Limits of lightness identification for real objects under natural viewing conditions

Rocco Robilotto SUNY College of Optometry, New York, NY, USA

Qasim Zaidi SUNY College of Optometry, New York, NY, USA

We examined whether observers veridically perceive the reflectances of real objects under natural viewing conditions. A new forced-choice paradigm was used to measure observers' abilities to identify (not match) the reflectance of randomly crumpled gray papers across two levels of illumination, and also to simultaneously measure brightness discrimination thresholds for the same objects. Accuracy of lightness identification differed qualitatively among observers. By explicitly manipulating observer strategies, we show that when observers use brightness dissimilarity, their performance is similar to lightness identification. A brightness adaptation model simulates how instead of extracting lightness, observers can rely on perceived relative brightness to achieve the measured degrees of lightness identification.

Keywords: lightness, brightness, constancy, natural stimuli, adaptation, material identification

1. Introduction

When viewing achromatic surfaces, in some instances, it is possible to separate the *lightness* of a surface from its *brightness*. Lightness is the mean reflectance, where reflectance is the fraction of incident light reflected back by the surface, and is solely a property of the surface (Evans, 1974). Brightness refers to apparent luminance, where luminance is the light reflected from the surface, and is a function of both the incident illumination and the surface reflectance. Perceived brightness differs from the physical quality of luminance because brightness is affected by adaptation (Craik, 1938; Helson, 1964) and lateral interactions (Chevreul, 1839; Zaidi, 1999). Reflectance is also a physical quality, and the lightness of a surface is inferred either visually or cognitively by separating the information in the scene into environmental and material changes (Helmholtz, 1962; Hering, 1964).

Numerous studies have treated lightness constancy as a perceptual phenomenon in which the lightness of a surface remains invariant to changes in illuminant, shape, or background. Despite the fact that Kardos (1934) showed that the surface of a target patch in shadow is always seen as darker than it actually is, and despite the fact that results vary widely across studies (Bruno, 1994; Gilchrist, 1988; Rutherford & Brainard, 2002; J. A. Schirillo & Arend, 1995), lightness constancy is generally thought to be well established. On the other hand, although light adaptation tends to reduce brightness differences across different scenes, and relative brightness within scenes is relatively constant, brightness constancy has been shown to be false in numerous studies, and does not even survive general everyday observations.

Beginning with Marzynski (1921), a few studies have made separate matches of brightness and lightness of objects under the same conditions (Arend & Goldstein, 1987; Arend & Spehar, 1993a, 1993b; Bloj & Hurlbert, 2002; Schirillo, Reeves, & Arend, 1990). These studies have been instrumental in distinguishing lightness constancy from brightness constancy. In this work, we suggest that lightness constancy can be defined in a direct performance-based manner as the ability to identify two objects as having the same lightness across physically different illumination conditions. This definition has the virtue of being functionally relevant, because the general connotation is that if the lightness of objects is constant across illumination conditions, it can be an aid to identifying objects (Adelson, 2000; Arend & Goldstein, 1987; Gilchrist & Jacobsen, 1984; MacEvoy & Paradiso, 2001). Furthermore, the task of identifying objects of similar lightness across physically different illuminants lets the observer apply all possible strategies to counter any appearance changes (Zaidi, 2001). The first aim of this work is to measure the accuracy of lightness identification for real objects under naturalistic viewing conditions. The second aim is to examine whether lightness judgments are based on extracting surface lightness, or are based on relative brightness.

Consider the demonstration in Figure 1. Look at the four crumpled objects in the two compartments. Three of the objects are made of identical gray paper while one is made of a different shade of gray paper. The compartment on the right is receiving half the illumination of the compartment on the left. Which is the odd object? To correctly perform this task, you can first use brightness discrimination to select the compartment that contains the pair of objects with reflectances different from each other. However, once this compartment has been chosen, brightness is no longer a sufficient cue and you have to identify the lightness within that compartment that is similar to the two objects in the other compartment, or equivalently, the one

doi:10.1167/4.9.9 Received February 12, 2004; published September 15, 2004 ISSN 1534-7362 © 2004 ARVO

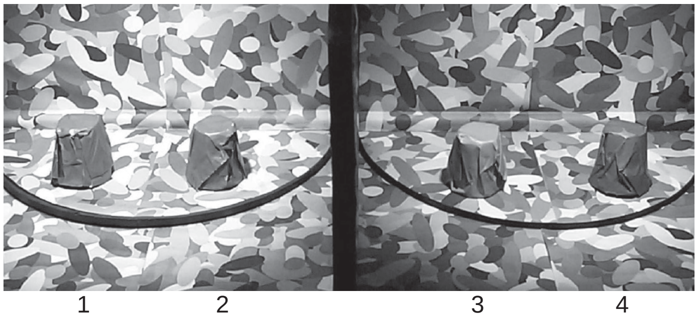

Figure 1. Setup for Experiments 1 and 2. Backgrounds in each compartment have the same mean reflectance and reflectance distribution. An independent spotlight illuminates each compartment with one compartment receiving half the illumination of the other (full illumination on the left in this example). Three of the crumbled objects are of the same reflectance and make up a standard set, while the one test object is of a different reflectance.

that is different. In other words, you first *discriminate within illuminants*, then *identify across illuminants*.

If observers could estimate lightness accurately, identification should be veridical except when the difference between test and standard in the right compartment is less than the threshold for brightness discrimination. In this paradigm, observers' abilities to identify the odd-lightness objects are measured simultaneously with brightness discrimination to obtain these lower bounds.

In Figure 2, the two objects from the right compartment in Figure 1 have been kept in the same place. The two objects from the left compartment have been placed behind them to facilitate comparison under a single illuminant. It is clear from this presentation that object 3 is the odd object, of a lighter shade than the other three. If you identified the object correctly, your response is consistent with lightness constancy.

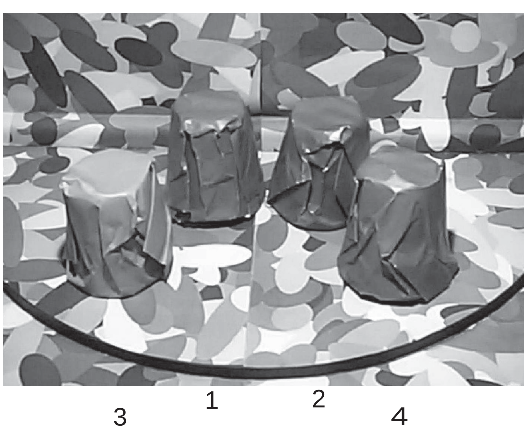

Figure 2. For demonstration purposes, objects 1 and 2 from the left compartment of Figure 1 have been placed behind objects 3 and 4 in the right compartment. When all four objects are under the same illumination, it becomes obvious that object 3 is the odd test object with a higher reflectance than the three identical standard objects.

It has been claimed that an object's material (Montag & Berns, 2000; Nayar & Oren, 1995; Nishida & Shinya, 1998), three-dimensional (3D) shape (Adelson & Pentland, 1996; Pessoa, Mingolla, & Arend, 1996; Sun & Perona, 1996), and spatial arrangement (Gilchrist, 1980; Gilchrist, Delman, & Jacobsen, 1983) provide many cues that help the observer ascertain physical properties such as reflectance. In this study, instead of presenting flat physical stimuli or stimuli generated on a computer monitor, crumpled paper was used to provide texture and facets that an observer might encounter in a natural setting. As in Figure 1, crumpled gray papers of varying levels of reflectance were presented side by side in two compartments, each illuminated by a separate light source differing in intensity by a factor of two. One compartment contained two standard objects; the other contained one standard and one test object. In Experiment 1, observers were asked to identify the object with the different reflectance. In Experiment 2, observers were asked to choose the object that differed most in brightness.

The purpose of this study is to present a direct method for quantifying observers' abilities to perform lightness identification tasks, and to examine whether observers estimate reflectances or use brightness cues in lightness identification. By simultaneously comparing the proportion of responses in which the correct object was chosen with the proportion of responses in which either object in the correct compartment was chosen, it is possible to compare lightness identification thresholds to brightness discrimination thresholds. See Khang and Zaidi (2002) for a similar method applied to identification of spectral transparency.

2. Experiment 1: Lightness identification

2.1 Stimuli

The stimuli consisted of crumpled papers varying in levels of reflectance. A 19-step commercial Color-aid gray scale set (Color Aid Corp., New York) with reflectances ranging from 3–90% was used as the template set. The sheets were copied on a Canon Color Laser Copier 2400 with seven different levels of copy darkness, creating papers of 133 different reflectances. The copies from the machine were cut into 20 x 14 cm pieces, and their reflectance was measured. The papers were then crumpled by hand around 7-ounce paper cups that measured 6-cm tall, 6-cm wide at the base, and 4-cm wide at the top.

To calibrate the reflectance of the stimuli paper, precrumpled sheets were laid flat in the middle of one compartment of the apparatus and a Spectra Scan 650 photometer (Photo Research Inc., Chatsworth, CA) was positioned 50 cm from the center of the paper at an angle of 55 deg from the normal. At this configuration under full illumination, the Color-Aid paper labeled "white," which had a known reflectance of 90%, had a luminance measured at 836 cd/m^2. A theoretical reflectance of 100% would therefore give a luminance reading of 930 cd/m^2. Luminance readings from the photometer for each pre-crumpled gray paper used as a stimulus were then taken and divided by 930 cd/m^2 to obtain individual reflectance ratios.

2.2 Apparatus

Figure 1 is a photograph of the observer's view of the stimuli presented in one trial. Figure 3 presents top and side drawings of the apparatus. The crumpled stimuli were presented in a rectangular wooden box with a partition dividing the box into two compartments. Each compartment had an open front and open top and measured 38-cm wide, 28-cm deep, and 25-cm high. The two outer walls and dividing partition were heightened with cardboard extensions to a total height of 67 cm. This prevented light from either compartment illuminating the other. Attached to the center of the floor of each compartment was a 1.5-cm thick rotating disk with a diameter equal to the width of its compartment. Each floor disk was bisected with a vertical wall 25-cm high, which acted as the back wall of its compartment when rotated to be perpendicular to the side walls. In this presentation position, the back walls and the floor were flush with the side walls. In each compartment, two pegs were mounted onto the disks 11 cm in front of the back wall and 14 cm from each other. Another pair of pegs was mounted on each rotating floor disk at the same positions behind the back wall, and was out of sight from the observer. Crumpled papers wrapped around cups were placed over the pegs. While the observers were viewing the stimuli of any given trial, a new set could be placed over the pegs behind the back wall. Once the observer signaled a response, the floor disks and back walls were rotated 180 deg and the new set was presented.

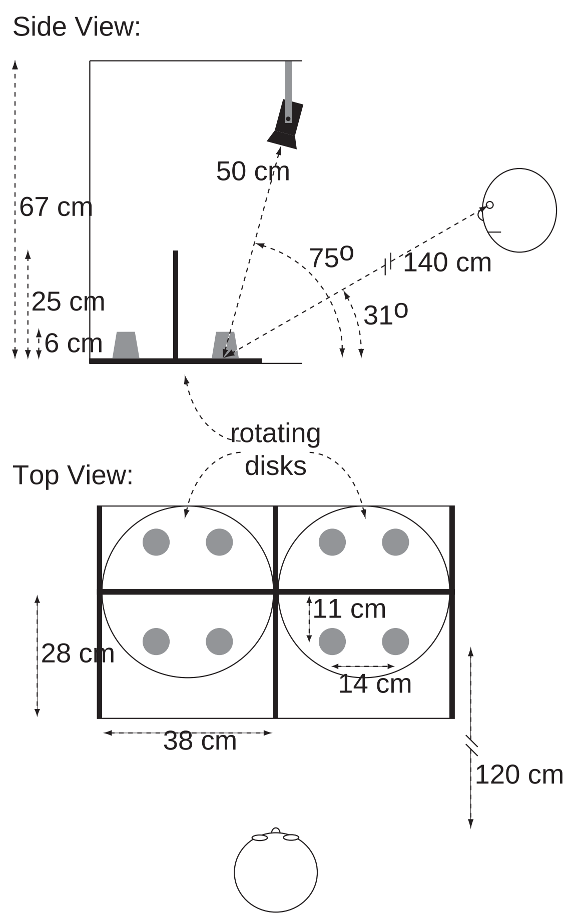

Figure 3. Schematic diagram of the experimental apparatus. Observers viewed the stimuli in an open-front box that was divided in half by a partition. Each compartment contained two objects and was illuminated from above by a separate light source. The back walls and objects were mounted on thin revolving floors disks. After each response by the observer, the disks were rotated 180 deg and a new stimulus set was presented.

The entire box was painted with matte black spray-paint (Krylon ultra-flat #1602, Cleveland, OH). In each compartment, the floor, the back wall, and each side wall were covered in paper consisting of a randomized pattern of overlapping gray-scale elliptical disks. The ellipses varied in axis size from 20 x 12 mm to 63 x 17 mm, and varied randomly in shade from near black to near white. The purpose of the randomly variegated background was to prevent observers from making direct brightness comparisons between stimuli and particular patches of background across the two compartments.

Each compartment was illuminated independently by a lamp (Tailored Lighting, Solar Simulator 38, Rochester, NY) with a color temperature of 4700 K and an angle of illumination 36-deg wide. Each lamp was suspended above and in front of its respective compartment. The central incident beams intercepted the floor in the center of each compartment 50 cm from the source, with a 75 deg angle of elevation. Observers could view the outside of the lamp casings, but the lamps themselves were angled away from the observers and could not be directly viewed. The room was otherwise dark. One compartment was left at full illumination while a 50% neutral density gelatin filter (Rosco, N.3 #3402, Stamford, CT) was placed immediately in front of the other compartment's light source, reducing its illumination in half. Photometric measurements were taken with the filter in place to ensure 50% transmittance and the filter was replaced at regular intervals to prevent heat damage from the lamp. The compartment containing the filter was randomly selected before each session.

2.3 Observers

Seven observers participated in Experiment 1. Observer KH had never participated in a psychophysical experiment, and did not know the issues behind the study. Observers RD, FK, ST, and SS had experience participating in psychophysical experiments but did not know the issues behind the study. Observers QZ and BK had extensive experience participating in psychophysical experiments, and understood the issues behind the study.

2.4 Procedure

A method of constant stimuli was used to measure discrimination and identification. Observers sat in front of the apparatus and viewed the stimuli binocularly, looking back and forth as they wanted. This presented a similar adaptation condition to a case where an observer in a natural scene is looking between directly illuminated objects and objects in the shade. The viewing distance from the stimuli was 140 cm with an angle of elevation of 31 deg. Three *standard* objects of equal reflectance were placed on three of the four pegs. On the fourth peg, a *test* object with a different reflectance was placed. The position of the test object was randomly assigned for each trial. Observers were instructed to identify the one object made from a different shade of gray paper. The following instructions were used:

> You will be presented with four pieces of gray crumpled paper. Three of the papers will be of an identical shade of gray. The fourth paper will be a slightly lighter or slightly darker shade of gray. You will be asked to decide which of the four papers is of a DIFFERENT MATERIAL from the other three. The papers will be placed in a box with two separate compartments. Each compartment will contain two of the papers. The compartments will be illuminated by different sources of light. In doing this task, you should first identify the side on which the papers are different from each other, and then decide which of those two papers is different from the two on the other side.

After the instructions were given, observers were shown a demonstration with stimuli that were not included in the actual trials. First, all objects were placed in the same compartment as shown in Figure 2. After the observers correctly identified the odd object made from the different shade of gray paper, they watched as two of the objects were moved to the other compartment under half illumination, as illustrated in Figure 1. It was explained that the odd-object just chosen was the same object to be chosen under the condition of dissimilar illuminations. A few of these practice demonstrations were run to ensure observers understood the task. On the back wall, above each object, a number was placed so that the observers could indicate their choice. Trials were self-paced with no time limit. At no time were the observers given any feedback on the accuracy of their responses.

Experiment 1 was separated into Experiment 1a and Experiment 1b with identical instructions and procedure. The only differences between them were the reflectance values of the stimuli used and the number of tests per standard. In Experiment 1a, 39 shades of stimuli divided into three groups of 13 were presented to two observers (ST and SS). Each of the three groups (Table 1) consisted of one set of standard objects plus six test objects of higher reflectance (+ deltas) and six test objects of lower reflectance (– deltas).

To confirm the results from Experiment 1a, Experiment 1b ran five additional observers (KH, QZ, BK, RD, and FK) under an identical experimental setup with a condensed number of conditions. In Experiment 1b, 27 shades of stimuli were divided into three groups of nine. Each of the three groups (Table 2) consisted of one set of standard objects plus four test objects of higher reflectance and four test objects of lower reflectance.

Trials for each condition were run over a series of sessions spanning several days. Each session used two of the three sets of standard objects. During a given session in Experiment 1a, sets of standards were presented with one of six test objects (3 higher reflectances and 3 lower reflectances). Separate sessions were run with the other six test objects. For each combination of standard and test, 8 repetitions were run under both high and low illumination, for two sessions, for a total of 16 trials per condition. For each observer, this totaled six sessions per experiment with 192 trials per session. During a given session in Experiment 1b, sets of standards were presented with one of four test objects (2 higher reflectances and 2 lower reflectances). Separate sessions were run with the other four test objects. For each combination of standard and test, 10 repetitions were run under both high and low illumination, for 10 trials per condition. For each observer, this totaled three sessions per

Tests (- deltas)						Standard	Tests (+ deltas)					
0.071	0.089	0.101	0.117	0.119	0.133	**0.142**	0.162	0.171	0.183	0.198	0.205	0.230
0.152	0.171	0.200	0.215	0.221	0.242	**0.268**	0.290	0.313	0.324	0.347	0.368	0.383
0.231	0.261	0.291	0.324	0.377	0.425	**0.435**	0.463	0.500	0.543	0.581	0.631	0.668

Table 1. Reflectance ratios of stimuli used in Experiment 1a. The three rows indicate the three sets of stimuli from low to high reflectance. Each standard set was matched with one of six test objects of higher reflectance, or one of six test objects of lower reflectance.

Tests (- deltas)				Standard	Tests (+ deltas)			
0.038	0.043	0.053	0.069	**0.078**	0.088	0.122	0.154	0.185
0.083	0.138	0.173	0.212	**0.230**	0.284	0.338	0.386	0.441
0.257	0.341	0.395	0.477	**0.516**	0.533	0.623	0.699	0.733

Table 2. Reflectance ratios of stimuli used in Experiment 1b. The three rows indicate the three sets of stimuli from low to high reflectance. Each standard set was matched with one of four test objects of higher reflectance, or one of four test objects of lower reflectance.

experiment with 160 trials per session. The order of the trials was randomized. Each session lasted approximately 40-60 min with a 10-min break in the middle.

2.5 Results

The four-alternative forced-choice (4AFC) paradigm used in this study can be analyzed like a standard 2 x 2 detection-discrimination AFC procedure (MacMillan & Creelman, 1991). First, observers discriminate brightness differences between pairs of objects under the same illumination and thus choose the side that contains the odd test object. Once a brightness difference has been discriminated, observers then identify which of the two objects on that side are different in reflectance from the two objects under the other illumination.

Figure 4a shows a hypothetical example of what data might look like for one set of conditions. For each test, the proportion of correct responses is plotted against the reflectance difference between standard and test. Open circles represent the proportion of correct side responses, that is, choosing either object in the compartment that contains the odd object. This is essentially a 2AFC task and results give psychometric brightness discrimination functions. Filled diamonds represent the proportion of correct object responses and give psychometric reflectance identification functions.

In Figure 4a, the proportion of correct responses, r, equals the proportion of discriminating or identifying, d, plus the probability of guessing correctly, γ, when discrimination or identification does not occur [i.e., $r = d(1\text{-}d)\gamma$]. γ is 0.5 for correct side and 0.25 for correct object. Threshold level for both brightness discrimination and lightness identification were set as the mid-point between γ and 100% correct (0.75 and 0.625, respectively). To normalize the response data so the two functions are equally scaled on the ordinate, detection and identification rates can be plotted after adjusting for guessing [i.e., $d = (r\text{-}\gamma)/(1\text{-}\gamma)$]. Both normalized functions (Figure 4b) are now equally scaled with chance at 0.0, threshold at 0.5, and 100% correct at 1.0. This makes it possible to directly compare both sets of response functions obtained simultaneously from the same sets of trials. Although the scaling of the two functions has been equated for proportions greater than chance, their normalized minimum response proportions are different. A 0% side correct response rate is equal to a normalized proportion of -1.0, whereas a 0% object correct response rate is equal to a normalized proportion of -0.33. Note that if response proportions are randomly lower than chance, it is probably due to noise. However, if they are systematically lower than chance, it is probably due to an incorrect strategy (see "Discussion").

Complete collections of hypothetical response data for a standard-test set are shown in Figure 5 for two photometer-based hypothetical observers. The normalized proportions of correct responses are plotted versus the reflectance differences between test and standard. Dashed curves represent the proportions of side correct responses, and solid curves represent proportions of object correct responses. The top plots represent conditions where the test objects are under the brighter illuminant, while the bottom plots

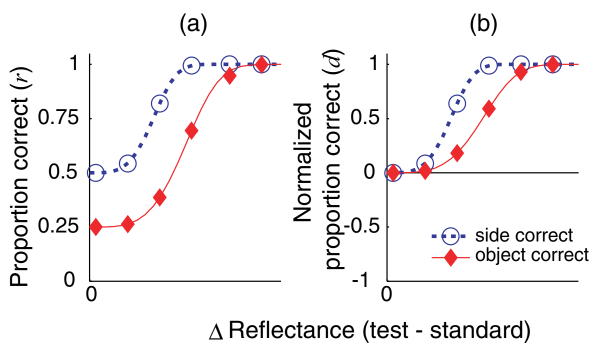

Figure 4. Hypothetical example of a single standard-test set where correct responses are plotted versus reflectance differences between standard and test. (a). Response rates are scaled as proportion correct. (b). The same response rates are corrected for guessing so that the two functions are equally scaled on the ordinate. Open circles represent correct side responses (choosing either object in the compartment that contains the test), while closed diamonds represent correct object responses (choosing the test object).

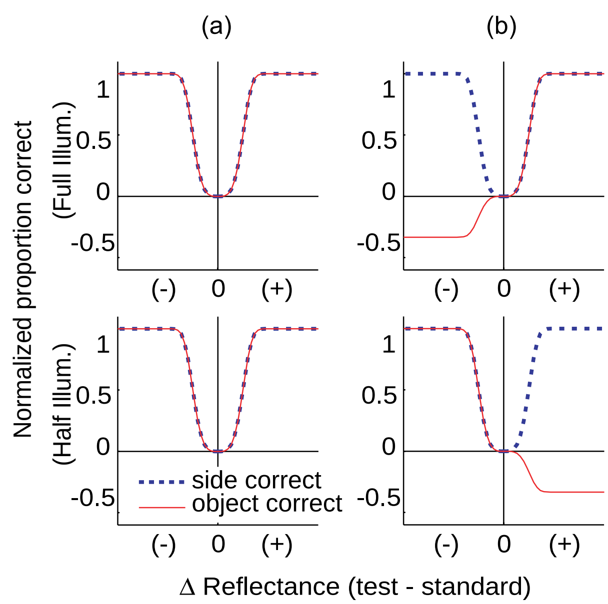

Figure 5. Hypothetical example of a set of conditions where proportions of correct responses (adjusted for guessing) are plotted versus reflectance difference between test and standard. Dashed lines represent proportion of correct side responses, while solid lines represent proportion of correct object responses. The top row indicates conditions in which the test objects are in the compartment under full illumination, while the bottom row indicates conditions in which the test objects are in the compartment under half illumination. (a). Responses of a hypothetical observer with perfect lightness identification limited only by brightness discrimination. (b). Responses of a hypothetical observer based on taking into account only luminances of the objects.

represent conditions where the test objects are under the darker illuminant. The vertical lines indicate the reflectance of the standard objects, with points to the right and left of it indicating test objects with higher or lower reflectances, respectively.

What information is available in the display, and what is the best any visual system could estimate? Backgrounds in the two boxes have similar statistics. The hypothetical observer in Figure 5a calculates the mean luminance of both backgrounds, and takes the ratio to obtain the ratio of illuminant intensities. This observer then measures the mean luminance of each of the four cups, and corrects for the relative illumination intensities to equate the reflectances of the standards across compartments. There is sufficient information to perform the lightness identification task; therefore, lightness constancy is limited only by the ability to discriminate within illuminants, and the response functions for brightness discrimination and lightness identification are superimposed. If, however, a photometer-based observer measured just the mean luminances of the four objects, and simply choose the object most different in luminance as the odd object, results would look like Figure 5b. When a test is under full illumination (top plot) and has a lower reflectance than the standards, its luminance will be lower than the standard in its compartment and closer to the luminance of the two standards in the other compartment. When a test is under half illumination (bottom plot) and has a higher reflectance than the standards, its luminance will be higher than the standard in its compartment and closer to the luminance of the two standards in the other compartment. Under these types of conditions, when the correct side is consistently chosen as containing the test, the standard on that side will consistently and incorrectly be chosen as the odd object. The correct object response rate will fall to 0%, which is equivalent to a normalized proportion of −0.33. In these conditions, if the reflectance difference continued to increase to a factor greater than the factor of illumination difference between the two compartments, the test would eventually become most different in luminance as well. This experiment, however, does not use sets of objects with reflectance differences of these magnitudes.

Figure 6 presents data from both observers in Experiment 1a in a format similar to Figure 5. The normalized proportion of correct responses is plotted versus the reflectance of the test objects. Both sets of data were fit with psychometric functions by means of maximum likelihood procedures (see "Appendix 5.1"). For each set of plots, the top row represents conditions where the test object was under full illumination, and the bottom row represents conditions where the test object was under half illumination. Each vertical column of plots represents a different standard object set, with the standard reflectance indicated by the vertical line. Each data point is the mean of 16 trials. Symbols that appear to be filled circles are points where side correct (open circles) and object correct (filled diamonds) responses fall on the same position.

Figure 7 presents response data for the five observers in Experiment 1b, plotted in an identical manner as Figure 6. Each data point is the mean of 10 trials. Because fewer test objects were used for each standard than in Experiment 1a, and fewer trials were run for each condition, psychometric functions were not fit to the data. Instead, data points are connected for grouping purposes only. For observer BK, conditions were only run with the low and middle reflectance standard sets.

All seven observers in both Experiments 1a and 1b displayed similar brightness discrimination. As test reflectances became progressively different from the standard reflectances, the proportions of side correct responses increased. In most cases, this increase occurred monotonically from chance to 100%. Occasional deviations from this trend did occur. For example, in Experiment 1b, when the high reflectance standard set was used and the test was in the compartment under full illumination (upper rightmost subplots), the correct response proportions never reached 100% with high reflectance test objects. In these conditions, the reflectance differences needed to reach threshold lay outside the range of stimuli used in this experiment.

In regards to lightness identification (i.e., choosing the correct object across different illuminants), observers fell into one of two categories. The first category includes the two observers from Experiment 1a and three of the five observers from Experiment 1b (KH, QZ, and BK). Their data are compatible with some degree of lightness constancy under almost all conditions. As the reflectance differences of the tests reached the level where brightness discrimination was achieved, lightness identification was usually achieved as well. In other words, when the observers could discriminate reflectance differences within a single illuminant, they could identify reflectance differences across different illuminants. Observers were using an estimate other than photometric luminance because in many conditions the object correctly chosen as most different in reflectance was not the object most different in luminance.

The difference in illumination is at least partially accounted for under these conditions and allows for veridical judgments of object material. However, even after the proportions are adjusted for guessing, in many cases lightness identification requires a larger reflectance difference to reach threshold level than does brightness discrimination. If constancy were perfect, the differences should be equal. There also seems to be an asymmetry in the data, as lightness identification seems to be systematically worse than brightness discrimination for lower reflectance tests under full illumination and higher reflectance tests under half illumination.

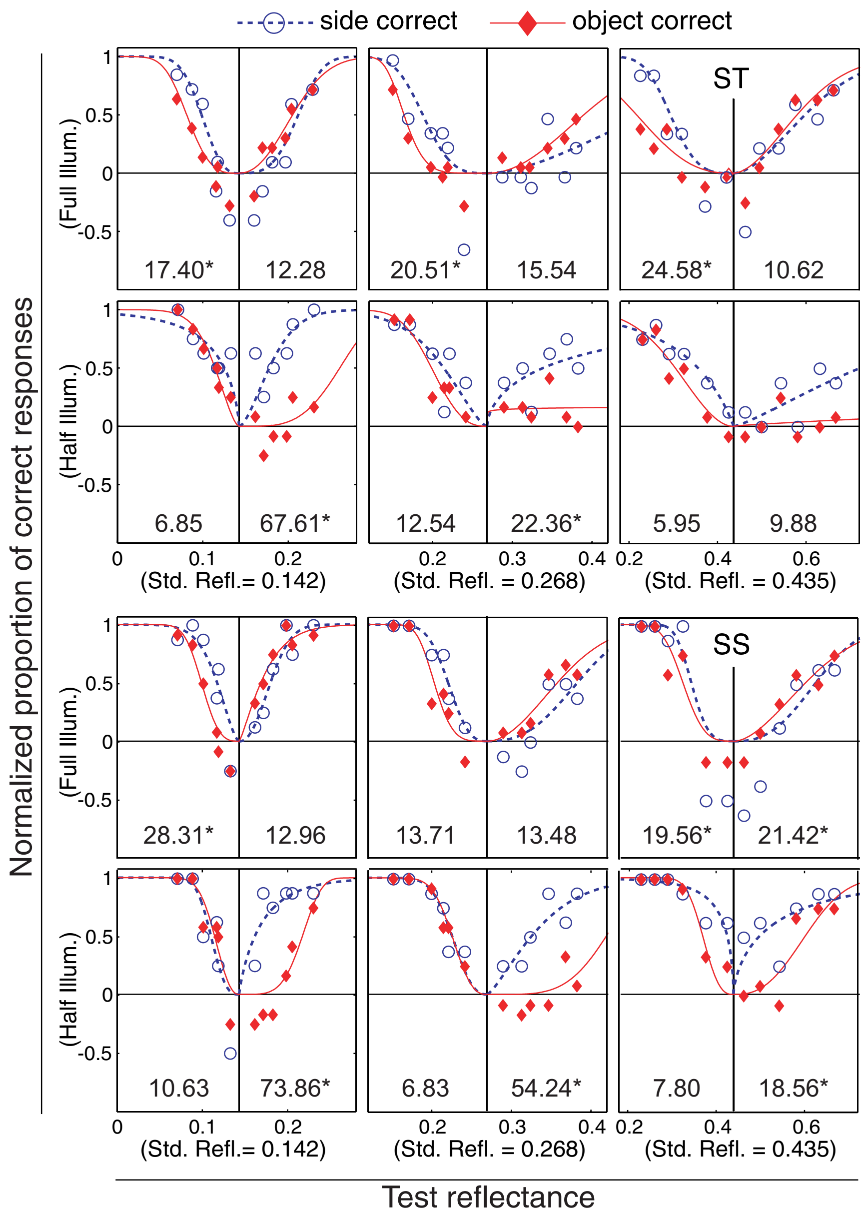

Figure 6. Results from the two observers in Experiment 1a. Normalized proportions of correct responses (adjusted for guessing) are plotted versus test reflectance. Brightness discrimination (proportion of correct side responses) is represented by open circles. Lightness identification (proportion of correct object responses) is represented by filled diamonds. For each set of plots, the top row indicates conditions where the test is under full illumination, while the bottom row indicates conditions where the test is under half illumination. The vertical line in each plot indicates the reflectance level of the standard set. Psychometric curves were fit to both sets of data using maximum likelihood ratios. For each lightness identification data set, the χ^2 value listed below it was calculated based on a separate likelihood ratio of observed response frequencies to model frequencies.

In contrast to the observers just described, the remaining two observers from Experiment 1b (RD and FK) displayed a reliable breakdown of lightness constancy under many conditions. The general trend for both of these observers involved identification response rates consistently less than chance for the following two types of conditions: (1) when test objects with higher reflectance than the standard were under full illumination, and (2) when test objects with lower reflectance than the standard were under half illumination. Normalized response rates of these conditions are consistently below zero (42 out of 48 data points across both observers in the above conditions). In fact, the larger the reflectance difference of the test in most of these conditions, the lower the proportion of correct responses. At the most extreme reflectance delta, the normalized proportion is often −0.33, equivalent to a non-normalized correct response rate of 0.0%. As shown by the brightness discrimination curves, observers RD and FK detected the side that

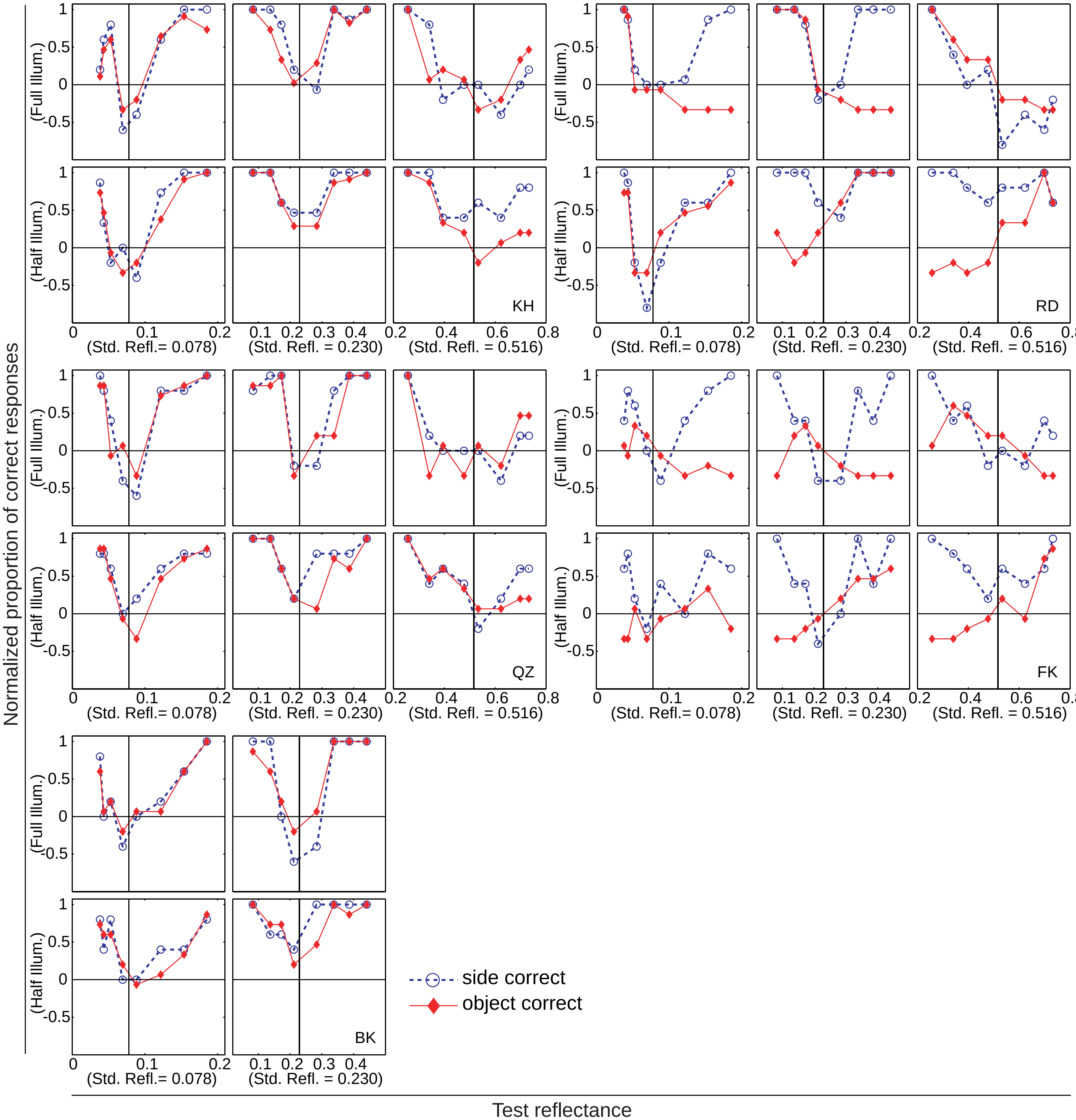

Figure 7. Results of five observers from Experiment 1b. Plots are formatted identically to Figure 6. Data points are connected for grouping purposes only.

contains the test object as well as the other observers. However, within that side, under the described conditions, they consistently chose the wrong object. This demonstrates a failure of lightness constancy that is opposite to the failure predicted by the photometer paradigm illustrated in Figure 5b. It appears that these observers used a simple strategy in which the effect of the illumination difference was overestimated, not underestimated or ignored.

2.6 Statistical tests

Experiment 1a contains a larger number of trials, with finer gradations of reflectance differences than Experiment 1b, and the data obtained are suitable for further statistical testing. Both ST and SS demonstrate some degree of lightness constancy. If lightness constancy were perfect across illuminants, the only limitation for choosing the correct object would be the ability to discriminate differences within illuminants. In other words, identification should be perfect given discrimination.

We tested this hypothesis by analyzing the distribution of incorrect responses. Under this hypothesis, the incorrect object identification responses should be randomly distributed among the three standard objects. In other words, if identification errors are only made when observers cannot discriminate between the objects, observers will choose at random, and the incorrect responses will be evenly distributed among the three wrong answers. The number of *incorrect side* responses should then be twice that of *incorrect object given correct side* responses (a fuller derivation can be found in "Appendix 5.2"). If this hypothesis is incorrect, a larger portion of responses will result in *correct side but incorrect object* responses.

Maximum likelihood estimates were found based on the hypothesis and then compared to the observed data with a chi-squared test (for details, see "Appendix 5.2"). With the given parameters, the critical value of χ^2 at the 0.01 significance level is 16.81. Each χ^2 value from Experiment 1a is shown in Figure 6 under the lightness identification function from which it was obtained. Values greater than the critical value are labeled by asterisks, and signify conditions where the hypothesis that identification is limited only by discrimination can be rejected.

The pattern of χ^2 results in Figure 6 shows a strong trend across the two observers. The above hypothesis is generally rejected in conditions when the test objects under full illumination have lower reflectances than the standards, and when test objects under half illumination have higher reflectances than the standards (10 out of 12 times combining both observers). The hypothesis generally fails to be rejected in conditions when test objects under full illumination have higher reflectances than the standards, and when test objects under half illumination have lower reflectances than the standards (rejected only 1 out of 12 times combining both observers). From the asymmetry among the conditions rejected, the question arises whether a single perceptual strategy can account for the failures of lightness constancy and its successes, depending on stimuli conditions. We address this question in Experiment 2.

3. Experiment 2: Brightness dissimilarity

3.1 Procedure

In the second experiment, we wanted to investigate whether observers in Experiment 1 were using a lightness-based strategy, or a brightness-based strategy. We also wanted to examine the constancy discrepancies between the two groups of observers found in Experiment 1. All physical parameters and procedures were kept the same. The only change was in the task. Instead of instructing the observers to choose the object with the different material, observers were asked to choose the object with the different brightness. The following instructions were used:

> As in the previous experiment, you will be presented with four pieces of gray, crumpled paper. The papers will be placed in a box with two separate compartments. Each compartment will contain two of the papers. Your task is to choose the paper that appears to be MOST DIFFERENT IN BRIGHTNESS. This paper may be the brightest or the least bright. Disregard the material of the paper and disregard the illumination differences. More than one paper may look different. Simply choose the one that looks most different.

It is important to note that the difference in instruction between Experiments 1 and 2 were made clear to the observers. Several practice demonstrations were run before trials began to ensure observers understood the task. Both observers from Experiment 1a were run in Experiment 2a with the same number of conditions using the same sets of standard and test objects. Four of the five observers from Experiment 1b were run in Experiment 2b, including the two observers who displayed poor lightness constancy, again with the same number of conditions using the same sets of standard and test objects. As in Experiments 1a and 1b, the only difference between Experiments 2a and 2b were the number of, and reflectance values of the test objects. Due to the constraint imposed by length of time needed to collect data, and the desire to use as many observers as possible, for three of the four observers in Experiment 2b, the high reflectance standard set was omitted.

3.2 Results

Results for the two observers in Experiment 2a are plotted in Figure 8. Normalized proportions of responses where the test object was chosen as most different in brightness are plotted versus test reflectance. As in Experiment 1, proportions of correct side (open circles) give psychometric brightness discrimination functions. Brightness dissimilarity functions across illuminants are given by the proportion of times the test object was chosen as the most different in brightness (filled diamonds). Both sets of data were fit with psychometric functions (for details, see "Appendix 5.1"). Each data point represents the mean of 16 trials. Results for the four observers in Experiment 2b are plotted in Figure 9 in an identical fashion to Figure 8. Because Experiment 2b did not contain enough conditions for reliable function fitting, data points are connected for grouping purposes only. Each data point represents the mean of 10 trials.

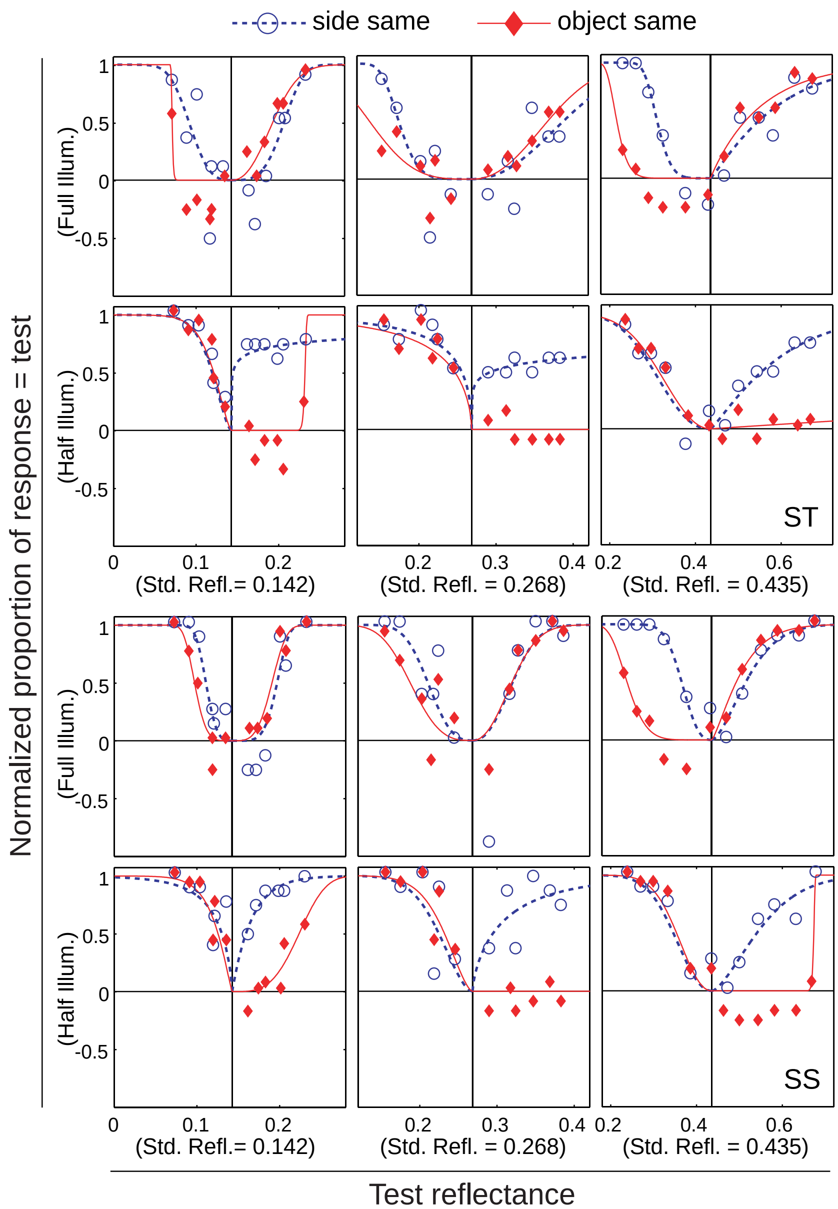

Figure 8. Results of two observers from Experiment 2a. Normalized proportions of responses where the object chosen as most different in brightness was the object of odd reflectance are plotted versus test reflectance. Open circles represent the proportion of responses that the object chosen was on the same side as the test; filled diamonds represent the proportion that the object chosen was the test. For each set of plots, the top row indicates conditions where the test is under full illumination, while the bottom row indicates conditions where the test is under half illumination. The vertical line in each plot indicates the reflectance level of the standard set. Psychometric curves were fit to both sets of data by using maximum likelihood ratios.

All observers in Experiment 2 gave brightness discrimination results similar to each other and similar to their brightness discrimination results in Experiment 1, as seen by comparing the sets of open circles in Figure 6 through Figure 9. This result is not surprising because the first step in either experiment's 2 x 2 AFC task is the same. No matter whether observers were asked to identify the object most different in lightness, or the object most different in brightness, they had to first choose the compartment that contained the two objects different from each other in both attributes under a single illuminant.

In terms of brightness dissimilarity performance, all observers in Experiment 2 gave similar results to each other. In addition, the two observers in Experiment 2a, and the two observers in Experiment 2b that demonstrated lightness constancy in Experiment 1b (KH and BK), gave brightness dissimilarity results similar to their lightness identification results in Experiment 1. Comparison between the lightness identification psychometric functions in Experiment 1a (solid line fits in Figure 6) and brightness dissimilarity psychometric functions in Experiment 2a (solid line fits in Figure 8) demonstrates the similarity be-

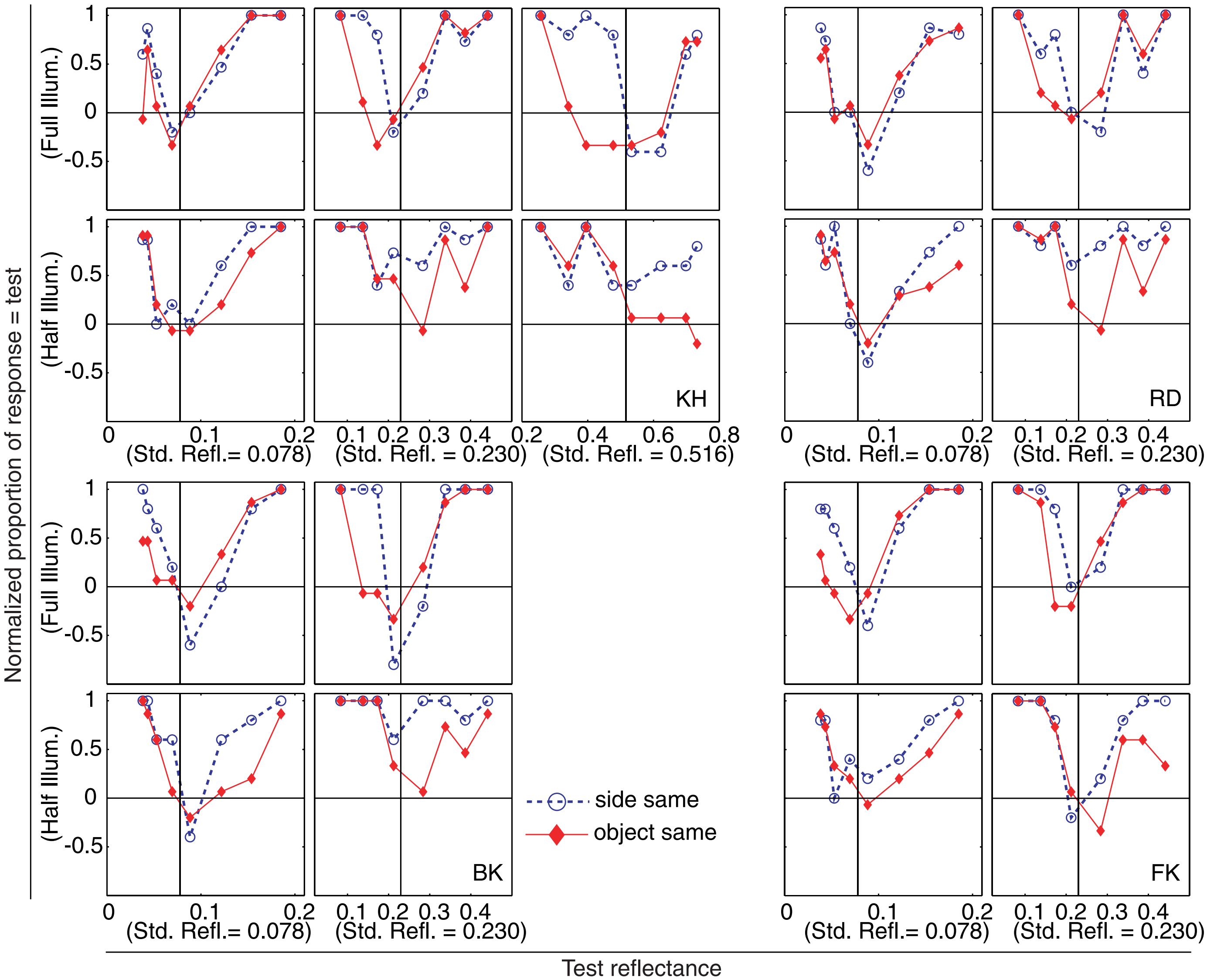

Figure 9. Results of four observers from Experiment 2b. Plots are formatted identically to Figure 8. Data points are connected for grouping purposes only.

tween the results of these two different tasks. Most importantly, the asymmetries seen in the lightness identification curves in Experiment 1a are similar to the asymmetries in the brightness dissimilarity curves in Experiment 2a.

As a quantitative comparison between equivalent condition responses in Experiments 1a and 2a, reflectance thresholds are plotted in Figure 10. Values were calculated from the psychometric functions fit to observers' responses in Figure 6 and Figure 8. Figure 10a shows the brightness discrimination thresholds of Experiment 1 versus Experiment 2. Figure 10b shows the lightness identification thresholds of Experiment 1 versus the brightness dissimilarity thresholds of Experiment 2. Within each subplot, data points are divided into conditions in which the test object was under full or half illumination. In Figure 6 and Figure 8, the psychometric curves do not fit the data points that fall below chance, but this does not affect the estimates of the threshold. The overall fits of the psychometric curves were tested using maximum likelihood chi-square tests, and they only failed in a few instances under the half illumination conditions. As a result, Figure 10b is missing two data points for ST and one for SS.

Figure 10 shows that threshold values are similar between equivalent conditions in the two experiments. This is expected for brightness discrimination responses because observers are most likely using the same strategy, but it is also true for the lightness identification and brightness dissimilarity thresholds. Across the two observers, in the 21 conditions where identification and dissimilarity curves could be fit, the lightness identification task had lower thresholds in 12 cases, while the odd-brightness task had lower thresholds in 9 cases. The fact that observers gave similar responses when identifying lightness differences in Experiment 1 as when identifying the object most different in brightness in Experiment 2 indicates that they could

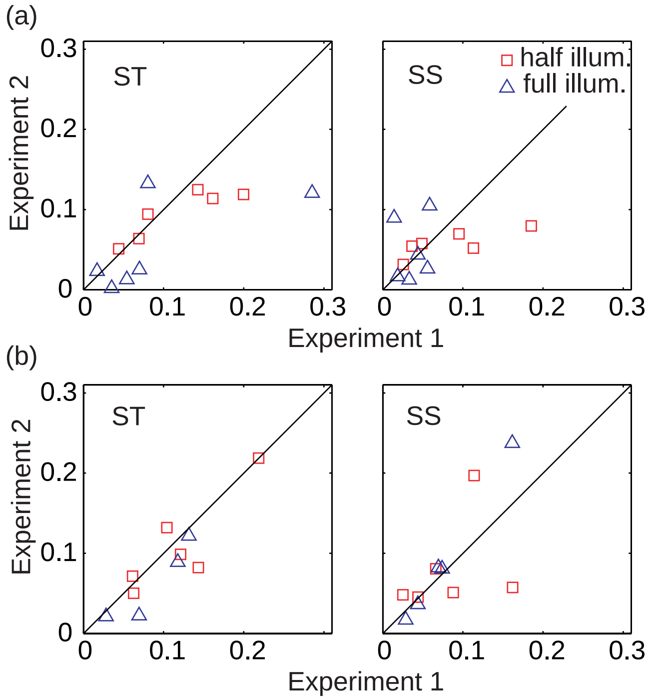

Figure 10. Reflectance threshold values for (a) brightness discrimination in Experiment 1 versus Experiment 2, and (b) lightness identification in Experiment 1 versus brightness dissimilarity in Experiment 2. Squares indicate conditions in which test objects were under full illumination, while triangles indicate conditions under half illumination.

have used brightness dissimilarity as the cue for reflectance identification in Experiment 1.

In contrast, the group of observers who displayed a consistent breakdown of lightness constancy in Experiment 1b (RD and FK) gave a different pattern of results for their brightness dissimilarity responses in Experiment 2b than for their lightness identification responses in Experiment 1b. This indicates that in the first experiment they did not use brightness dissimilarity as the cue for reflectance identification. The brightness dissimilarity responses of these two observers in Experiment 2b were, however, similar to both the lightness identification and brightness dissimilarity results of the other group of observers. This suggests that if these two observers had used a brightness dissimilarity strategy in Experiment 1, their lightness identification performance would have been comparable to that of the other observers.

4. Discussion

Most recent studies of lightness perception use computer-generated stimuli. With the advent of computer-simulated displays, many complex scenes can be easily rendered and parameters can be altered with fine precision and high accuracy. There are, however, important advantages to using natural materials (Koenderink, 1999). Natural stimuli require much more effort to obtain consistency in properties such as reflectance and texture. Natural stimuli also require much more effort in terms of presentation, because after each trial new stimuli must be manually adjusted or replaced. However, the advantage of binocularly viewed, complex, 3D real stimuli is that they provide many cues missing from computer simulations, and even from flat natural stimuli. The crumpled materials used in this study exhibit physical phenomena such as shadows, highlights, and inter-reflections between points on surfaces. All of these cues can help in extracting reflectance (Montag & Berns, 2000; Nayar & Oren, 1995; Nishida & Shinya, 1998) and estimating illuminants (Maloney, 2002). Shape cues, such as occlusion, object curvature, and perspective, give information regarding illumination direction that also aid lightness perception (Adelson & Pentland, 1996; Pessoa et al., 1996; Sun & Perona, 1996). These naturalistic cues can potentially contribute to the correct classification of reflection versus illumination variations across the scene, classifications for which 2D interpretations such as the Retinex model often fail (Land & McCann, 1971).

The methods used in this study provide performance-based measurements of lightness constancy. In Experiment 1, by presenting four materials in a 2 x 2 AFC manner, lightness identification and brightness discrimination were simultaneously measured from the same trials. Given that observers in the current study had the cues described above to draw from, if they were capable of perfect lightness constancy, they should have successfully identified identical reflectances across illuminants whenever they successfully discriminated within illuminants.

Previous studies have attempted to measure lightness constancy by separating appearance judgments between reflectance matching and intensity matching on separate trials (Arend & Goldstein, 1987; Arend, Reeves, Schirillo, & Goldstein, 1991; Arend & Reeves, 1986). These studies have shown fairly accurate constancy in paper-matching tasks where observers were matching reflectance, whereas brightness matches varied substantially as a function of illumination. All of these experiments, however, used flat Lambertian displays (simulated or real). In flat displays, contrast of the test with respect to the background is immune to full-field luminance changes, and could be used as a cue to match reflectance without extracting lightness. Most objects in the real world, however, are neither flat nor perfectly Lambertian. Contrast with respect to the background may not be an invariant for 3D situations involving different slants with respect to the observer. In Experiment 2 of our study, observers used an explicit brightness-based strategy and did not need to estimate lightness. The similarities in the results of Experiments 1 and 2 indicate that the majority of observers could have used brightness differences as a cue and not estimated lightness in the lightness identification task of Experiment 1.

Khang, Koenderink, and Kappers (2003) have used a variant of our method to measure veridical perception of the reflectance of rotating dodecahedra presented under collimated, hemispherical-diffuse, and ambient illumination on a CRT screen. Pairs of polyhedra were simulated under the same illuminant. On each trial, observers had to choose the polyhedron that had reflectances of six if its faces randomly perturbed. Under ambient lighting there is no impression of 3D shape, and the task reduces to brightness discrimination. Thresholds for discriminating nonuniform polyhedra were significantly higher for the collimated and hemispherical-diffuse illuminants than for the ambient case. Their results for regular polyhedra thus corroborate ours for crumpled objects.

The most common method used in the past to express the degree of approximation to veridical perception has been "constancy ratios" derived from appearance matches (Thouless, 1931; Woodworth, 1938). When matching standard surface materials across illuminants, reflectance match settings usually lie between two extremes, one conforming to the standard reflectance, and the other conforming to the reflectance that would give equal luminous intensity under the match illuminant. Because perceived brightness, unlike luminance, is affected by the surrounding conditions, these ratios do not address the question of where the matches lie between brightness and lightness settings.

The bias in the identification and dissimilarity curves (i.e., shifted to the left in the full illumination conditions, and shifted to the right in the reduced illumination curves) is present both in Experiments 1a (Figure 6, lightness identification functions) and 2a (Figure 8, brightness dissimilarity functions). The lightness identification curves of the photometer-like observer discussed earlier were based on judging luminance dissimilarities. They have similar biases, but do not resemble the curves from Experiments 1 or 2. However, it is well known that light adaptation affects brightness discrimination and appearance (Craik, 1938; Helson, 1964). Therefore, we tested whether a photometer-like observer could give results similar to our observers if we incorporated a mechanism of adaptation, where the brightness value of a stimulus is equal to the mean luminance of the stimulus multiplied by a scalar gain whose value is a monotonically decreasing function of mean luminance of the compartment (Hayhoe, Benimoff, & Hood, 1987; Zaidi, Shapiro, & Hood, 1992). Because the backgrounds of the compartments have equal mean reflectances, their luminance values can be related to the amount of illumination they receive. According to this model (Figure 11), the gain, g, is equal to 1.0 when the compartment's mean luminance, L, is equal to 0.0 and declines monotonically as illumination increases. The rate of decline is governed by the free-parameter, κ [i.e., $g = \kappa/(\kappa+L)$].

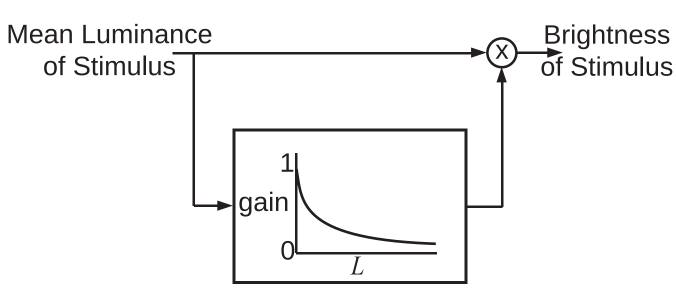

Figure 11. Model of adaptation where the luminance of a stimulus is scaled by an adapting mechanism to give perceived brightness. The gain of adaptation is controlled by the mean luminance of the compartment, L, and a free parameter, κ.

Figure 12 illustrates hypothetical brightness dissimilarity responses based on this model for three different values of κ. Normalized proportions of correct responses are plotted versus reflectances of test objects. Standard reflectances are marked by x. If there were no adaptation (i.e., $\kappa = \infty$) (Figure 12a), responses would be equivalent to the photometer paradigm illustrated in Figure 5b, and judgments would be based solely on luminance values of the objects. Systematically incorrect responses will cross over to correct responses at points where the reflectance difference between test and standard is great enough to overcome the illumination difference. At these points, the absolute luminance difference between standards across the two compartments is equal to the absolute luminance difference between test and standard across compartments. With an illumination difference of two, and $\kappa = \infty$, these crossover points are reached when a test under full illumination is lowered to a reflectance of 0, or when a test under reduced illumination is raised to a reflectance of $3x$. As κ decreases (Figures 12b and 12c), adaptation increases, and the crossover points move closer to the standard reflectance. The model's brightness dissimilarity responses now start to approximate the reflectance identification responses of the observers: in particular, the asymmetry of the lightness identification curves compared to the brightness discrimination curves. Note also that the curves in Figure 12 predict points systematically below chance levels for conditions where test objects under full illumination have lower reflec-

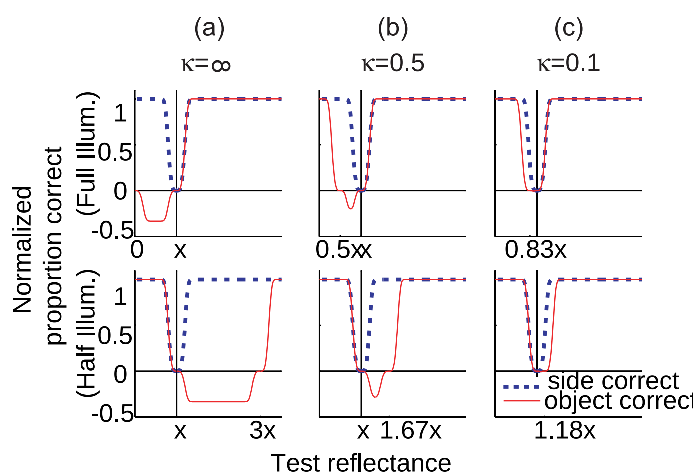

Figure 12. Normalized proportion of correct responses based on a model of brightness dissimilarity plus adaptation. (a), (b), and (c) represent models based on three levels of gain with decreasing κ values. Standard reflectances are denoted by x.

tances than the standards, and where test objects under half illumination have higher reflectances than the standards. These predictions based on brightness dissimilarity judgments are confirmed in the asymmetry of the points below chance for lightness identification data in Figure 6.

Many models of lightness and color constancy assume that the visual system estimates the scene illuminant and uses this estimate to determine material reflectance (Beck, 1961; Katz, 1935; Koffka, 1935; Kozaki & Noguchi, 1976; Logvinenko & Menshikova, 1994; Noguchi & Kozaki, 1985; Woodworth, 1938). In terms of this illumination-estimation hypothesis, the questions arising about this study are, what information is available in the display, and what is the best any visual system could estimate? Backgrounds in the two boxes have similar statistics. Theoretically, a system could calculate the mean luminance of both backgrounds, then measure the mean luminance of each of the four cups, and take the ratios. According to this design, there is sufficient information to perform the lightness identification task perfectly (Figure 5a), but it is clear that observers are unable to do this. Observers can tell which side has the brighter illuminant, but either they are unable to calculate the ratio exactly or else they cannot put it to use. In Figure 13, results are shown for a hypothetical observer who does not use brightness dissimilarity of objects, but rather estimates the illumination ratio between the two compartments and consciously factors it into the luminance of the objects to obtain their reflectance. The brightness discrimination functions are not affected by this estimation. Assuming reflectance can be perfectly factored out of luminance when the correct illumination ratio (2:1) has been estimated, lightness identification functions will overlie brightness discrimination functions. Depending on whether the illumination ratio is overestimated or underestimated, asymmetries in opposite directions will result. The two observers who did not use brightness differences as the primary cue (RD and FK) were most likely using a strategy in Experiment 1 where the illumination difference between the two compartments was consistently over-estimated (i.e., the right two columns in Figure 13).

As more investigators look at lightness of 3D objects, it is becoming clear that some observers try to do conscious corrections that lead to individual differences. For example, Ripamonti et al. (2004) find ranges of individual differences when people try to match lightness across different slants. These individual differences have been modeled in terms of different estimates for the ambient illumination (Bloj et al., 2004; Boyaci, Doerschner, & Maloney, 2004; Boyaci, Maloney, & Hersh, 2003; Doerschner, Boyaci, & Maloney, 2004). Individual differences thus are likely to be due to attempts to infer a non-sensory quality, rather than due to the particular task or instruction.

The illumination-estimation hypothesis has also been challenged by the results of Rutherford and Brainard (2002), who used a two-stage appearance-matching paradigm. In their study, observers first adjusted the illumination in one chamber to match that in another chamber with different mean background reflectance. Then, with the illumination matches fixed, observers adjusted the perceived reflectance of a test patch in one chamber to match the surface reflectance of a patch in the other. After both matches, neither the illumination nor the reflectance of the test patches was matched correctly. Furthermore, the luminance of the light reflected from the two test patches was significantly different.

Under everyday conditions, observers consistently judge surfaces as having a certain lightness or grayness. This subjective impression points out the tendency to use reflectance in mental representations of surfaces. This phenomenological experience, however, is not sufficient evidence that the visual system has access to the reflectance or lightness of materials. To estimate relatively veridical reflectance, a scene would have to contain sufficient information

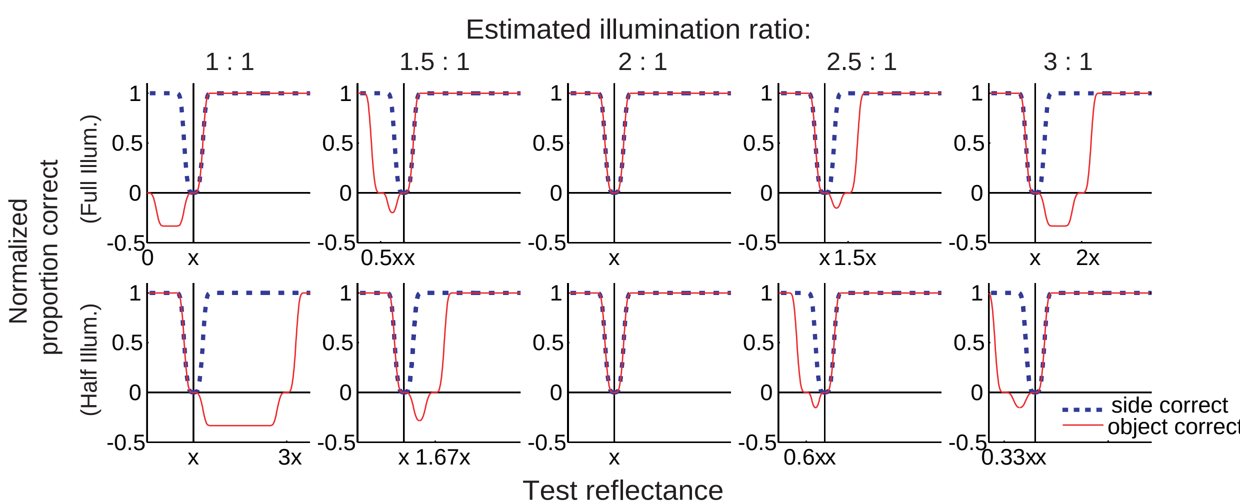

Figure 13. Normalized proportion of correct responses based on a model where the illumination ratio between the two compartments is estimated and used to factor out object reflectance. The format of the plots is identical to Figure 12.

to perform this task, and a visual system would have to have the capability of utilizing this information. In our study, we have tried to maximize the information available to observers by providing a rich background and objects that are 3D, containing facets and highlights. Our results show that despite naturalistic binocular viewing of these information-rich situations, under many conditions, observers cannot identify objects of the same lightness across different illumination levels. Further, observers seem to use brightness-based strategies to try to identify lightness. This conclusion is based on the fact that the psychometric curves for lightness identification (and brightness dissimilarity) are systematically asymmetric as compared to brightness discrimination curves measured simultaneously for the same objects, and this asymmetry can be predicted quantitatively from observers' choices of most dissimilar brightnesses, and qualitatively from the choices of a photometer-like model observer incorporating brightness adaptation. Note that reflectance dissimilarities are not asymmetric around the standard reflectance, so lightness percepts could not be the basis of the asymmetric brightness dissimilarity judgments in Experiment 2.

To summarize, we show some conditions where lightness identification is limited solely by the limen of brightness discrimination, and other conditions where lightness identification is considerably worse. More importantly, we show that the same relative brightness-based strategy reproduces both sets of results. The visual system may have evolved to identify object properties, but this identification can only be based on sensory information. Mean reflectance of a surface is a physical quality, just like the spectral reflectance of a light. In color matching, an observer does not match spectra, but rather the outputs of cones. We wanted to find out the proximal quality that is used in lightness identification of surfaces. We suggest that, for our 3D objects, this quality is perceived brightness, because (1) the results are similar if observers are explicitly asked to identify brightness dissimilarities, and (2) a model of brightness as multiplicative adaptation applied to luminance, generates curves that reproduce the asymmetries in the lightness identification data and the points below chance.

5. Appendix

As stated earlier, the observer's responses simultaneously provide percentage correct for a pair of 2AFC tasks [i.e., (1) choosing the side that contains the dissimilar object, and (2) then choosing one of the objects on that side that is different from the two on the other side]. As such, it is a variant of the classical 2 x 2 detection-discrimination task and can be statistically analyzed by modifications of well-established procedures (Khang & Zaidi, 2002; Sachtler & Zaidi, 1995; Watson, 1979; Watson & Robson, 1981).

5.1 Psychometric function fitting

Here, we fit psychometric curves to method of constant stimuli data. These fits allow for the visualization of trends in data, and provide estimates of reflectance thresholds for brightness discrimination, lightness identification, and brightness dissimilarity. The following notation is adopted:

- Δ – difference between standard and test reflectance
- δ – index for Δ
- d_δ – probability of detecting a difference at level Δ
- γ_δ – probability of guessing correctly if detecting a difference does not occur at level Δ
- r_δ – probability of correct responses at level Δ
- n_δ – # of trials at level Δ
- m_δ – # of correct responses at level Δ
- α – magnitude parameter of psychometric function
- β – steepness parameter of psychometric function

Discrimination responses were judged as correct when either object in the compartment that contained the test was chosen, whereas correct identification required that the test itself be chosen. Both sets of data can be fit with a psychometric function (Equation 1) modified from the function proposed by Quick (1974), which is different from the Weibull (1951) distribution only in using 2 instead of *e* as the base of the exponentiation:

$$d_\delta = 1 - 2^{-(\Delta/\alpha)^\beta}, \quad (1)$$

where d_δ is the probability of detecting a difference at level Δ. Depending on the set of data being analyzed, this can be either the probability of discriminating the correct side, or the probability of identifying the correct object. α and β, respectively, represent the function's magnitude and steepness parameters. Because the probability of correct response, r_δ, equals the probability of detecting a difference, d_δ, plus the probability of guessing correctly when detection does not occur, γ, Equation 1 can be substituted into Equation 2 to obtain Equation 3:

$$r_\delta = d_\delta + (1 - d_\delta)\gamma \quad (2)$$

$$r_\delta = 1 - (1-\gamma)2^{-(\Delta/\alpha)^\beta} \quad . \quad (3)$$

Equation 3 is normalized for guessing by γ equal to 0.5 for brightness discrimination and 0.25 for lightness identification. This function can now be fit to the data by maximizing overall likelihoods based on the binomial probability density function:

$$L_\delta = \frac{n_\delta!}{m_\delta!(n_\delta - m_\delta)!} r_\delta^{m_\delta} (1 - r_\delta)^{n_\delta - m_\delta} \quad . \quad (4)$$

where m_δ is the number of correct responses and n_δ is the number of trials for a given reflectance delta. Because each psychometric function is fit from trials over six different Δ

levels, the overall likelihood is the product of each Δ level's individual likelihood:

$$L = \prod_{\delta} L_{\delta} \quad . \tag{5}$$

For each of the pairs of discrimination and identification data, the maximum likelihood estimates of α and β were obtained that generated r_{δ} values that when substituted into Equation 4, maximized Equation 5. These fittings were used to construct psychometric curves plotted with their respective response data in Figure 6 and Figure 8. In addition, for each data set, the α value obtained corresponds to the threshold reflectance defined as the reflectance differences between standard and test that gives a correct response rate at the midpoint between chance and 100% correct. These threshold reflectance values from Experiments 1a and 2a are plotted versus each other in Figure 10.

5.2 Hypothesis testing

Ability to discriminate brightness differences versus ability to identify lightness differences can also be studied by analyzing the proportion of responses in the three possible response categories. The additional notation is adopted:

- k – index for response category:
- k = 1 – side correct, object correct
- k = 2 – side correct, object incorrect
- k = 3 – side incorrect, object incorrect
- δ – probability of discriminating correct side
- γ_{side} – probability of guessing correct side when discrimination does not occur = 0.5
- γ_{object} – probability of guessing correct object when identification does not occur, given correct side = 0.5
- J – # of Δ in each data set

The likelihood of getting $m_{\delta k}$ responses in each category is given by the multinomial probability distribution for each Δ:

$$L_{\delta} = \frac{n_{\delta}!}{\prod_{k} m_{\delta k}!} \prod_{k} \pi_{\delta k}^{m_{\delta k}} \quad . \tag{6}$$

The likelihood of the whole data set of standard-test comparisons is then:

$$L = \prod_{\delta} L_{\delta} \quad . \tag{7}$$

The simplest hypothesis about the distribution of $r_{\delta k}$ is H_o: assume the best estimate for each $r_{\delta k}$ is $m_{\delta k}/n_{\delta}$. In other words, the probability of each response is simply given by the ratio of responses from the observed data. The total number of parameters for this hypothesis is $2J$, because for each Δ level, once the number of responses has been determined for two of the categories, the number of responses in the third category must be fixed to make the total sum equal to n_{δ}, i.e.,

$$\sum_{\delta k} m_{\delta k} = n_{\delta} \, . \tag{8}$$

The hypothesis we want to test against is H_1: *Probability of lightness identification given brightness discrimination = 1.0.* This assumes that once Δ is large enough to discriminate two objects as different within a single illumination, observers will always be able to identify the odd object across illuminants. Possible routes to each response category are given in Figure 14.

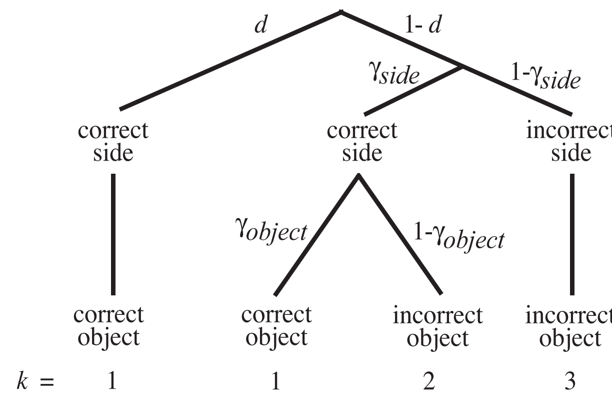

Figure 14. Routes to obtain one of the three response categories: *[side correct, object correct]*, *[side correct, object incorrect]*, or *[side incorrect, object incorrect]*. The decision tree assumes that brightness discrimination is the limitation for lightness identification, and that if the correct side can be discriminated, the correct object will always be identified. δ is the probability of discriminating the correct side, γ_{side} is the probability of guessing the correct side when discrimination does not occur, and γ_{object} is the probability of guessing the correct object when identification does not occur, given correct side has been chosen.

An assumption made in both hypotheses is, because the position of the test was randomized on each trial, if Δ is below the appropriate discrimination threshold, γ_{side} = 1-γ_{side} = 0.5 (i.e., no guessing biases). Another assumption is that if the correct side cannot be discriminated, the correct object cannot be identified, therefore γ_{object} = 1-γ_{object} = 0.5.

According to H_1, if lightness identification responses across illuminants are limited only by the ability to discriminate brightness differences within illuminants, the incorrect object identification responses should be randomly distributed among the three standard objects. For a given reflectance delta, if it is not possible to discriminate the correct side, the probability of guessing the incorrect object on the correct side is 0.25(1-d), while the probability of guessing the incorrect side (and hence the incorrect object) is 0.5(1-d). Under these assumptions, the number of incorrect side responses should be twice that of correct side/incorrect object responses. From Figure 14, for a given

Δ, the probabilities of the three response categories can be written as:

$$r_{\delta 1} + r_{\delta 2} + r_{\delta 3} = 1.0 \tag{9}$$

$$r_{\delta 1} = d + (1\text{-}d)\gamma_{side}\,\gamma_{object} = d + (1\text{-}d)0.25 = 1\text{-}3r_{\delta 2} \tag{10}$$

$$r_{\delta 2} = (1\text{-}d)\gamma_{side}\,(1\text{-}\gamma_{object}) = (1\text{-}d)0.25 = r_{\delta 2} \tag{11}$$

$$r_{\delta 3} = (1\text{-}d)(1\text{-}\gamma_{side}) = (1\text{-}d)0.5 = 2r_{\delta 2} \tag{12}$$

Equation 6 now becomes a trinomial probability distribution function with the three response probabilities defined by $r_{\delta 2}$:

$$L_\delta = \frac{n_\delta\,!}{\prod_k m_{\delta k}\,!}\left(1-3r_{\delta 2}\right)^{m_{\delta 1}} r_{\delta 2}^{m_{\delta 2}} \left(2r_{\delta 2}\right)^{m_{\delta 3}} \tag{13}$$

Maximum likelihood estimates were found by obtaining values for $r_{\delta 2}$ that when substituted into Equation 13 maximized Equation 7. The number of parameters under this hypothesis is J, because for each Δ level, the only free parameter is the probability of discrimination, d. To test whether H_1 fits as well as H_0, we used the statistic:

$$\lambda = -2\ln\left(L_1/L_0\right), \tag{14}$$

where L_1 and L_0 are the maximum likelihoods under H_1 and H_0, respectively. In a theorem of Wilks (1938), λ has been shown to be asymptotically distributed as χ^2 with degrees of freedom equal to the difference in the number of free parameters between the two hypotheses (i.e., $2J\text{-}J = J = 6$) (Hoel, Port, & Stone, 1971). With degrees of freedom equal to 6, the critical value of χ^2 at the 0.01 significance level is 16.81. If λ exceeds the critical value, we conclude that the likelihood based on the model is significantly worse than the likelihood based on the observed data. This is equivalent to stating that discrimination is not the only limitation of lightness identification. λ values from Experiment 1a are listed as equivalent χ^2 values in Figure 5 under each lightness identification function. Values that exceed the critical value are labeled by and asterisk.

Acknowledgments

We thank Fuzz Griffiths, Andrea Li, Hal Sedgwick, William Swanson, Hannah Smithson, and Byung-Geun Khang for their helpful comments, and John Robilotto for help with constructing the experimental apparatus. Portions of this work were presented at the 2000 annual meeting of the Association for Research in Vision and Ophthalmology. This work was supported by a T-35 grant to RR and National Eye Institute Grant EY07556 to QZ.

Commercial relationships: none.
Corresponding author: Rocco Robilotto.
Email: rrobilotto@sunyopt.edu.

References

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339-351). Cambridge, MA: MIT Press.

Adelson, E. H., & Pentland, A. P. (1996). The perception of shading and reflectance. In D. Knill & W. Richards (Eds.), *Perception as* Baysian *inference* (pp. 409-423). New York: Cambridge University Press.

Arend, L. E., & Goldstein, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America A, 4*(12), 2281-2285. [PubMed]

Arend, L. E., Jr., Reeves, A., Schirillo, J., & Goldstein, R. (1991). Simultaneous color constancy: Paper with diverse Munsell values. *Journal of the Optical Society of America A, 8*(4), 661-672. [PubMed]

Arend, L. E., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A, 3*(10), 1743-1751. [PubMed]

Arend, L. E., & Spehar, B. (1993a). Lightness, brightness, and brightness contrast: 1. Illuminance variation. *Perception & Psychophysics, 54*(4), 446-456. [PubMed]

Arend, L. E., & Spehar, B. (1993b). Lightness, brightness, and brightness contrast: 2. Reflectance variation. *Perception & Psychophysics, 54*(4), 457-468. [PubMed]

Beck, J. (1961). Judgments of surface illumination and lightness. *Journal of Experimental Psychology, 61*, 368-375. [PubMed]

Bloj, M., Ripamonti, C., Mitha, K., Greenwald, S., Hauck, R., & Brainard, D. H. (2004). An equivalent illuminant model for the effect of surface slant on perceived lightness. *Journal of Vision*, 4(9), 747-763. http://journalofvision.org/4/9/6/, doi:10.1167/4.9.6. [PubMed][Article]

Bloj, M. G., & Hurlbert, A. C. (2002). An empirical study of the traditional Mach card effect. *Perception, 31*(2), 233-246. [PubMed]

Boyaci, H., Doerschner, K., & Maloney, L. T. (2004). Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity. *Journal of Vision,* 4(9), 664-679, http://journalofvision.org/4/9/1/, doi:10.1167/4.9.1. [PubMed][Article]

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision,* 3(8), 541-553, http://journalofvision.org/3/8/2/, doi:10.1167/3.8.2. [PubMed][Article]

Bruno, N. (1994). Failures of lightness constancy, edge integration, and local edge enhancement. *Vision Research, 34*(17), 2205-2214. [PubMed]

Chevreul, M. E. (1839). *De la loi du contraste simultane des couleurs*. Paris: Pitois-Levreault.

Craik, K. J. W. (1938). The effect of adaptation on differential brightness discrimination. *Journal of Physiology, 92*, 406-421.

Doerschner, K., Boyaci, H., & Maloney, L. T. (2004). Human observers compensate for secondary illumination originating in nearby chromatic surfaces. *Journal of Vision, 4*(2), 92-105, http://journalofvision.org/4/2/3/, doi:10.1167/4.2.3. [PubMed][Article]

Evans, R. M. (1974). *The perception of color*. New York: Wiley & Sons.

Gilchrist, A. L. (1980). When does perceived lightness depend on perceived spatial arrangement? *Perception & Psychophysics, 28*(6), 527-538. [PubMed]

Gilchrist, A. L. (1988). Lightness contrast and failures of constancy: A common explanation. *Perception & Psychophysics, 43*(5), 415-424. [PubMed]

Gilchrist, A. L., Delman, S., & Jacobsen, A. (1983). The classification and integration of edges as critical to the perception of reflectance and illumination. *Perception & Psychophysics, 33*(5), 425-436. [PubMed]

Gilchrist, A. L., & Jacobsen, A. (1984). Perception of lightness and illumination in a world of one reflectance. *Perception, 13*(1), 5-19. [PubMed]

Hayhoe, M. M., Benimoff, N. I., & Hood, D. C. (1987). The time-course of multiplicative and subtractive adaptation process. *Vision Research, 27*(11), 1981-1996. [PubMed]

Helmholtz, H. v. (1962). *Physiological Optics* (J. P. C. Southhall, Trans. Vol. 3). New York: Dover.

Helson, H. (1964). Adaptation-*level theory*. New York: Harper and Row.

Hering, E. (1964). *Outlines of a theory of the light sense* (L. M. Hurvich & D. Jameson, Trans.). Cambridge, MA: Harvard University Press.

Hoel, P., Port, S., & Stone, C. (1971). *Introduction to statistical theory*. Boston: Houghton Mifflin.

Kardos, L. (1934). *Ding und schatten*. Leipzig: Barth.

Katz, D. (1935). *The world of color* (R. B. MacLeod & C. W. Fox, Trans.). London: Kegan Paul, Trench, Truber & Co.

Khang, B. G., Koenderink, J. J., & Kappers, A. M. (2003). Perception of surface reflectance of 3-D geometrical shapes: Influence of the lighting mode. *Perception, 32*(11), 1311-1324. [PubMed]

Khang, B.-G., & Zaidi, Q. (2002). Cues and strategies for color constancy: Perceptual scission, image junctions and transformational color matching. *Vision Research, 42*(2), 211-226. [PubMed]

Koenderink, J. J. (1999). Virtual psychophysics. *Perception, 28*(6), 669-674. [PubMed]

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace, and Company.

Kozaki, A., & Noguchi, K. (1976). The relationship between perceived surface lightness and perceived illumination: A manifestation of perceptual scission. *Psychological Research, 39*(1), 1-16. [PubMed]

Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. Journal *of the Optical Society of America, 61*(1), 1-11. [PubMed]

Logvinenko, A., & Menshikova, G. (1994). Trade-off between achromatic colour and perceived illumination as revealed by the use of pseudoscopic inversion of apparent depth. *Perception, 23*(9), 1007-1023. [PubMed]

MacEvoy, S. P., & Paradiso, M. A. (2001). Lightness constancy in primary visual cortex. *Proceedings of the National Academy of* Science *U.S.A., 98*(15), 8827-8831, doi:10.1073/pnas.161280398. [PubMed] [Article]

MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Maloney, L. T. (2002). Illuminant estimation as cue combination. *Journal* of *Vision, 2*(6), 493-504, http://journalofvision.org/2/6/6/, doi:10.1167/2.6.6. [PubMed] [Article]

Marzynski, G. (1921). Zeitschrift *fur Psychologie, 87*, 45-72. (Described in R. S. Woodworth [1938], *Experimental Psychology*. New York: Henry Holt.)

Montag, E. D., & Berns, R. S. (2000). Lightness Dependencies and the effect of texture on suprathreshold lightness tolerances. Color *Research and Application, 25*(4), 241-249.

Nayar, S. K., & Oren, M. (1995). Visual appearance of matte surfaces. Science, *267*(5201), 1153-1156. [PubMed]

Nishida, S., & Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. *Journal of the* Optical *Society of America A, 15*(12), 2951-2965. [PubMed]

Noguchi, K., & Kozaki, A. (1985). Perceptual scission of surface-lightness and illumination: An examination of the Gelb effect. *Psychological Research, 47*(1), 19-25. [PubMed]

Pessoa, L., Mingolla, E., & Arend, L. E. (1996). The perception of lightness in 3-D curved objects. *Perception & Psychophysics, 58*(8), 1293-1305. [PubMed]

Quick, R. F., Jr. (1974). A vector-magnitude model of contrast detection. Kybernetik, *16*(2), 65-67. [PubMed]

Ripamonti, C., Bloj, M., Mitha, K., Greenwald, S., Hauck, R., Maloney, S. I., (2004). Measurements of the effect of surface slant on perceived lightness. *Journal of Vision*, 4(9), 747-763. http://journalofvision.org/4/9/7/, doi:10.1167/4.9.7. [PubMed][Article]

Rutherford, M. D., & Brainard, D. H. (2002). Lightness constancy: A direct test of the illumination-estimation hypothesis. *Psychological Science, 13*(2), 142-149, doi:10.1111/1467-9280.00426. [PubMed]

Sachtler, W. L., & Zaidi, Q. (1995). Visual processing of motion boundaries. *Vision Research, 35*(6), 807-826. [PubMed]

Schirillo, J., Reeves, A., & Arend, L. (1990). Perceived lightness, but not brightness, of achromatic surfaces depends on perceived depth information. *Perception & Psychophysics, 48*(1), 82-90. [PubMed]

Schirillo, J. A., & Arend, L. E. (1995). Illumination change at a depth edge can reduce lightness constancy. *Perception &* Psychophysics, *57*(2), 225-230. [PubMed]

Sun, J., & Perona, P. (1996). Early computation of shape and reflectance in the visual system. *Nature, 379*(6561), 165-168. [PubMed]

Thouless, R. H. (1931). Phenomenal regression to the "real" object. *British* Journal *of Psychology, 22*, 1-30.

Watson, A. B. (1979). Probability summation over time. *Vision Research, 19*(5), 515-522. [PubMed]

Watson, A. B., & Robson, J. G. (1981). Discrimination at threshold: Labelled detectors in human vision. *Vision Research, 21*(7), 1115-1122. [PubMed]

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics, 18*, 292-297.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics, 9*, 60-62.

Woodworth, R. S. (1938). Experimental *Psychology*. New York: Henry Holt and Company.

Zaidi, Q. (1999). Color and brightness induction: From Mach bands to three-dimensional configurations. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), *Color vision: From genes to perception*. New York: Cambridge University Press.

Zaidi, Q. (2001). Color constancy in a rough world. *Color Research and Application, 26*(S1), S192-S200.

Zaidi, Q., Shapiro, A., & Hood, D. (1992). The effect of adaptation on the differential sensitivity of the S-cone color system. *Vision Research, 32*(7), 1297-1318. [PubMed]