Perception, 2002, volume 31, pages 247 – 263

DOI:10.1068/p08sp

Surface-illuminant ambiguity and color constancy: Effects of scene complexity and depth cues

James M Kraft¶, Shannon I Maloney, David H Brainard
Department of Psychology, University of California at Santa Barbara, Santa Barbara, CA 93106, USA
Commissioned for the Special issue on Articulation

Abstract. Two experiments were conducted to study how scene complexity and cues to depth affect human color constancy. Specifically, two levels of scene complexity were compared. The low-complexity scene contained two walls with the same surface reflectance and a test patch which provided no information about the illuminant. In addition to the surfaces visible in the low-complexity scene, the high-complexity scene contained two rectangular solid objects and 24 paper samples with diverse surface reflectances. Observers viewed illuminated objects in an experimental chamber and adjusted the test patch until it appeared achromatic. Achromatic settings made under two different illuminants were used to compute an index that quantified the degree of constancy. Two experiments were conducted: one in which observers viewed the stimuli directly, and one in which they viewed the scenes through an optical system that reduced cues to depth. In each experiment, constancy was assessed for two conditions. In the valid-cue condition, many cues provided valid information about the illuminant change. In the invalid-cue condition, some image cues provided invalid information. Four broad conclusions are drawn from the data: (a) constancy is generally better in the valid-cue condition than in the invalid-cue condition; (b) for the stimulus configuration used, increasing image complexity has little effect in the valid-cue condition but leads to increased constancy in the invalid-cue condition; (c) for the stimulus configuration used, reducing cues to depth has little effect for either constancy condition; and (d) there is moderate individual variation in the degree of constancy exhibited, particularly in the degree to which the complexity manipulation affects performance.

1 Introduction

Useful vision depends on perceptual representations that make explicit the properties of the environment (ie the *distal stimulus*) rather than the properties of the retinal image (ie the *proximal stimulus*). In the case of color, the distal stimulus may be taken as the surface reflectances of the objects in the scene along with the spectral power distribution of the illumination incident on the objects.[1] The process of reflection confounds surface and illuminant information in the retinal image so that variation in one may masquerade as variation in the other. Classically, color constancy refers to the ability of the visual system to process the retinal image and produce a perceptual representation of surfaces that is stable against variation in the illumination. Such an ability is critical if color appearance is to be a useful perceptual indicator of object properties.

Both introspection and empirical studies indicate that under some conditions human vision exhibits excellent color constancy. Typical color constancy experiments are designed to determine how varying the illuminant affects object color appearance (eg Burnham et al 1957; McCann et al 1976; Arend and Reeves 1986; Brainard and Wandell 1992; Lucassen and Walraven 1993; Bauml 1994). Indeed, we have conducted experiments of this sort in our laboratory and shown that, if a test object is viewed amongst a fixed set of contextual objects, the color appearance of the test object changes only modestly as the illuminant is varied (eg Brainard et al 1997a; Brainard 1998).

¶ Present address: Vision Sciences Laboratory, Department of Optometry and Neuroscience, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK; e-mail: j.kraft@umist.ac.uk

[1] In this description, we neglect consideration of geometrical factors that can influence the spectrum of the light reflected to the eye.

Although experiments where the independent variable is the illuminant are useful for demonstrating that color constancy exists, they cannot provide a complete description of how well the visual system separates illuminant variation from surface variation. As a number of theorists have pointed out, color constancy is a difficult computational problem because the visual system must separate these two components of the distal stimulus using only the confounded representation provided by the proximal stimulus (eg Buchsbaum 1980; Brainard and Wandell 1986; Maloney and Wandell 1986; D'Zmura and Iverson 1993; D'Zmura et al 1995; Brainard and Freeman 1997; see also Hurlbert 1998; Maloney 1999). Closely related is the idea that, to understand how the visual system separates illuminant from surface variation, both factors must be manipulated in the experimental design (Gilchrist and Jacobsen 1984; McCann 1994; Kraft and Brainard 1999; Brainard et al 2001). For example, we (Kraft and Brainard 1999; see also Hurlbert 1999; Brainard et al 2001) showed that the effect of an illuminant change on object color appearance depends strongly on whether the other objects in the scene remain constant (as in typical studies) or whether they are varied so as to silence some potential cues to the illumination change. Because we varied both the illuminant and the surface reflectances of objects, we were able to perform strong tests of a number of models that had been proposed to explain color constancy.

In this paper, we report additional experiments where we assess color appearance under changes of both illuminant and contextual surfaces. In the first experiment we assessed the effect of increasing scene complexity on color constancy. A number of authors have suggested that more complex, or well-articulated, scenes lead to greater constancy (Gelb 1929/1938; Burzlaff 1931; MacLeod 1932; Henneman 1935; Gilchrist et al 1999; see also Maloney and Wandell 1986; Brainard and Freeman 1997). Our complexity manipulation compared two specific scenes. The low-complexity scene contained two walls with the same surface reflectance and a test patch which itself provided no information about the experimental illuminant. In addition to the surfaces visible in the low-complexity scene, the high-complexity scene contained a rectangular solid object and 24 paper samples with diverse surface reflectances. Neither of these scenes contained objects having strong specular reflections, smooth gradients of surface reflectance, or a wide variety of differently oriented surfaces—other factors thought to increase scene complexity. In the second experiment we examined whether manipulating the information available about the three-dimensional structure of the image affects the ability of the visual system to separate illuminant and surface variation. In both experiments, we included conditions where we expected, on the basis of our previous work, constancy to be good (contextual surfaces held constant) and where we expected it to be poor (contextual surfaces manipulated to reduce cues to the illuminant change).

2 Experiment 1: Effect of scene complexity

2.1 *Methods*

2.1.1 *Overview.* The purpose of this experiment was to determine whether color constancy is better for scenes that are more complex, in the sense that they contain more objects and a wider range of chromaticities. We measured color constancy for two stimulus configurations. In the *low-complexity configuration*, observers looked into a rectangular chamber lined with cardboard of a single reflectance (see figure 1a). In the *high-complexity configuration*, additional objects were placed in the chamber to increase its complexity (see figure 1b).

For each level of complexity (low/high), we made two measurements of color constancy. In the *valid-cue condition*, the surfaces in the chamber were held constant across the illuminant change. When the scene objects remain constant, there are many valid cues to the illuminant change. In the *invalid-cue condition*, the objects in the scene

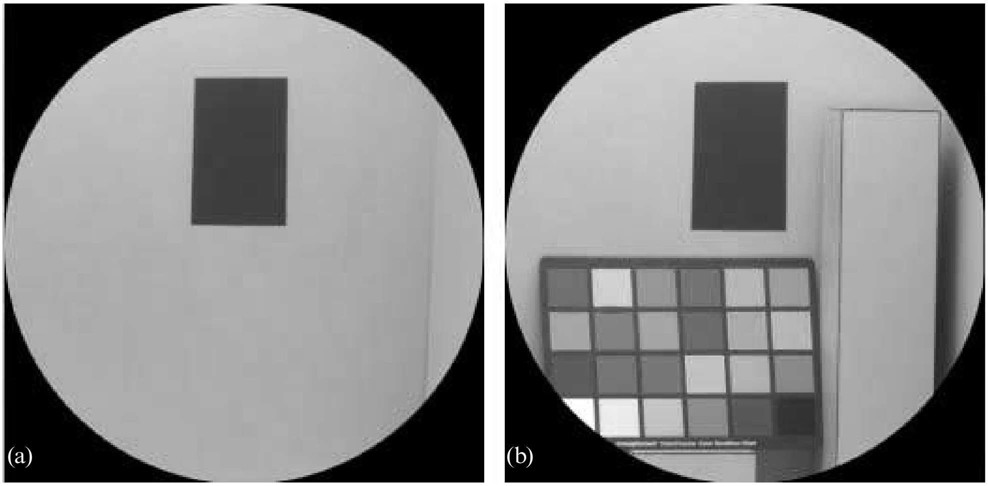

Figure 1. Observer's view of the experimental chamber. (a) Low-complexity configuration. The walls of the chamber were lined with cardboard of a single reflectance. The test patch is visible on the back wall of the chamber. (b) High-complexity condition. Additional objects were placed in the chamber. A color version of this figure may be viewed on the world wide web at http://color.psych.ucsb.edu/complexity/figure1.html and also at http://www.perceptionweb.com/perc0202/kraft.html.

were manipulated in conjunction with the illuminant change to reduce the validity of cues to the change in illuminant.

To measure constancy, we asked observers to adjust a test patch so that it appeared achromatic (somewhere on the perceptual continuum from black to gray to white). We measured how the achromatic settings varied with the illuminant and composition of surfaces in the scene. The physical difference between observers' settings under two illuminants indicates how observers adapted to the illuminant change. The general method of achromatic adjustment has been used extensively in the study of color appearance (eg Helson and Michels 1948; Werner and Walraven 1982; Bauml 1994; Chichilnisky and Wandell 1996; Brainard 1998; Kraft and Brainard 1999). Speigle and Brainard (1999) showed that conclusions about color constancy drawn from achromatic adjustments match those drawn from more extensive asymmetric matching experiments.

2.1.2 *Observers*. Ten observers (four male and six female; mean age 23.4 years) were tested. All observers were color normal as classified by the American Optical Company H-R-R (Hardy, Rand, and Rittler) pseudoisochromatic plates and the Ishihara color plates. None of the observers wore glasses or contacts during the experiment. Eight of the ten observers did not routinely wear glasses or contact lenses and reported having good uncorrected vision.[2] Nine of the ten observers were completely naïve about the design and purpose of the experiment.[3] The naïve observers were paid for their participation. Data from an eleventh observer were excluded because of a calibration error that occurred during his sessions.

2.1.3 *Stimuli*. An experimental chamber (102 cm high × 70 cm wide × 73 cm deep) contained the viewed scene. The chamber was illuminated by three theater stage lamps. Each lamp had a different filter (red, green, or blue) and the output of the three lamps was passed through a diffuser at the top of the chamber. The chromaticity and luminance of the diffuse scene illumination were controlled by varying the intensity of

(2) Observer LNR used glasses during lectures but not while driving. Observer MAE was not asked if she had good uncorrected vision, but she did not routinely wear glasses or contacts.

(3) Author DHB was a non-naïve observer in these experiments.

the three lamps. This type of illuminant hardware and its associated control software are described in detail elsewhere (Brainard et al 1997a; Brainard 1998).

All visible surfaces of the chamber were lined with matte cardboard, which could be changed between sessions. By varying the cardboard and other objects placed in the chamber, we could vary the reflectance of the surfaces in the scene. In the low-complexity configurations, the chamber was lined with a single type of cardboard and no additional objects were placed in it (see figure 1a). In the high-complexity configurations, a Macbeth Color Checker Chart and two rectangular solid objects, made from the same cardboard as that on the walls, were placed in the chamber (see figure 1b). Adding the objects increased the spatial complexity of the scene and also increased the number of distinct chromaticities visible to the observer.

Three illuminant spectral power distributions and two cardboard surface reflectances were used in the experiments. We take the descriptive liberty of using color names to refer to both the illuminants and the cardboards. We refer to the illuminants as 'aqua', 'orange-low', and 'orange-high'. We refer to the cardboards as 'gray' and 'blue'. Physical specifications of both illuminants and cardboards are provided in table 1. Information about which illuminants were paired with which cardboards is provided below (see section 2.1.7) and in table 2.

Table 1. Specification of illuminants and cardboard surface reflectances. The table provides specification of the illuminants and cardboards used in the experiments. For illuminants, CIE xy chromaticities and luminance in cd m^{-2} are specified. For cardboards, CIE object-color xy chromaticities and luminance factors are specified, as computed with respect to an equal-energy illuminant. For backgrounds, the chromaticities and luminance in cd m^{-2} of the light reflected from the area immediately adjacent to the test patch are specified. See section 3.1 for a discussion of how stimulus luminance was specified for TVS viewing.

	Experiment 1	Experiment 2
Illuminant		
'Aqua'	(0.303, 0.363, 19.3)	(0.288, 0.363, 19.0)
'Orange-low'	(0.437, 0.409, 19.8)	(0.419, 0.408, 18.5)
'Orange-high'	(0.421, 0.415, 65.8)	(0.419, 0.412, 60.5)
Cardboard		
'Gray'	(0.345, 0.341, 34.5)	
'Blue'	(0.246, 0.253, 11.5)	
Background		
'Aqua – gray'	(0.318, 0.370, 6.3)	(0.303, 0.369, 6.4)
'Orange – gray'	(0.451, 0.409, 6.8)	(0.435, 0.407, 6.4)
'Orange – blue'	(0.315, 0.378, 6.9)	(0.310, 0.371, 6.3)

A test patch was located on the back wall of the chamber. The light reaching the eye from the test patch was manipulated by varying the illumination cast on it by a projection colorimeter (see Brainard et al 1997a; Brainard 1998; Kraft and Brainard 1999). The illumination from the colorimeter formed a rectangular spot focused on the test patch.[(4)] It was not explicitly visible to the observer, and the test patch appeared as a surface illuminated only by the diffuse scene illuminant. By varying the illumination from the colorimeter, we could vary the apparent surface color of the test patch. Different test patch materials were used with the two cardboards to ensure that the

[(4)] To reduce apparatus alignment demands, the projected spot was slightly smaller than the test patch, leaving a thin (approx. 1.6 mm, 0.1 deg) border. We do not believe that the presence of this border substantially affects achromatic judgments for our stimulus configuration (see control experiments reported in Brainard 1998).

Table 2. Summary of stimulus conditions. The table summarizes the properties of the two scenes used in each of the four comparisons made. Note that scene 1 is the same for both valid-cue and invalid-cue comparisons. When objects were added, they consisted of two rectangular solids made of the same cardboard as the walls and a Macbeth Color Checker Chart.

Condition	Low complexity		High complexity	
	scene 1	scene 2	scene 1	scene 2
Valid cue				
Illuminant	'Aqua'	'Orange-low'	'Aqua'	'Orange-low'
Wall cardboard	'Gray'	'Gray'	'Gray'	'Gray'
Objects added?	no	no	yes	yes
Invalid cue				
Illuminant	'Aqua'	'Orange-high'	'Aqua'	'Orange-high'
Wall cardboard	'Gray'	'Blue'	'Gray'	'Blue'
Objects added?	no	no	yes	yes

apparatus gamut included the achromatic locus for each experimental combination. When the 'gray' cardboard was placed in the chamber, the test patch (10 cm wide × 15 cm high; 6.3 deg × 10.1 deg) was made of matte dark-gray Munsell paper (N3/). When the 'blue' cardboard was placed in the chamber, the test patch was made of matte dark-blue-green Munsell paper (5B3/2). The test patch was changed merely to optimize the effective range of the projection colorimeter for the probable locus of the achromatic settings, determined in pilot studies on one of the authors. Each test patch was calibrated separately (see below) to compensate for the different underlying surface reflectances.

2.1.4 *Procedure*. Observers viewed the interior of the chamber monocularly (right eye) through a 17 cm circular aperture in the front of the chamber. An exterior hood shielded the observers from any stray light in the experimental room. The observers' field of view into the chamber was 45 deg in diameter. This allowed them to see the test patch and portions of the back wall and right wall of the chamber (see figure 1). In each session, observers made 8 achromatic settings at each of 2 luminances (3.8 and 4.9 cd m^{-2})[5] by using a game controller to manipulate the CIELAB a^* and b^* chromaticity coordinates of the stimulus. Control software held the luminance of the test patch approximately constant during the adjustment. A short break separated two blocks within a session. Each observer made settings in one or more sessions per condition. Observers were encouraged to look around the apparatus before finalizing each setting (see instructions below).

Each session consisted of two blocks of 8 trials each. The starting chromaticity of the test for the first adjustment in each block was chosen at random from a region of color space centered on the chromaticity of the local surround of the test patch. On subsequent adjustments in a block, the starting chromaticity was chosen as a function of the preceding settings.

[5] The luminances available for use in this experiment were limited by our desire to match the viewing conditions of experiment 2. In experiment 2, light loss in the optical viewing system reduced the luminance gamut of the apparatus. The two test-patch luminances used lie near the top of the mesopic range. The background had slightly higher luminance of approximately 6.5 cd m^{-2} and occupied the majority of the central visual field. Thus it probably determined the overall adaptive state of the retina, particularly given our instructions that asked observers to look at different parts of the scene frequently. While not guaranteeing the exclusion of rod effects, these luminances would certainly minimize them. In a recent CRT-based asymmetric matching experiment, Delahunt and Brainard (2000) found little if any difference in matching performance when test-patch luminances were increased by about 2 orders of magnitude from the range 4 – 32 cd m^{-2}.

We took precautions to avoid providing observers with any information about the illuminant other than that which could be gleaned through viewing the experimental chamber. First, observers were seated under a black hood, so that essentially no light from the experimental chamber scattered to their eyes from the walls or floor of the experimental room. Second, observers entered and left the experimental room under incandescent room illumination and with the experimental illuminants turned off. When the observers were first seated at the apparatus, a shutter prevented them from seeing the contents of the experimental chamber. After observers were seated under the hood, the room lights were turned off and the experimental illuminant was set. Only at this point was the shutter opened so that the observer could view the stimulus. Each observer was asked to adapt to the stimulus before beginning the first adjustment, but no fixed adaptation time was enforced.

2.1.5 *Instructions.* The instructions provided to observers can affect their judgments of color appearance; under some circumstances, observers can distinguish the appearance of a surface (apparent surface color) from the appearance of the light reflected from the surface (unasserted color; Arend and Reeves 1986; Arend 1994). This phenomenon is often taken to indicate observers' ability to distinguish the identity of surfaces from the appearance of the light reflected from those surfaces, but the range of conditions where observers might be able to make this distinction is not yet well-understood (but for related ideas see eg Metelli 1974; Gerbino et al 1990; D'Zmura et al 1997). We did not explore instructional effects in the experiments reported here. We did, however, read a set of instructions to each observer before the experiment began. The instructions explained how to change the color of the test patch and told the observer to try to make the test patch appear gray. The three passages quoted below were intended to standardize behavior across observers.

> "... After the patch looks perfectly gray to you, look at the different parts of the scene to give your eyes a little rest. Then look at the patch again to make sure it still looks perfectly gray. If it doesn't, adjust it until it does. Keep looking around and then adjusting the patch until it looks perfectly gray. You've finished making a setting when the patch looks gray AFTER you have looked around ..."

> "We want your judgment to be made about the color you see, not what color of paper the rectangular patch looks like it's made out of. For example, if the patch looks yellow because there seems to be yellow light falling on gray paper, adjust the patch until the yellow sensation is gone."

> "While you're doing the experiment, it might be tempting to just assume that one of the surfaces you see is gray and then adjust the test patch until it looks like that surface. It's very important that you do not do that. We want you to adjust the patch so that it looks gray, not so that it looks like some other surface that you see ..."

Note that observers were instructed to judge unasserted color rather than apparent surface color.

2.1.6 *Calibration.* Calibration measurements were made to characterize the chamber illuminants and projection colorimeter (see Brainard et al 1997a; Brainard 1998). Illumination measurements were made separately for each chamber cardboard. Projection colorimeter measurements were made separately for each test patch. The calibration data were used to choose device settings for the experiments. To compensate for apparatus drift and quantization error, the experimental stimuli and observer settings were measured after each experimental session. The data we report are based on these direct measurements.

2.1.7 *Conditions.* For both low-complexity and high-complexity configurations of the chamber, achromatic settings were measured for three combinations of illuminant and cardboard: 'aqua' illuminant paired with 'gray' cardboard (*aqua – gray combination*), 'orange-low' illuminant paired with 'gray' cardboard (*orange – gray combination*), and 'orange-high' illuminant paired with 'blue' cardboard (*orange – blue combination*). The light reflected from the cardboard in the aqua – gray and orange – blue combinations had almost the same chromaticity and luminance. The light reflected from the cardboard in the orange – gray combination had approximately the same luminance as that in the other two combinations, but a different chromaticity (see table 1).

Assessment of performance for the valid-cue condition is obtained by comparing achromatic settings from the orange – gray combination with those from the aqua – gray combination. Assessment of performance for the invalid-cue condition is obtained by comparing achromatic settings from the orange – blue combination with those from the aqua – gray combination. Table 2 summarizes properties of the scenes used in each comparison.

2.2 *Results*

2.2.1 *Valid-cue condition.* Figure 2 shows the data for observer LSI in the low-complexity, valid-cue condition. The filled circles indicate the CIE xy chromaticities for the 'aqua' and 'orange-low' illuminants used in the two combinations. The solid line connects two points which give LSI's achromatic loci from the aqua – gray and orange – gray combinations. Each locus was computed from the individual settings as described in detail by Brainard (1998). We have previously established that, for conditions similar to ours, the achromatic chromaticity does not vary systematically with luminance (Brainard 1998). The chromaticity of the achromatic locus under the 'aqua' illuminant lies near the chromaticity of the 'aqua' illuminant (lower left), and the chromaticity of the achromatic locus under the 'orange-low' illuminant lies near the chromaticity of the 'orange-low' illuminant (upper right).

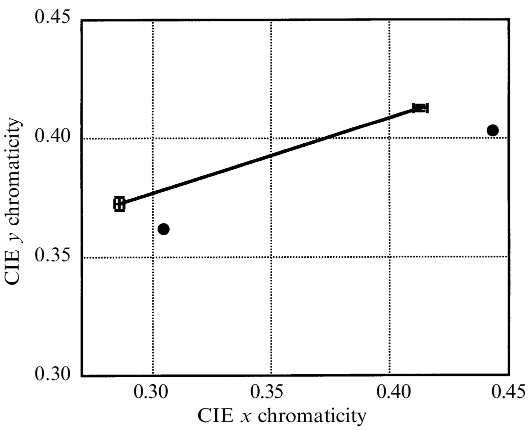

Figure 2. Experiment 1. Single-observer data for the low-complexity configuration, valid-cue condition. The filled circles show the chromaticities of the 'aqua' and 'orange-low' illuminants. Each end of the solid line indicates the chromaticity of a measured achromatic locus for observer LSI. The lower left end of the line indicates the locus for the aqua – gray combination. The upper right end indicates the locus for the orange – gray combination. Each locus was obtained by averaging the chromaticities of individual settings at two luminances. Error bars show standard errors.

In figure 2, the span between the achromatic chromaticities (solid line) and the span between the illuminant chromaticities (filled circles) have similar lengths and directions, indicating that the chromaticity difference between the proximal stimuli that looked achromatic to observer LSI was similar to the difference between the illuminants. This pattern of results indicates good constancy (see Brainard 1998; Kraft and Brainard 1999; Brainard et al 2001). To understand why, consider that, when the illuminant incident on a surface changes, the light reflected from the surface changes similarly. For the surface to retain the same appearance, the visual system has to adjust so that the light reflected under the second illuminant looks the same as the light reflected under the first illuminant. The shift in achromatic loci shown in figure 2 reveals just this type of adjustment.

To quantify the degree of color constancy, we use the chromaticities of the achromatic loci and illuminants to compute a constancy index. Intuition regarding the index may be obtained by considering the ratio of the shift in the achromatic chromaticity to the shift in the illuminant chromaticity. This ratio takes on a value of zero when there is no adjustment to the illuminant change and a value near one when the adjustment is consistent with perfect color constancy. Our constancy index is based on a similar calculation.(6) Although the index provides a useful summary of performance, it is important to keep in mind that it condenses eight numbers (chromaticity coordinates of two achromatic loci and two illuminants) to one number and thus can potentially obscure interesting patterns in the data. Here, however, conclusions drawn from consideration of the constancy indices are consistent with conclusions drawn from the complete data set.

The constancy index for the data shown in figure 2 is 0.88.

How does increasing scene complexity affect constancy? Figure 3a shows data for observer LSI for the high-complexity, valid-cue condition. The unfilled circles show the illuminant chromaticities, while the endpoints of the dashed line show the achromatic loci. For comparison, the data from the low-complexity condition are reproduced from figure 2. Recall that the difference between the high-complexity and low-complexity conditions was the addition of objects to the experimental chamber. Figure 3a shows

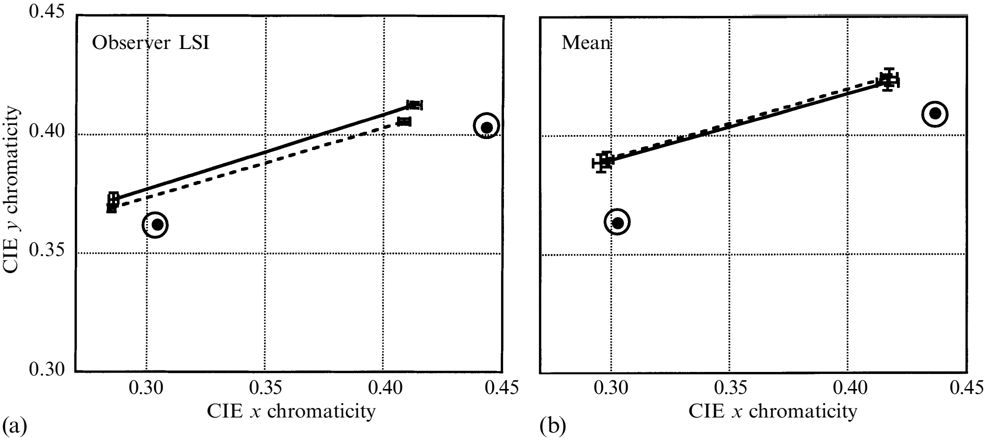

Figure 3. Experiment 1. (a) Single-observer data and (b) mean data for both complexity configurations, valid-cue condition. The unfilled circles show the chromaticities of the 'aqua' and 'orange-low' illuminants. Each end of the dashed line indicates the chromaticity of the measured achromatic locus from the high-complexity configuration. In (a), data from figure 2 are replotted for comparison in the same format as they were shown in figure 2. In (b), mean data are plotted in the same format. In the figure, the unfilled circles overlay closely with the filled circles. Recall that we measured the illuminant after each experimental session. For this reason, it is possible for the illuminants in two nominally matched combinations to differ slightly. Such differences are small here but are visible in other figures of this paper. Error bars show in (a) the standard errors of the individual settings, and in (b) the between-observer standard errors.

(6) Although the idea of taking the ratio of chromaticity shifts captures the intuition underlying our constancy index, we do not compute it in this way. The actual computation is designed to produce more reasonable results than would a simple ratio in the case where the shift of achromatic chromaticity is not parallel to the shift in illuminant chromaticity. The details of the index computation are given in Brainard (1998). The procedure described there requires defining one scene as the standard and the other as the comparison. Since there is no a priori reason to assign the role of standard to either experimental scene, we follow Kraft and Brainard (1999) and compute the index using both assignments, then report the average.

that adding these objects had little effect on constancy for observer LSI. The achromatic loci from the high-complexity configuration (dashed line) are shifted slightly relative to those from the low-complexity configuration (solid line), but the spans between the corresponding settings have similar length and direction. Consistent with the similarity in the settings, LSI's constancy indices are also similar (0.85 high-complexity, 0.88 low-complexity).

Figure 3b shows mean data for the ten observers plotted in the same format as figure 3a. As with observer LSI, the data indicate that there is little difference between the two complexity configurations. Figure 4 shows the individual constancy indices for all ten observers. The mean constancy indices were 0.86 in the high-complexity configuration and 0.87 in the low-complexity configuration. These are not significantly different (paired two-tailed t-test, $t_9 = -0.45$). Adding spatial and chromatic complexity to our simple scene does not improve the ability of the visual system to perceive the colors of objects as invariant across an illuminant change.

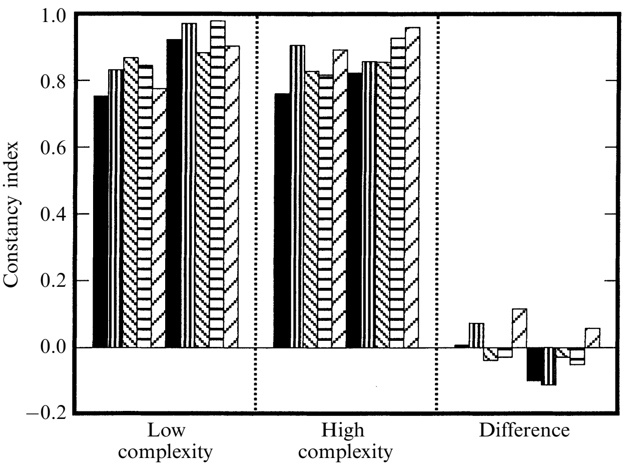

Figure 4. Experiment 1. Individual constancy indices for the valid-cue condition. The figure shows the constancy indices obtained for each observer in the low-complexity configuration (left) and high-complexity configuration (center). The difference in indices obtained in the two configurations is shown on the right.

2.2.2 *Invalid-cue condition.* Figure 5a shows data for observer LSI for the invalid-cue condition (both low-complexity and high-complexity configurations). The data are in the same format as figure 3a. A number of differences from the results from the valid-cue conditions are apparent. First, the degree of constancy is much lower than for the

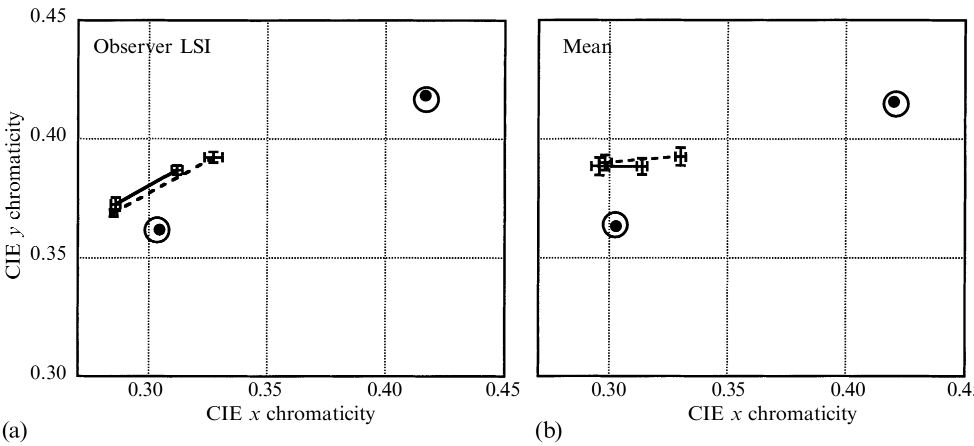

Figure 5. Experiment 1. (a) Single-observer data and (b) mean data for both complexity configurations, invalid-cue condition. Same format as figure 3.

valid-cue conditions. For both low-complexity and high-complexity, the chromaticities of the achromatic loci have shifted away from the chromaticity of the 'orange-low' illuminant and towards the chromaticity of the 'aqua' illuminant. This shift in the loci is accompanied by a corresponding decrease in the constancy indices. These are 0.22 for the low-complexity configuration and 0.36 for the high-complexity configuration. Second, unlike in figure 3a, the data for the low-complexity and high-complexity configurations differ substantially. The achromatic locus measured for the high-complexity configuration reveals better constancy than that for the low-complexity configuration.

The mean data, shown in figure 5b, are similar to those for observer LSI. Figure 6 shows the constancy indices for the individual observers. The mean constancy indices are 0.14 in the low-complexity condition and 0.25 in the high-complexity condition. Although there is considerable individual variation in the constancy indices shown in figure 6, increasing complexity leads to increased constancy for nine out of the ten observers and the mean constancy indices are significantly different (paired two-tailed t-test, $t_9 = 3.30$, $p < 0.01$). In the invalid-cue condition, increasing scene complexity does lead to better color constancy.

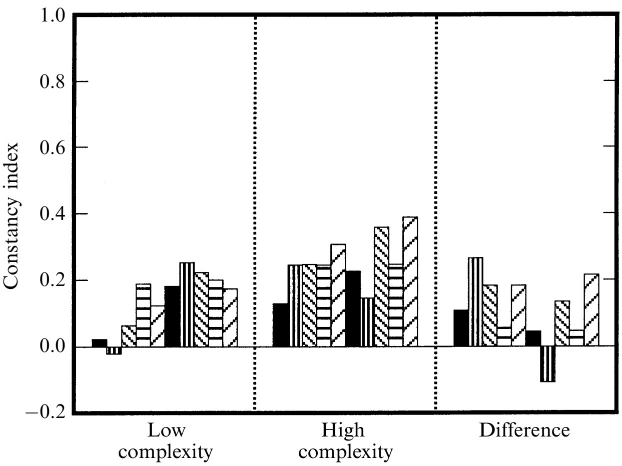

Figure 6. Experiment 1. Individual constancy indices for the invalid-cue condition. Same format as in figure 4.

3 Experiment 2: Effect of cues to depth

3.1 *Methods*

3.1.1 *Overview.* To investigate the effect of cues to depth on color constancy, we repeated experiment 1 under conditions where cues to the three-dimensional structure of the scene were reduced. Rather than viewing the chamber directly, observers viewed it through a telescopic viewing system (TVS). Viewing through the TVS eliminated depth cues provided by parallax (small head movements) and accommodation in the direct view conditions. The effect of removing depth cues was evaluated by comparing results from experiment 2 with those obtained for the same observers in experiment 1.

3.1.2 *Observers.* Nine observers (four male and five female; mean age 23.7 years) participated in experiment 2. Eight of these also participated in experiment 1 (see footnotes 2 and 3). The additional observer had normal acuity and color vision (assessed as described previously), was naïve to the purpose of the experiments, and was paid for his participation.

3.1.3 *Stimuli.* The chamber and its configurations were the same as for experiment 1. Observers viewed the chamber monocularly (right eye) through the TVS. The TVS was based on a pair of Nikon 5X15D CF binoculars and two front-surface mirrors. The binoculars were mounted rigidly to control their position and orientation. They had a magnification of 5×, a distal field of view of 9 deg, and an apparent field of view of 45 deg. The objective lens had an effective diameter of 15 mm, and the exit pupil had a diameter of 3 mm. The image of the scene was relayed to the binoculars by reflection off two front-surface mirrors. To compensate for the 5× magnification of the binoculars, the optical distance from the test patch to the binocular position was 5 times the optical distance from the scene to the direct view position. This arrangement made the angular sizes of the objects in the scene close to the same for the direct and TVS viewing modes.

To match the spectral properties of the stimuli to those used in experiment 1, the illuminant and projection colorimeter calibrations were corrected for the spectral transmission of the binoculars. To measure the relative spectral transmission of the binoculars, light from a slide projector was passed though the binoculars, imaged on a Spectralon (PhotoResearch PR650) plate, and measured. The spectrum so obtained was divided by that of the same light source imaged on the plate without the binoculars in place. To measure the absolute transmission of the binoculars at a reference wavelength, light from a diode laser (Alpec-Team 5 mW laser pointer) was passed through the binoculars and then imaged on a radiometric sensor (United Detector Technology 350 Optometer). The power measured by the sensor was divided by that from the same light source measured without passing though the binoculars. This quotient was taken as the absolute transmission of the binoculars at 672 nm. (Approximately 80% of the power from the laser is confined to 672 ± 5 nm.) The absolute spectral transmission of the binoculars was calculated by normalizing the relative spectral transmission to the correct value at 672 nm. This normalization indicated that the binoculars transmitted nearly all of the available light at their peak transmission wavelength (roughly 455 nm). All apparatus calibrations and stimulus measurements for the TVS conditions were performed with the radiometer at the TVS observing position, but without the binoculars in place. These calibrations were then corrected for the absolute spectral transmission of the binoculars.

Even when the spectral transmission of the binoculars is accounted for, matching the luminances of stimuli viewed directly and through the TVS does not match their retinal illuminances. This is because for our viewing conditions the exit pupil of the binoculars is smaller than the entrance pupil of the eye. We took this effect into account when specifying the stimuli used in experiment 2. The luminances provided in table 1 for the experiment 2 stimuli have been corrected: a stimulus of the specified luminance seen directly would produce the same retinal illuminance as the actual stimulus seen through the TVS. To make the correction, we estimated observers' pupil diameters to have been 4.08 mm in experiment 1 and assumed that they were larger than 3 mm in experiment 2. Thus the measured spectra, corrected for binocular transmission, were multiplied by 0.54 to obtain the specified values. Our estimate of pupil diameter was based on the formula provided by Trezona (1983) and the luminance of the area immediately surrounding the test patch.

As specified in table 1, the 'aqua', 'orange-low', and 'orange-high' illuminants were not perfectly matched between experiments 1 and 2. The differences between the illuminants used in the two experiments may be taken as an indication of the precision of our stimulus control across viewing modalities. To compensate for the light loss in TVS viewing (see above) in experiment 2, we used six theater lamps (two for each primary) above the chamber rather than three as in experiment 1. This difference in configuration and light levels made it difficult to match the illuminants more precisely. The variation of illumination across sessions in both experiments 1 and 2 was considerably smaller than the between-experiment variation.

3.1.4 *Procedure.* The procedure was the same as in experiment 1.

3.2 *Results*

Figure 7a shows observer LSI's data from experiment 2 for the valid-cue condition. These data are similar to the corresponding data obtained with direct viewing (experiment 1, figure 3a). For TVS viewing, the constancy index for the low-complexity configuration is 0.74 and for the high-complexity configuration is 0.83, compared with 0.88 and 0.85 for direct viewing. Figure 7b shows that the same pattern holds for the mean data from all observers. Figure 8 provides the constancy indices obtained for the nine individual observers who participated in experiment 2. The lower constancy index shown by observer LSI in experiment 2 for the low-complexity configuration is not a general feature of the data. The mean valid-cue condition indices from experiment 2 were 0.82 and 0.85 for the low-complexity and high-complexity configurations, respectively. The corresponding indices from experiment 1 were 0.87 and 0.86. Figure 9a shows the difference in individual observer, valid-cue constancy indices for the eight observers who participated in both experiments 1 and 2. A comparison of these indices showed no significant difference (paired two-tailed t-test, $t_7 = 1.19$).

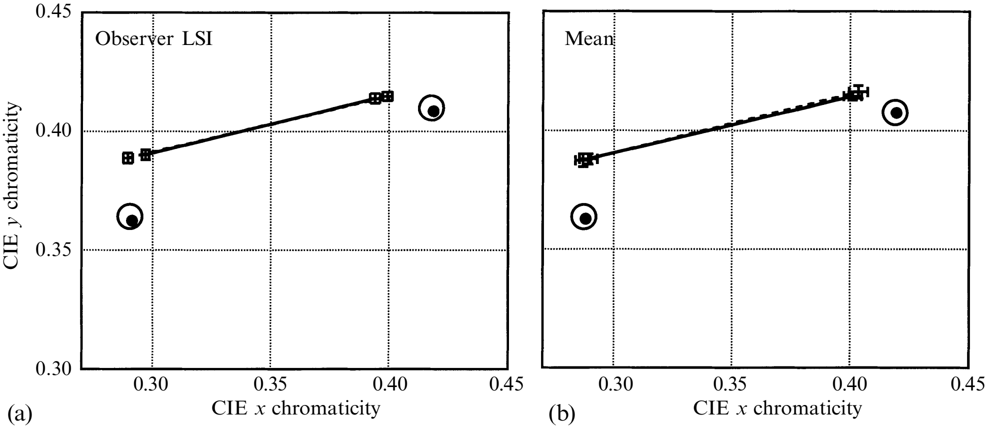

Figure 7. Experiment 2. (a) Single-observer data and (b) mean data for both complexity configurations, valid-cue condition. Same format as figure 3.

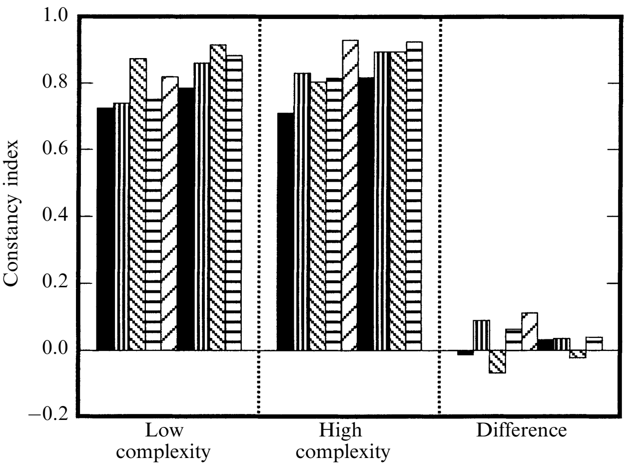

Figure 8. Experiment 2. Individual constancy indices for the valid-cue condition. Same format as figure 4.

Figures 10a and 10b summarize the data from experiment 2 for the invalid-cue condition. These figures are in the same format as figures 7a and 7b above. Observer LSI's data (compare figures 5a and 10a) differ somewhat between the two experiments. Her constancy indices were 0.07 and 0.21 for the low-complexity and high-complexity

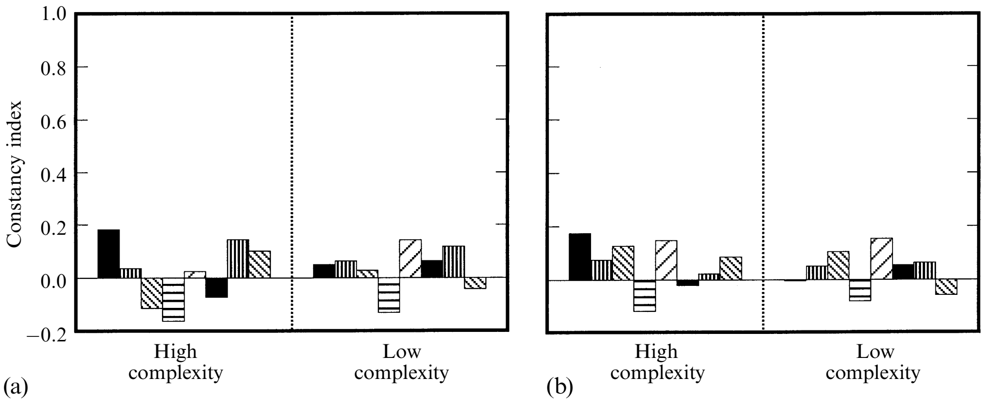

Figure 9. Differences in individual-observer constancy indices between experiments 1 and 2: (a) valid-cue condition; (b) invalid-cue condition. Data are for the eight observers who participated in both experiments.

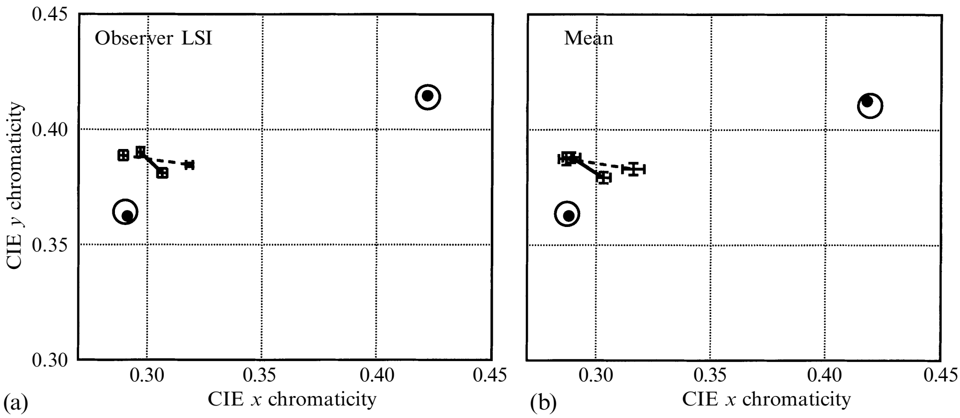

Figure 10. Experiment 2. (a) Single-observer data and (b) mean data for both complexity configurations, invalid-cue condition. Same format as in figure 3.

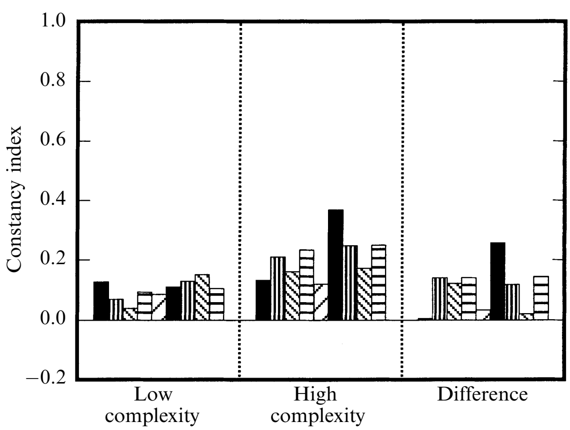

Figure 11. Experiment 2. Individual constancy indices for the invalid-cue condition. Same format as figure 4.

configurations, respectively, compared with the values of 0.22 and 0.36 obtained in experiment 1. The mean data, however, do not reveal much of a difference (compare figures 5b and 10b). Figure 11 shows the individual constancy indices in experiments 2 for the invalid-cue condition. The mean index was 0.10 for the low-complexity configuration and 0.21 for the high-complexity configuration, compared with the values of

0.14 and 0.25 obtained in experiment 1. Figure 9b shows the difference in single-observer, invalid-cue constancy indices for the eight observers who participated in both experiments 1 and 2. A comparison of these indices showed no significant difference (paired two-tailed t-test, $t_7 = 1.78$).

4 Discussion

The results from experiment 1 are quite straightforward. Under conditions where there are many valid cues to the illuminant change, increasing scene complexity does not increase constancy. When the number of valid cues to constancy is decreased, however, a different result emerges. Here, increasing scene complexity leads to improved constancy.

One way to understand the difference between the valid-cue and invalid-cue conditions of experiment 1 might be as follows. In the valid-cue conditions, many cues present in the image are consistent with the actual illuminant change. Because there is little inconsistency between cues, any additional consistent cues provided by increasing scene complexity have little effect. In essence, performance is already at ceiling. In the invalid-cue conditions, however, the situation is different. The manipulation of the cardboard produces a cue-conflict situation. The cue provided by the average light reflected to the observer, for example, does not provide valid information about the illuminant change. Constancy is quite poor in this condition. Our scene complexity manipulation provides additional, valid, cues. For example, the light reflected from the 'white' square of the Macbeth Color Checker provides a good indicator of the illuminant. When performance is not at ceiling, the additional cues matter.

Our results answer directly the question whether scene complexity is necessary for good color constancy. It is not. For low-complexity configuration, valid-cue condition, measured constancy was excellent and comparable with the high levels of constancy found for very rich, nearly natural scenes (Brainard 1998). Our conclusion here is consistent with that of Valberg and Lange-Malecki (1990), who compared constancy for uniform backgrounds and Mondrian scenes in what was essentially a valid-cue condition.

At the same time, we were able to identify circumstances where scene complexity did affect constancy (the invalid-cue condition). Whether complexity improves color constancy depends on the stimulus conditions. Our results thus suggest that it is worth pursuing a sharper formulation of when and how complexity affects human color constancy. A few authors have suggested that the main role of complexity is to improve the ability of observers to segment a scene into differently illuminated regions (Adelson 1999; Gilchrist et al 1999). Our experiments employed diffusely illuminated stimuli and did not place strong segmentation demands on the observer. See below for further discussion of this point.

Our complexity manipulation consisted of adding objects to the chamber and was not intended to allow isolation of a single cue. It is currently not known what image cues mediate human color constancy (but see McCann 1994; Kraft and Brainard 1999; Yang 1999; Maloney and Yang 2001). Given an arbitrary set of potential cues (ie statistics computed from an image), it is not necessarily possible to manipulate each one in isolation. For example, if we vary the chromaticity of the most luminous image region we will also affect the chromaticity of the average light reflected to the observer and the covariance of the chromaticities in the scene. Although one can define a set of image statistics that can be manipulated independently, there is no guarantee that these are in fact the ones used as cues by the human visual system. To advance our understanding of the effect of image complexity on constancy, we feel that further theoretical consideration of how best to operationalize complexity is required. Whatever the result of such consideration, our current results emphasize the importance of studying the effect of image manipulations on constancy across a range of conditions (see also Kraft and Brainard 1999; Brainard et al 2001). Note that in the literature most studies of constancy are analogous to our valid-cue condition. Data from such studies cannot provide a complete picture of how the human visual system separates illuminant and surface variation.

Examination of figures 4, 6, 8, and 11 reveals considerable between-observer variation in the effect of complexity on color constancy. Particularly striking is the variation in the effect for the invalid-cue condition. It may be that, for this condition, different observers are using different cues. Although the overall effect of complexity emerges clearly, it is important to keep in mind that it is not revealed for all observers. Some caution in interpretation is also required as we do not have direct experimental measurements of the variability of the individual observer constancy indices.

The results of experiment 2 are also straightforward. Here, the manipulation of depth cues implemented by changing from direct to TVS viewing has no effect on performance, either for valid-cue or invalid-cue conditions. Note that we did not manipulate all of the depth information contained in the scene. Indeed, most monocular static depth cues were unaffected by the change from direct to TVS viewing. It may be that the individual variation in the effect of the depth manipulation (see figure 9) is due to individual observers placing different weights on the various available depth cues. In addition, it may be that more aggressive manipulations would have produced an overall effect. Our choice of manipulation was driven by a desire to understand various reports, including a preliminary report from our laboratory, that perception of color for images presented on CRT monitors differs from that of directly viewed scenes (Schirillo et al 1990; Berns and Gorzynski 1991; Agostini and Bruno 1996; Brainard et al 1997b; but see Savoy and O'Shea 1993). As one step towards this goal, we wanted to determine whether the absence of dynamic or accommodative depth cues in scenes rendered on monitors could play a role in explaining the purported differences. Our current results show that they do not, at least for the class of scenes we studied.

There are additional reasons for caution against generalizing from our results to the conclusion that depth information never plays a role in color perception. First, it is quite clear that under some circumstances the perceived three-dimensional shape of objects can influence perceived color (eg Knill and Kersten 1991; Bloj et al 1999).[7] These effects seem to be mediated by the perception of shadows or of mutual reflection between adjacent surfaces. Presumably such effects were small for our stimuli. Second, in the experiments reported here, we measured successive color constancy—the effect of an illumination change over time—and there was little change in the spatial distribution of the illumination across the stimulus combinations for which we assessed constancy. Other investigators have studied simultaneous color constancy, where the comparisons of interest are across differently illuminated regions of a single scene (eg Gilchrist 1977, 1980; Arend and Reeves 1986; Brainard et al 1997a; Bauml 1999). In simultaneous constancy, manipulation of depth cues can have a large effect on perceived color (Gilchrist 1977, 1980). This effect is presumably mediated through a change in the way the visual system segments the scene into regions of different apparent illumination. Such segmentation processes would be expected to play a crucial role in simultaneous color constancy (see Adelson 1999; Gilchrist et al 1999) but less so in successive color constancy (see Brainard et al 2001).

Acknowledgements. Supported by EY 10016 and EY 006745. We thank James Schirillo for organizing a symposium on the role of scene complexity in constancy and for providing references to the early literature.

[7] We use the term color in an inclusive sense: many of the specific experiments that study the effect of depth cues employ isochromatic stimuli and assess the lightness dimension of color appearance.

References

Adelson E H, 1999 "Lightness perception and lightness illusions", in *The New Cognitive Neurosciences* 2nd edition, Ed. M Gazzaniga (Cambridge, MA: MIT Press) pp 339 – 351

Agostini T, Bruno N, 1996 "Lightness contrast in CRT and paper-and-illuminant displays" *Perception & Psychophysics* **58** 250 – 258

Arend L E, 1994 "Surface colors, illumination, and surface geometry: intrinsic-image models of human color perception", in *Lightness, Brightness and Transparency* Ed. A L Gilchrist (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 159 – 213

Arend L E, Reeves A, 1986 "Simultaneous color constancy" *Journal of the Optical Society of America A* **3** 1743 – 1751

Bauml K H, 1994 "Color appearance: effects of illuminant changes under different surface collections" *Journal of the Optical Society of America A* **11** 531 – 542

Bauml K H, 1999 "Simultaneous color constancy: how surface color perception varies with the illuminant" *Vision Research* **39** 1531 – 1550

Berns R S, Gorzynski M E, 1991 "Simulating surface colors on CRT displays: the importance of cognitive clues", paper presented at the AIC Conference: *Colour and Light* Sydney, Australia, 25 – 28 June

Bloj M, Kersten D, Hurlbert A C, 1999 "Perception of three-dimensional shape influences colour perception through mutual illumination" *Nature* **402** 877 – 879

Brainard D H, 1998 "Color constancy in the nearly natural image. 2. Achromatic loci" *Journal of the Optical Society of America A* **15** 307 – 325

Brainard D H, Brunt W A, Speigle J M, 1997a "Color constancy in the nearly natural image. 1. Asymmetric matches" *Journal of the Optical Society of America A* **14** 2091 – 2110

Brainard D H, Freeman W T, 1997 "Bayesian color constancy" *Journal of the Optical Society of America A* **14** 1393 – 1411

Brainard D H, Kraft J M, Longère P, 2001 "Color constancy: developing empirical tests of computational models", to appear in *Colour Perception: From Light to Object* Eds R Mausfeld, D Heyer (Oxford: Oxford University Press)

Brainard D H, Rutherford M D, Kraft J M, 1997b "Color constancy compared: experiments with real images and color monitors" *Investigative Ophthalmology & Visual Science* **38**(4) S476

Brainard D H, Wandell B A, 1986 "Analysis of the retinex theory of color vision" *Journal of the Optical Society of America A* **3** 1651 – 1661

Brainard D H, Wandell B A, 1992 "Asymmetric color-matching: how color appearance depends on the illuminant" *Journal of the Optical Society of America A* **9** 1433 – 1448

Buchsbaum G, 1980 "A spatial processor model for object colour perception" *Journal of the Franklin Institute* **310** 1 – 26

Burnham R W, Evans R M, Newhall S M, 1957 "Prediction of color appearance with different adaptation illuminations" *Journal of the Optical Society of America* **47** 35 – 42

Burzlaff W, 1931 "Methodologische Beiträge zum Problem der Farbenkonstanz" *Zeitschrift für Psychologie* **119** 177 – 235

Chichilnisky E J, Wandell B A, 1996 "Seeing gray through the on and off pathways" *Visual Neuroscience* **13** 591 – 596

Delahunt P B, Brainard D H, 2000 "Control of chromatic adaptation: signals from separate cone classes interact" *Vision Research* **40** 2885 – 2903

D'Zmura M, Colantoni P, Knoblauch K, Laget B, 1997 "Color transparency" *Perception* **26** 471 – 492

D'Zmura M, Iverson G, 1993 "Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces" *Journal of the Optical Society of America A* **10** 2148 – 2165

D'Zmura M, Iverson G, Singer B, 1995 "Probabilistic color constancy", in *Geometric Representations of Perceptual Phenomena: Papers in Honor of Tarow Indow's 70th Birthday* Eds R D Luce, M D'Zmura, D Hoffman, G Iverson, A K Romney (Mahwah, NJ: Lawrence Erlbaum Associates) pp 187 – 202

Gelb A, 1929/1938 "Die 'Farbenkonstanz' der Sehdinge", in *Handbuch der normalen und pathologischen Physiologie* Band 12, 1. Hälfte *Receptionsorgane II* Eds W A von Bethe, G von Bergmann, G Embden, A Ellinger (1938, Berlin: Springer) pp 594 – 678 [selection translated in *A Source Book of Gestalt Psychology* Ed. W D Ellis (1938, New York: Harcourt Brace; London: K Paul, Trench, Trubner) pp 196 – 209]

Gerbino W, Stultiens C I F H J, Troost J M, Weert C M M de, 1990 "Transparent layer constancy" *Journal of Experimental Psychology: Human Perception and Performance* **16** 3 – 20

Gilchrist A L, 1977 "Perceived lightness depends on perceived spatial arrangement" *Science* **195** 185

Gilchrist A L, 1980 "When does perceived lightness depend on perceived spatial arrangement?" *Perception & Psychophysics* **28** 527 – 538
Gilchrist A, Jacobsen A, 1984 "Perception of lightness and illumination in a world of one reflectance" *Perception* **13** 5 – 19
Gilchrist A, Kossyfidis C, Bonato F, Agostini T, Cataliotti J, Li X, Spehar B, Annan V, Economou E, 1999 "An anchoring theory of lightness perception" *Psychological Review* **106** 795 – 834
Helson H, Michels W C, 1948 "The effect of chromatic adaptation on achromaticity" *Journal of the Optical Society of America* **38** 1025 – 1032
Henneman R, 1935 "A photometric study of the perception of object color" *Archives of Psychology* **27**(179) 5 – 89
Hurlbert A C, 1998 "Computational models of color constancy", in *Perceptual Constancy: Why Things Look as They Do* Eds V Walsh, J Kulikowski (Cambridge: Cambridge University Press) pp 283 – 322
Hurlbert A C, 1999 "Is color constancy real?" *Current Biology* **12** R558 – R561
Knill D C, Kersten D, 1991 "Apparent surface curvature affects lightness perception" *Nature* **351** 228 – 230
Kraft J M, Brainard D H, 1999 "Mechanisms of color constancy under nearly natural viewing" *Proceedings of the National Academy of Sciences of the USA* **96** 307 – 312
Lucassen M P, Walraven J, 1993 "Quantifying color constancy: evidence for nonlinear processing of cone-specific contrast" *Vision Research* **33** 739 – 757
MacLeod R B, 1932 "An experimental investigation of brightness constancy" *Archives of Psychology* **24**(135) 5 – 102
McCann J J, 1994 "Psychophysical experiments in search of adaptation and the gray world", paper presented at the IS&T's 47th Annual Conference, Rochester, NY
McCann J J, McKee S P, Taylor T H, 1976 "Quantitative studies in retinex theory: A comparison between theoretical predictions and observer responses to the 'Color Mondrian' experiments" *Vision Research* **16** 445 – 458
Maloney L T, 1999 "Physics-based approaches to modeling surface color perception", in *Color Vision: From Genes to Perception* Eds K T Gegenfurtner, L T Sharpe (Cambridge: Cambridge University Press) pp 387 – 416
Maloney L T, Wandell B A, 1986 "Color constancy: a method for recovering surface spectral reflectances" *Journal of the Optical Society of America A* **3** 29 – 33
Maloney L T, Yang J N, 2001 "The illuminant estimation hypothesis and surface color perception", to appear in *Colour Perception: From Light to Object* Eds R Mausfeld, D Heyer (Oxford: Oxford University Press)
Metelli F, 1974 "Perception of transparency" *Scientific American* **230**(4) 90 – 98
Savoy R L, O'Shea R P, 1993 "Color constancy with reflected and emitted light" *Perception* **22** Supplement, 61
Schirillo J, Reeves A, Arend L, 1990 "Perceived lightness, but not brightness, of achromatic surfaces depends on perceived depth information" *Perception & Psychophysics* **48** 82 – 90
Speigle J M, Brainard D H, 1999 "Predicting color from gray: the relationship between achromatic adjustment and asymmetric matching" *Journal of the Optical Society of America A* **16** 2370 – 2376
Trezona P W, 1983 "Luminance level conversions to assist lighting engineers to use fundamental visual data" *Lighting Research and Technology* **15**(2) 83 – 88
Valberg A, Lange-Malecki B, 1990 "'Color constancy' in mondrian patterns: a partial cancellation of physical chromaticity shifts by simultaneous contrast" *Vision Research* **30** 371 – 380
Werner J S, Walraven J, 1982 "Effect of chromatic adaptation on the achromatic locus: the role of contrast, luminance and background color" *Vision Research* **22** 929 – 944
Yang J N, 1999 *Illuminant Estimation in Surface Color Perception* unpublished PhD thesis, New York University, New York