Is Color Constancy Task Independent?

Jon M. Speigle and David H. Brainard
Department of Psychology
University of California, Santa Barbara, California

Abstract

We present data for which the effect of context on color appearance is independent of what task is used to measure appearance.

Introduction

The goal of color appearance models[1-3] is to predict how observers perceive, describe, and match colors. Two difficulties prevent the easy formulation of these models. First, the color appearance of a light can depend strikingly on the context in which it is viewed. Thus color appearance models must incorporate transformations to account for context effects. Second, there are multiple ways that observers can assess and report color appearance. A complete model must predict performance for all appearance tasks.

Color appearance models can be simplified if the effects of context may be separated from how appearance is measured. One way to achieve such separation is to formulate the model in two distinct stages. The first stage incorporates context effects by transforming tristimulus coordinates to *appearance coordinates*. This transformation is allowed to depend on viewing context but not on how appearance is measured. The second stage maps between the appearance coordinates and actual observer responses. Different mappings are allowed for different tasks, but each mapping is restricted to be context independent. The RLAB color appearance model is consistent with this type of separability.[3]

Separating the effect of context from task specific mappings allows model parameters to be determined economically. The effects of context need only be measured using one task, while the form of the task dependent mappings need only be characterized for one context. Of course, it only makes sense to incorporate such *context-task separability* only if models of this form actually describe performance. In this paper we report initial empirical tests.

Introspection suggests that context-task separability is likely to hold. As we experience the world we are generally aware of a unitary color percept at each image location. It is appealing to think that this subjective impression mediates our performance on all color tasks. There are several reasons why we feel an empirical check is desirable. First, although introspection is useful it is important to build our models on the firmer foundations of experimental data. Second, there are conditions under which cognitive factors such as instructions[4] or perceptual set[5] can influence color judgments. The influence of cognitive factors may be both context and task dependent and their influence could provide a mechanism through which separability could fail. Third, Troost and DeWeert[6] have presented data which, while not conclusive, call the separability hypothesis into doubt.

In our experiments, we use three tasks to measure the effect of changing the illuminant on color appearance. In the first task, observers adjusted a surface to appear achromatic. In the second task, observers performed asymmetric matches. In the third task, observers named colors. We used the same observers and stimulus conditions for all three tasks. We then asked whether the effect of the illuminant was independent of task.

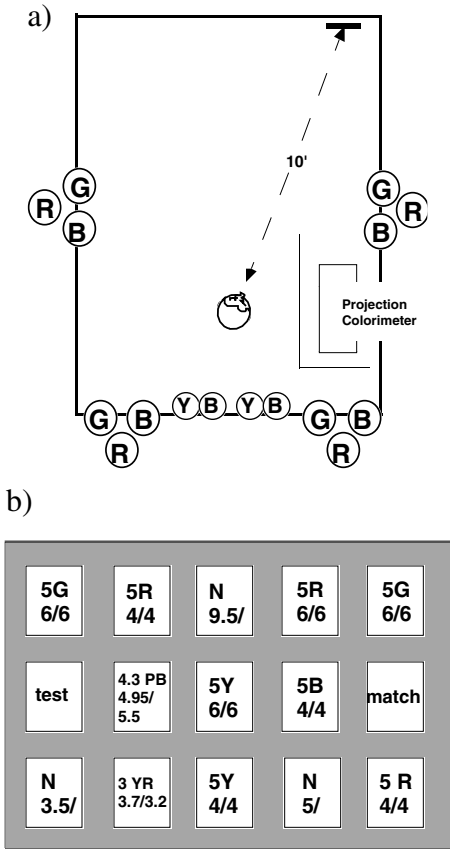

Figure 1. a) Plan view of the experimental room. b) Schematic representation of the observer's view of the far wall of the experimental room.

Methods

Apparatus

The apparatus for our experiments is an entire experimental room, depicted in Figure 1. The room is painted a neutral grey and is diffusely illuminated by triads of chro-

matically filtered stage lamps (labeled RGB). A second bank of directional lights (labeled BY) can create illumination gradients on the far wall of the room. The lights are under computer control and allow setting a broad range of illumination conditions. In all tasks, the observer judged the appearance of an 8.5" by 11" *match* panel located on the far wall of the room (see figure). The match panel is spot illuminated by a projection colorimeter which is also under computer control. During the asymmetric matching task, various panels were placed at the *test* panel position. During the achromatic adjustment and naming experiments the test panel was set to Munsell paper 2.2 YR 6.47/4.1.

The colorimeter is calibrated to take into account the ambient component of illumination reflected from the match panel as well as the shift in bulb spectra as a function of intensity. Calibration measurements allow us to specify the illuminant and the light reaching the observer from the match panel in CIE tristimulus coordinates. Despite the apparatus calibration, all data reported in this paper are the result of direct spectral measurements of the stimuli.

Illumination Conditions

We tested a pair of illumination conditions which approximated direct and indirect sunlights. The measured CIE 1931 chromaticities and luminances at the test and match positions are given in Table 1. In the *bluish* condition the entire room was set to roughly the same diffuse illumination. In the *gradient* condition, the color of the illuminant at the test position was the same as in the bluish condition, while at the match position it was yellowish. The asymmetric matching task was conducted under the gradient illuminant, with the observer producing matches to stimuli across the illumination gradient. For the achromatic adjustment and naming tasks the experiments were repeated for the two illuminant conditions. The observer adapted to the illumination condition at the beginning of each session for five minutes.

Table 1. CIE xyY Coordinates of the Experimental Illuminants.

	x	y	Y
bluish, left	0.342	0.392	23.8
bluish, right	0.342	0.373	19.1
gradient, left	0.344	0.371	20.0
gradient, right	0.452	0.403	84.9

Achromatic Adjustment

Observers adjusted the chromaticity of the match panel until it appeared achromatic.[7-9] Stimulus luminances were 1-17 cd/m^2 in steps of 2 cd/m^2 under the bluish illuminant and 3-15 cd/m^2 in steps of 2 cd/m^2 under the gradient. Each trial began with the colorimeter initialized to a random, within-gamut color. Four replications were made for each luminance/illumination condition. At the end of each session, the observer's settings were measured directly with a spectraphotometer.

Asymmetric Matching

Test panels were placed on the left side of the far wall of the room as shown in Figure 1. The observer was instructed to adjust the match panel until its appearance was the same as that of the test panel. The adjustments were made using three knobs, each of which controlled the output of a different colorimeter channel. The observer also indicated the quality of the final match. When the match was satisfactory, both the test and match spectra were measured by placing a spectraphotometer at the position of the observer. When the match was unsatisfactory, the experimenter replaced the match panel with an alternative likely to shift the colorimeter gamut to include the test color. If a satisfactory match was still unattainable, the data for the particular test panel were discarded. For each observer, matches were made to approximately thirty different Munsell panels with 2-3 replications per panel.

Color Naming

Our color naming method utilizes the eleven basic color categories: red, yellow, green, blue, orange, purple, pink, brown, white, grey, and black.[10] A number of psychophysical studies have examined the properties of these color terms. Boynton and colleagues' work used the OSA uniform color space as the stimulus set.[11-18] In our experiment, observers viewed stimuli generated by the colorimeter and described them using a series of color names and ratings.[19] were instructed to "rate how good an *X* the stimulus is", where *X* was each of the eleven color category names. For each stimulus the observer named and rated each of the eleven color names which applied. The ratings were from 0-9. A response of "0" indicated that the stimulus did not represent the color name at all. A response of "1" indicated that the stimulus was an extremely poor example and a response of "9" indicated that it was an extremely good example. Observers were asked to use the full rating scale range.[20]

Observers entered their responses using a keyboard which had keys labeled with the names of the eleven basic color terms. After indicating a name, the rating was assigned using the numeric keypad. Before advancing to the next trial, the observer's entries were read back using a speech synthesizer and the observer was given the opportunity to correct any recording errors.

The stimulus set sampled the colorimeter gamut at 0.01 intervals in x and y chromaticity and at 2 cd/m^2 luminance intervals. The gamut was extended by using different Munsell surfaces as the match panel. The complete stimulus sets for the naming task consisted of 300-400 distinct stimuli which were replicated 2-14 times. The variable number of replications resulted from the overlap in gamuts for different choices of match panel.

Results

Achromatic Adjustments

For two observers, Figure 2 shows the mean chromaticities of the achromatic settings under the bluish and gradient illuminants (circles and triangles, respectively). Under the bluish illuminant each point is the mean of 48

measurements, computed across replications and luminance levels. Under the gradient illuminant each point is the mean of 36 measurements. The achromatic chromaticities were independent of luminance. The standard error of the mean for each point was between 0.001 and 0.002 when expressed in CIE xy chromaticity.

The achromatic settings clearly vary with the illuminant. The effect is roughly the same for the two observers. For a color constant observer, we would expect the achromatic chromaticities to lie near the illuminant chromaticities. For our observers, this is approximately true. We interpret the data in terms of color constancy in the discussion.

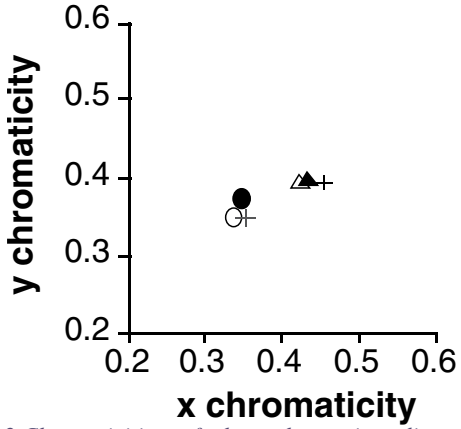

Figure 2.Chromaticities of the achromatic adjustments for observers JMS and ASH under the blue and gradient illuminant conditions.Circles represent the adjustments for the bluish illuminant.Triangles represent the adjustments for the gradient illuminant.The open symbols are for ASH; the closed, for observer JMS. Crosses represent the illuminant chromaticities at the match location for the bluish and gradient conditions.

Asymmetric Matching

The asymmetric matching data consist of the measured spectra of the test panel and the observer's appearance match. We present a subset of the matches for observer JMS in Figure 3. The figure illustrates the shift in chromaticity and luminance of the measured match/test pairs. The open squares correspond to the test; the closed, to the match. Each closed symbol represents the mean of 2-3 matches. Because of the illumination gradient, the data are not colorimetric matches.

We examined the CIELAB ΔE* deviations between each match and the mean match to that test panel. In calculating the error, we used the measured illuminant at the match location as our white point. The mean ΔE* deviations across the entire set of matches were 2.42 for ASH and 1.89 for JMS.

A good description of the matches was obtained using a simple von Kries transformation. This transformation accounts for the effect of the illuminant by independently scaling the L, M, and S cone responses. We used numerical search to fit such a transformation to our data. The fit minimized the CIELAB ΔE* error between the predicted and measured matches. The model accounts for most of the variance in the data. The mean prediction errors for the model were 4.43 ΔE* units for observer ASH and 3.61 ΔE* units for observer JMS.

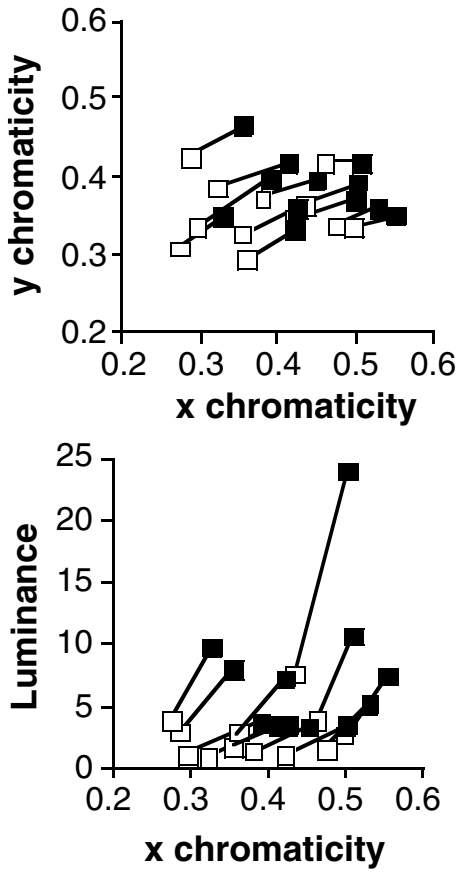

Figure 3. Asymmetric matches for observer JMS. Open squares represent test coordinates; closed squares, the observer's matches.

Color Naming

A qualitative sense for the naming data may be obtained from Figure 4. We present the stimulus chromaticities for four landmark colors (red, green, blue and yellow) under the bluish and gradient illuminants. Each symbol represents a stimulus which received a rating of seven or greater. The change in illumination affects the locus of points for each color term. The largest difference is in the locus of yellow. Blue expands toward higher y-chromaticities while both red and green become more compact. Exemplars for other color names undergo similar shifts.

One way to summarize color naming data is with the centroid of the stimuli that elicit a given color name. Such centroids have been used to demonstrate the positions of the color terms for a fixed stimulus set under a single illuminant[11,13,16,17] and to analyze the effect of context on color naming.[6,12,15,18] We considered this approach for our data. After preliminary analyses, however, we concluded that the centroid location is heavily influenced by the stimulus gamut. This is revealed in the data by the fact that the best exemplars of most categories are found at the edges of our gamut. Because there is no principled way to equate stimulus gamuts across changes of illumination, we do not in general use the centroid locations to measure the effect of changing the illuminant.

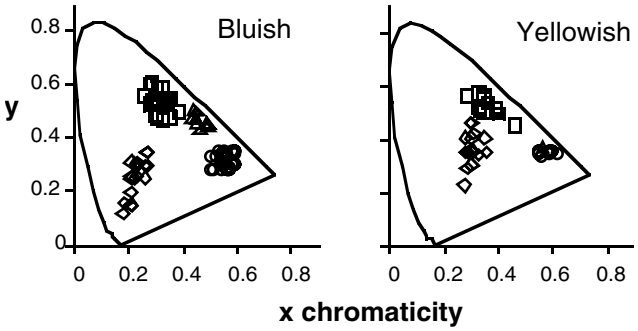

Figure 4. Naming summary for two illuminants, for observer JMS. The symbols represent the stimuli rated red (circles), green (squares), blue (diamonds), and yellow (triangles) which received ratings of seven or greater.

To get a quantitative handle on our naming data, we treated the naming data for each stimulus as an eleven-dimensional *naming vector*. Each entry of the vector gives the observer's rating for one of the eleven basic color categories. Typically from one to four entries of the naming vector were non-zero. To summarize our raw data, we averaged the naming vectors for stimuli contained in rectangular regions of our stimulus space. Each region was 0.01 in x chromaticity by 0.01 in y chromaticity by 2 cd/m^2 in luminance. This averaging procedure allows us to associate naming vectors with each small region of stimulus space.

We use a simple city-block metric for evaluating the difference between naming vectors. For two naming vectors **a** and **b** we take

$$e_{overall} = \Sigma \left| a_i - b_i \right|$$

as the difference between them, where

$$e_i = a_i - b_i$$

is the difference in ratings for a single color name. We can use our metric to examine how consistently observers named colors. For each named stimulus, we calculate how close its naming vector is to the mean naming vector in the corresponding region of the stimulus space. Histograms for these differences are shown in Figure 5. Observer ASH was more consistent than JMS.

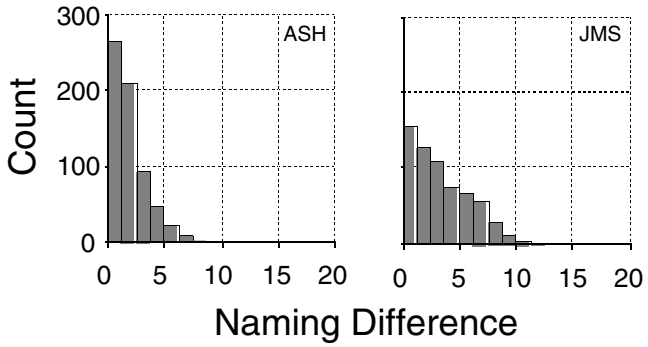

Figure 5. Naming consistency collapsed across color term and illuminant, computed using our naming difference measure. Each panel shows a histogram of naming differences for one observer.

The histograms of Figure 5 are useful because we can compare them to histograms that result from other ways of predicting naming vectors. For example, we can compute the consensus between observers by asking how well the mean naming data for one observer predicts the individual names given by another observer. Figure 6 shows how well ASH's mean naming data predicts JMS's individual names. Although the histogram is broader than those of Figure 5, the difference is not overwhelming, particularly in view of the fact that it is affected by variability in two observers' data rather than just one.

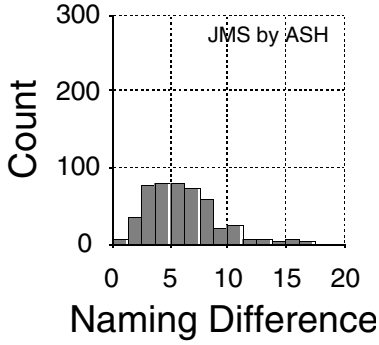

Figure 6. Naming consensus collapsed across color term and illuminant, computed using our naming difference measure. We used ASH's mean naming data for each stimulus region to predict the names given by JMS to individual stimuli in that region.

Analysis and Discussion

Agreement between Tasks

If color appearance assessed by the different tasks is based on a common appearance representation, the effect of the illuminant should be the same for each. We evaluate the agreement between methods in two ways. First, we consider the shift in stimuli indicated as achromatic by all three tasks. Second, we ask whether we can use the matching data to predict how color names change with the illuminant.

Figure 7 compares the achromatic chromaticities. For the achromatic adjustments, these were computed as described above. For the asymmetric matches, we extracted and averaged the data corresponding to neutral test surfaces (Munsell N 2.75/, N3/, N3.5/ and N5/). These surfaces appeared approximately achromatic. For the naming data, we computed the centroid of all stimuli named black, white, or grey. The centroid is a valid summary for these data because the relevant stimuli lie away from the edges of our stimulus gamuts. The differences between the three tasks are small, especially when compared to the size of the shift induced by changing the illuminant. In addition, the small differences between tasks are not systematic across observers.

To compare the matching and naming data, we proceeded as follows. We began with the von Kries transformation fit to the matching data. This transformation allows us to map the mean naming data under the blue illuminant into a target region of the naming stimulus space under the gradient illuminant. We then predicted the mean naming data under the blue illuminant with the mean naming vector for its target region. Figure 8 shows a histogram of the

naming errors computed in this way. This histogram may be compared with those in Figures 5 and 6. We see that our ability to predict the change in color names using the matching data is not as good as our individual observers' consistency (Figure 5) but is comparable to the consensus between observers (Figure 6). Given that the accuracy of the match based prediction is subject to error in two tasks, we regard this agreement as very good.

Color Constancy

Our conclusion is that all three tasks reveal similar and perhaps identical effects of the illuminant on color appearance. This agreement supports the use of context-task separability in the formulation of color appearance models. We caution, however, that we have not examined all possible tasks. In particular, it remains possible that tasks that involve reasoning about relations among colors may reveal high-level effects of context not tapped by our experiments.[4]

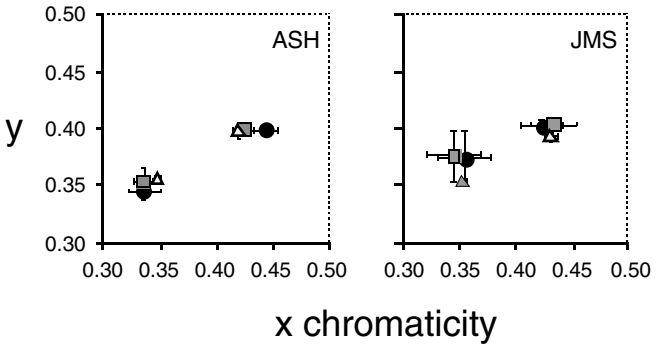

Figure 7. Comparison of achromatic stimuli under the bluish and gradient illuminant conditions. Squares represent the mean achromatic adjustments; triangles, the mean match to Munsell neutrals; circles, the mean chromaticity of stimuli named achromatic. Error bars represent +/- 1 standard deviation.

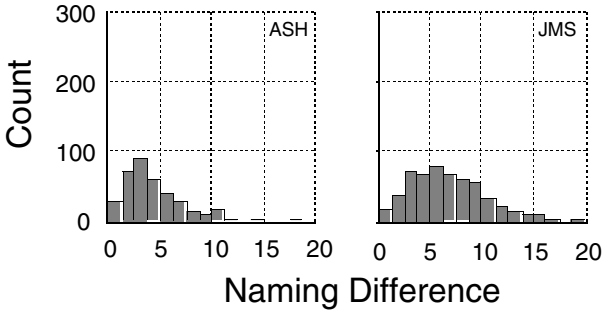

Figure 8. Naming error for match-based predictions using the naming difference measure.

The agreement across tasks in our data means that we can use the data from one task to assess how color constant our observers are for all three. For this purpose, we will analyze the asymmetric matching data. The matching data assess the effect of the illuminant for many test colors and are more precise than the naming data. We have already described the model and fitting procedure which we applied to the matching data. From the von Kries model, we can calculate what stimulus would have been matched to a perfectly reflecting surface. We then compared the shift revealed by this calculation to the shift that would have been seen given perfect constancy. In this sense, observer ASH showed 77% constancy while observer JMS showed 67% constancy. These degrees of constancy are higher than typically seen when the experiments are conducted using simulations presented on CRT displays.[4,21]

References

1. R. W. G. Hunt, "Revised colour-appearance model for related and unrelated colours," *Color Research And Application* **16**, 146-165 (1991).
2. Y. Nayatani, "Revision of the chroma and hue scales of a nonlinear color-appearance model," *Color Research and Application* **20**, 143-155 (1995).
3. M. D. Fairchild and R. S. Berns, "Image color-appearance specification through extension Of CIELAB," *Color Research And Application* **18**, 178-190 (1993).
4. L. E. Arend and A. Reeves, "Simultaneous color constancy," *Journal of the Optical Society of America A* **3**, 1743-1751 (1986).
5. A. L. Gilchrist, S. Delman and A. Jacobsen, "The classification and integration of edges as critical to the perception of reflectance and illumination," *Perception and Psychophysics* **33**, 425-436 (1983).
6. J. M. Troost and C. M. de Weert, "Naming versus matching in color constancy," *Perception & Psychophysics* **50**, 591-602 (1991).
7. H. Helson and W. C. Michels, "The effect of chromatic adaptation on achromaticity," *Journal of the Optical Society of America* **38**, 1025-1032 (1948).
8. J. S. Werner and J. Walraven, "Effect of chromatic adaptation on the achromatic locus: the role of contrast, luminance and background color," *Vision Research* **22**, 929-944 (1982).
9. E. J. Chichilnisky and B. A. Wandell, "Seeing gray through the on and off pathways," *Visual Neuroscience* **13**, 591-596 (1996).
10. B. Berlin and P. Kay *Basic color terms: Their universality and evolution*, (University of CA Press: 1969).
11. R. M. Boynton and C. X. Olson, "Locating basic colors in the OSA space," *Color Research and Application* **12**, 94-105 (1987).
12. R. M. Boynton and K. F. Purl, "Categorical colour perception under low-pressure sodium lighting with small amounts of added incandescent illumination," *Lighting Research and Technology* **21**, 23-27 (1989).
13. R. M. Boynton, R. E. MacLaury and K. Uchikawa, "Centroids of color categories compared by two methods," *Color Research and Application* **14**, 6-15 (1989).
14. R. M. Boynton and C. X. Olson, "Salience of chromatic basic color terms confirmed by three measures," *Vision Research* **30**, 1311-1317 (1990).
15. R. M. Boynton, L. Fargo and B. L. Collins, "Categorical color rendering of four common light sources," *Color Research and Application* **15**, 222-230 (1990).
16. P. K. Kaiser and R. M. Boynton, "Naming, ordering, and recognizing surface colors." In *Human Color Vision*. (Optical Society of America: Washington, D.C., 1996).
17. K. Uchikawa and R. M. Boynton, "Categorical color perception of Japanese observers: comparison with that of Americans," *Vision Research* **27**, 1825-1833 (1987).
18. K. Uchikawa, H. Uchikawa and R. M. Boynton, "Partial color constancy of isolated surface colors examined by a

color-naming method," *Perception* **18**, 83-91 (1989).
19. L. Sivik and C. Taft, "Color naming: A mapping in the NCS of common color terms," *Scandinavian Journal of Psychology* **35**, 144-164 (1994).
20. G. Borg and E. Borg, "Principles and experiments in category-ratio scaling," *Reports from the Department of Psychology, Stockholm University* (1994).
21. D. H. Brainard and B. A. Wandell, "A bilinear model of the illuminant's effect on color appearance." In *Computational Models of Visual Processing*; J. A. Movshon and M. S. Landy eds., (MIT Press: Cambridge, MA, 1991).