observers collected data: DC and EC. Data from both conditions are on the same plot, with data from condition DC shown by the left-hand bar in each pair, and the data from condition EC shown by the right-hand bar. Error bars show 95% confidence intervals, as before.

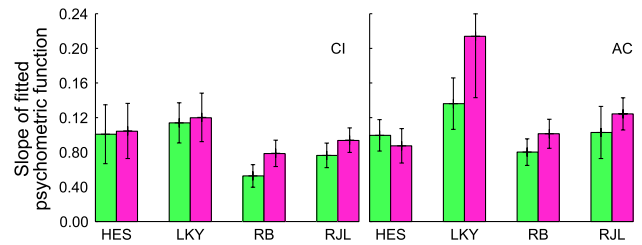


Fig. 4. Slopes of the fitted psychometric functions at $f(r_A) = 0.5$ and $f(r_B) = 0.5$ for conditions CI (left panel) and AC (right panel). Error bars again show 95% confidence intervals on these fitted values. Each pair of bars shows the slopes for one observer, and within each pair the green and pink bars show the slope of the curves fitted to the classification probabilities for the green- and red-biased sets, respectively.

($F_{(1,7)} = 9.49$, $p = 0.02$). Larger shifts are obtained when the same set is used for at least two consecutive trials. This implies that evidence about chromatic context is accumulated over trials.

4. DISCUSSION

We discuss the circumstances in which we see differences between the chromaticities of the points of subjective equality in the red–green judgments of samples from the two sets. It is clear from the first comparison that we make in the results, between the condition with consistent backgrounds and test patches that are spatiotopically and retinotopically separated and the condition with none of those features, that we have succeeded in generating situations when all observers make reliably different judgments about stimuli from the two sets, and situations when they do not. In further comparisons, we

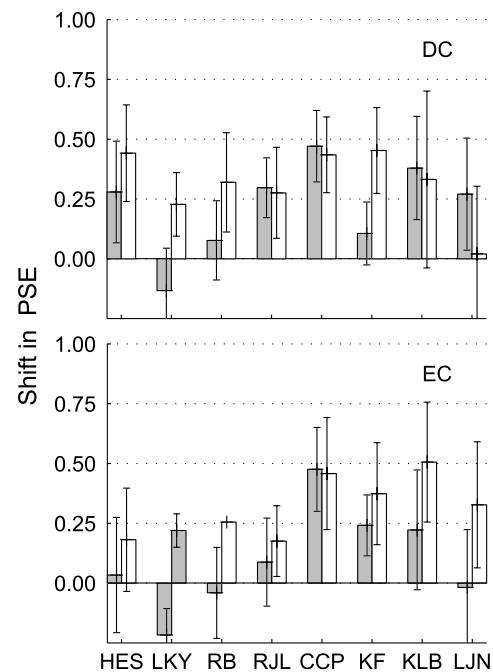


Fig. 5. Calculated shifts in the PSE for conditions DC (top plot) and EC (bottom plot). Within each pair of bars, the left-hand (gray) bar indicates the shift calculated using only classifications made on presentations in which the test chromaticity was from the set not used in the previous presentation. The right-hand (white) bar indicates the shift calculated using only classifications made on presentations in which the test chromaticity was from the same set as in the previous presentation. Error bars show 95% confidence intervals, as before.

consider individual features of the stimulus to pin down the effect of each.

A. Instantaneous Constancy

The addition of a spatial surround with chromatic statistics that are defined by the set from which the test sample is drawn is not sufficient to produce a displacement in the PSEs derived from the two sets. The chromaticities surrounding the test patch are insufficient to specify how the test chromaticity should be judged. This might seem surprising in relation to the well-studied effects of color induction, in which the surround chromaticity has a strong effect on the chromatic appearance of a target stimulus. However, these effects tend to be strongest with uniform [39] or regularly patterned fields [40] and not with randomly varying patterns like the ones we used here. The strength of color induction depends on the bias in the inducer, and it may be that the chromatic displacement between the two sets was insufficient to elicit strong spatial effects, although it was of a magnitude that elicited robust effects in other conditions. A more plausible explanation for the lack of difference here is the time course of the stimulus presentation. An instantaneous shift in the mean chromaticity of the surround is not sufficient to signal that the reference should be changed. Color-constancy experiments, however, have shown that observers can be instantaneously aware of a change in the global statistics of a scene that are associated with an illumination change, and can behaviorally identify violations in this global change [41]. Whether or not this ability can be called color constancy might be debated [42], but it does suggest that observers should be able to respond to the global set-change in our experiment. There is also evidence that color-constancy mechanisms are active even when the observer is unaware of the change in illumination [43]. However, we have evidence from a separate study [27] that observers do not make immediate adjustments to their achromatic point after a simulated change in illumination. The lag on updating appearance judgments implies that the perceptual reference that is used depends on a time-extended process. It also implies that coding relative to the statistics of spatially distributed samples is not achieved instantaneously, or at least that, even if a relative signal is extracted, the absolute signals are not discarded.

B. Adaptation

The results indicate a bigger shift in the PSE in all conditions in which test patches from the two sets are retinotopically separated than in the equivalent conditions in which the test patches are interleaved on the same position on the retina. The relative differences between conditions are predicted in large part by the relative differences in time-averaged chromaticities for the locations in which stimuli from the two sets are presented. This is the most robust of our findings. In the context of multiple references, this implies that the most effective way in which the visual system keeps track of different chromatic references is by assigning a different reference to a different region of retinotopic space. However, this cannot be the only method of maintaining more than one reference, since we also show that symbolic cues are effective to some degree (see Subsection 4.C).

The finding that observers use different references for different regions of retinotopic space could be explained by any process of adaptation that renormalized signals to a retinoto-

pically localized mean. As discussed in the introduction, chromatic adaptation can be highly localized retinotopically, at about the scale of individual photoreceptors. We do not wish to suggest, however, that the adaptation underlying our behavioral results is confined to the retina, as chromatic adaptation has also been demonstrated at later stages in the system [7–9]. Our results simply emphasize the importance of different adaptation levels for different parts of the retina. To first order, the effect of adaptation would be to normalize the responses associated with a particular retinotopic region to the average chromaticity of the stimulus in the corresponding location. So, in our experimental conditions, the retinotopic regions corresponding to the green-biased stream of test patches will be green adapted and those corresponding to the red-biased set will be red adapted. After adaptation, the averages of the green and red sets will elicit similar neural responses in the corresponding separated regions of the retina, or in the later neural mechanisms that correspond to those regions.

Since retinotopic chromatic adaptation can be highly localized, this mechanism could maintain a very great number of references, if each had a different retinotopic location. Morgan *et al.* [32] demonstrated that as many as eight references could be maintained for line separation, and the resolution at which differential chromatic adaptation is possible is even finer than the size required for his stimuli. However, any eye movements reduce this resolution or render it ineffective for anything other than the average chromaticity of the field sampled by successive fixations.

The difference we observe between the separation of the PSEs in conditions AI and AC is roughly consistent with the fact that the time-averaged chromaticities at the two locations differ by only half as much in condition AI as in condition AC. Similarly, the relatively small shifts we observe in conditions C* and D* might be expected since, in these conditions, there is no difference in time-averaged chromaticities at the retinal location used for the test patches. However, these shifts are not all zero, and for condition BC, which again is matched in terms of time-averaged chromaticity at the retinal locations of the test patches, we see relatively large shifts. Additional factors are required to account for these shifts.

C. Symbolic Cues

A retinotopically localized time-averaged signal is clearly a strong determinant of color judgments. However, a constancy mechanism that is driven purely by such signals is vulnerable. First, such a mechanism will fail to recognize chromatic bias in reflectance samples in different regions of the visual field (e.g., forest or ground, which provide strong counterexamples to the gray-world heuristic [44]). Second, such a mechanism will confound different regions of illumination when eye movements convert spatially distributed signals into retinotopically localized ones. Third, such a mechanism will mix chromaticity samples with different characteristics if self-motion or motion of objects or illuminants cause movement in the retinal image. These issues can be probed by drawing comparisons between some of our stimulus conditions. The increased separation of PSEs obtained in condition BC compared to DC implies that, even when retinotopic adaptation is controlled, a difference in the spatial location of two test streams contributes to the maintenance of separate standards. The

simultaneous maintenance of separate perceptual references on the basis of symbolic cues has previously been shown for judgments of line separation and orientation. The results we present here indicate that the spatial location of a stimulus stream can be used to indicate separate perceptual references for color, even when the time-averaged chromaticity is controlled so that retinotopic adaptation does not additionally vary with the stream from which samples are drawn.

Although our experiment has parallels with the task of estimating surface color for illuminated surfaces, our stimuli are very impoverished compared to real-world scenes. We gave no instructions other than asking observers to classify the appearance of the test patch as reddish or greenish. This is the most conservative way to test the potential utility of “symbolic cues” in setting perceptual references for color. Such cues could be important in maintaining perceptual constancy during visual exploration of scenes with multiple regions of illumination and regions of chromatically biased reflectances.

D. Eye Movements

In conditions DI and DC (compared to CI and CC, respectively) there is a hint that a saccade across an illumination boundary does not completely confound stimuli from the two sets, even though they are retinotopically interleaved. We repeated conditions DC and EC with additional observers and found reliable separation of the PSEs obtained with the two stimulus sets despite retinotopic interleaving of samples from the two sets. Comparison between conditions DC and EC indicates that the execution of a saccade is not itself critical for this perceptual separation.

Cornelissen and Brenner [45] measured observers' eye movements while they performed an asymmetric color match between test patches embedded in Mondrian displays under different illuminants. Their observers looked back and forth between the two test patches immediately before making a match. One of the aims of their study was to determine whether differences between settings obtained for a “paper match” and those obtained for a “hue match” could be predicted by differences in looking behavior and local chromatic adaptation. They found that differential eye movements in conjunction with adaptation had some influence on observers' settings, but that there was a strong influence of instruction over and above these effects (their Figure 9). So, it is the case that observers move their eyes back and forth between different regions of illumination before making an asymmetric color match, and more so for a “paper match” than a “hue match,” but looking across an illumination boundary is not causal in generating the difference in observers' settings. In a later study, Brenner *et al.* [46] specifically investigated the relevance for color appearance of where one fixates, and they found that the perceived color of a surface depends heavily on what one looked at last. So their results, like ours, indicate that local adaptation to the recent stimulus history is important, but that additional factors can further influence observers' judgments.

ACKNOWLEDGMENTS

The experimental data collection took place at the Department of Psychology, Durham University, UK, to which the authors were affiliated. This work was supported by Wellcome Trust grant WT094595AIA to H. E. Smithson.

REFERENCES

1. M. J. Morgan, “On the scaling of size judgements by orientational cues,” *Vis. Res.* **32**, 1433–1445 (1992).
2. H. Helson, “Adaptation-level as frame of reference for prediction of psychophysical data,” *Am. J. Psychol.* **60**, 1–29 (1947).
3. H. Helson, “Adaptation-level as a basis for a quantitative theory of frames of reference,” *Psychol. Rev.* **55**, 297–313 (1948).
4. P. Lennie and M. D’Zmura, “Mechanisms of color vision,” *Crit. Rev. Neurobiol.* **3**, 333–401 (1988).
5. J. von Kries, “Beitrag zur Physiologie der Gesichtsempfindungen [Transl. Physiology of visual sensations],” in *Sources of Color Science*, D. L. MacAdam, ed. (MIT Press, 1878), pp. 101–108.
6. A. Stockman, M. Langendörfer, H. E. Smithson, and L. T. Sharpe, “Human cone light adaptation: from behavioral measurements to molecular mechanisms,” *J. Vision* **6**(6), 1194–1213 (2006).
7. E. N. J. Pugh and J. D. Mollon, “A theory of the P₁ and P₂ color mechanisms of stiles,” *Vis. Res.* **19**, 293–312 (1979).
8. J. Krauskopf, D. R. Williams, and D. W. Heeley, “Cardinal directions of color space,” *Vis. Res.* **22**, 1123–1131 (1982).
9. C. Tailby, S. G. Solomon, N. T. Dhruv, and P. Lennie, “Habituation reveals fundamental chromatic mechanisms in striate cortex of macaque,” *J. Neurosci.* **28**, 1131–1139 (2008).
10. A. Kohn, “Visual adaptation: physiology, mechanisms, and functional benefits,” *J. Neurophysiol.* **97**, 3155–3164 (2007).
11. I. J. Murray, A. Daigirdiene, R. Stanikunas, H. Vaitkevicius, and J. J. Kulikowski, “Cone contrasts do not predict color constancy,” *Vis. Neurosci.* **23**, 543–547 (2006).
12. M. A. Webster and J. A. Wilson, “Interactions between chromatic adaptation and contrast adaptation in color appearance,” *Vis. Res.* **40**, 3801–3816 (2000).
13. D. Katz, *The World of Colour* (K. Paul, Trench, Trubner, 1935).
14. B. H. Crawford, “Visual adaptation in relation to brief conditioning stimuli,” *Proc. R. Soc. Lond. Ser. B* **134**, 283–302 (1947).
15. D. Jameson, L. M. Hurvich, and F. D. Varner, “Receptoral and postreceptoral visual processes in recovery from chromatic adaptation,” *Proc. Natl. Acad. Sci. USA* **76**, 3034–3038 (1979).
16. J. Neitz, J. Carroll, Y. Yamauchi, M. Neitz, and D. R. Williams, “Color perception is mediated by a plastic neural mechanism that is adjustable in adults,” *Neuron* **35**, 783–792 (2002).
17. P. B. Delahunt, M. A. Webster, L. Ma, and J. S. Werner, “Long-term renormalization of chromatic mechanisms following cataract surgery,” *Vis. Neurosci.* **21**, 301–307 (2004).
18. D. H. Foster and S. M. C. Nascimento, “Relational colour constancy from invariant cone-excitation ratios,” *Proc. R. Soc. Lond. Ser. B* **257**, 115–121 (1994).
19. H. E. Smithson, “Sensory, computational and cognitive components of human colour constancy,” *Philos. Trans. R. Soc. Lond. Ser. B* **360**, 1329–1346 (2005).
20. D. I. A. MacLeod and S. He, “Visible flicker from invisible patterns,” *Nature* **361**, 256–258 (1993).
21. S. He and D. I. A. MacLeod, “Local nonlinearity in S-cones and their estimated light-collecting apertures,” *Vis. Res.* **38**, 1001–1006 (1998).
22. D. I. A. MacLeod, D. R. Williams, and W. Makous, “A visual nonlinearity fed by single cones,” *Vis. Res.* **32**, 347–363 (1992).
23. S. J. Schein and R. Desimone, “Spectral properties of V4 neurons in the macaque,” *J. Neurosci.* **10**, 3369–3389 (1990).
24. A. C. Hurlbert and T. A. Poggio, “Synthesizing a color algorithm from examples,” *Science* **239**, 482–485 (1988).
25. J. D. Mollon, B. C. Regan, and J. K. Bowmaker, “What is the function of the cone-rich rim of the retina?” *Eye* **12**, 548–552 (1998).
26. M. D’Zmura and P. Lennie, “Mechanisms of color constancy,” *J. Opt. Soc. Am. A* **3**, 1662–72 (1986).
27. R. J. Lee, K. A. Dawson, and H. E. Smithson, “Slow updating of the achromatic point after a change in illumination,” *J. Vision* (to be published).
28. H. E. Smithson and Q. Zaidi, “Colour constancy in context: Roles for local adaptation and levels of reference,” *J. Vision* **4**(8), 693–710 (2004).
29. A. D. D’Antona and S. K. Shevell, “Induced steady color shifts from temporally varying surrounds,” *Vis. Neurosci.* **23**, 483–487 (2006).

30. K. Koffka, *Principles of Gestalt Psychology* (Harcourt, Brace, and World, 1935).
31. A. Gelb, "Die Farbenkonstanz der Sehdinge," in *Handbuch der normalen und pathologischen Psychologie*, A. Bethe, G. V. Bergmann, G. Embden, and A. Ellinger, eds. (Springer-Verlag, 1929), pp. 594–678.
32. M. J. Morgan, S. N. Watamaniuk, and S. P. McKee, "The use of an implicit standard for measuring discrimination thresholds," *Vis. Res.* **40**, 2341–2349 (2000).
33. S. X. Xian and S. K. Shevell, "Changes in color appearance caused by perceptual grouping," *Vis. Neurosci.* **21**, 383–388 (2004).
34. D. I. A. MacLeod and R. M. Boynton, "Chromaticity diagram showing cone excitation by stimuli of equal luminance," *J. Opt. Soc. Am.* **69**, 1183–1186 (1979).
35. A. Stockman and L. T. Sharpe, "The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype," *Vis. Res.* **40**, 1711–1737 (2000).
36. J. Golz and D. I. A. MacLeod, "Influence of scene statistics on colour constancy," *Nature* **415**, 637–640 (2002).
37. F. A. Wichmann and N. J. Hill, "The psychometric function: I. Fitting, sampling, and goodness of fit," *Percept. Psychophys.* **63**, 1293–1313 (2001).
38. D. H. Foster, "Color constancy," *Vis. Res.* **51**, 674–700 (2011).
39. Q. Zaidi, B. Spehar, and J. DeBonet, "Color constancy in variegated scenes: role of low-level mechanisms in discounting illumination changes," *J. Opt. Soc. Am. A* **14**, 2608–2621 (1997).
40. P. Monnier and S. K. Shevell, "Large shifts in color appearance from patterned chromatic backgrounds," *Nat. Neurosci.* **6**, 801–802 (2003).
41. D. H. Foster, S. M. C. Nascimento, K. Amano, L. Arend, K. J. Linnell, J. L. Nieves, S. Plet, and J. S. Foster, "Parallel detection of violations of color constancy," *Proc. Natl. Acad. Sci. USA* **98**, 8151–8156 (2001).
42. D. H. Foster, "Does colour constancy exist?," *Trends Cognit. Sci.* **7**, 439–443 (2003).
43. J. L. Barbur and K. Spang, "Colour constancy and conscious perception of changes of illuminant," *Neuropsychologia* **46**, 853–863 (2008).
44. R. O. Brown, "The world is not gray," *Investig. Ophthalmol. Vis. Sci.* **35**, 2165 (1994).
45. F. W. Cornelissen and E. Brenner, "Simultaneous colour constancy revisited: an analysis of viewing strategies," *Vis. Res.* **35**, 2431–48 (1995).
46. E. Brenner, J. J. M. Granzier, and J. B. J. Smeets, "Perceiving colour at a glimpse: the relevance of where one fixates," *Vis. Res.* **47**, 2557–2568 (2007).