Lightness perception in high dynamic range images: Local and remote luminance effects

Sarah R. Allred — Department of Psychology, Rutgers, The State University of New Jersey, Camden, NJ, USA

Ana Radonjić — Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

Alan L. Gilchrist — Department of Psychology, Rutgers, The State University of New Jersey, Newark, NJ, USA

David H. Brainard — Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

We measured the perceived lightness of target patches embedded in high dynamic range checkerboards. We independently varied the luminance of checks immediately surrounding the test and those remote from it. The data establish context transfer functions (CTFs) that characterize perceptual matches across checkerboard contexts. Several features of the CTFs are broadly consistent with previous research: Matched luminance decreases when overall context luminance decreases; matched luminance increases when overall context luminance increases; manipulating context locations near the target has a greater effect than manipulating locations far from the target patch. The measured CTFs are not well described, however, by changes with context in multiplicative gain alone or by changes in both multiplicative and subtractive adaptation parameters. We were able to fit the data with a three-parameter model of adaptation. This allowed us to characterize the CTFs by specifying the luminances that appeared white, black, and gray (white point, black point, and gray point, respectively). The white and black points depended additively on the local and remote contrasts, but accounting for the gray point required an interaction term. Analysis of this effect suggests that the target patch itself must be included in a description of the visual context.

Keywords: lightness/brightness perception, contrast, high dynamic range display

Citation: Allred, S. R., Radonjić, A., Gilchrist, A. L., & Brainard, D. H. (2012). Lightness perception in high dynamic range images: Local and remote luminance effects. *Journal of Vision, 12*(2):7, 1–16, http://www.journalofvision.org/content/12/2/7, doi:10.1167/12.2.7.

Introduction

The perceived lightness of an object depends on the scene in which it is viewed. Part of this dependence is simple to understand. Across scenes, the illumination impinging on an object can change, and this in turn causes a change in the intensity of light reflected from the object to the observer. However, even when the intensity of the reflected light is held constant, context can still affect perceived lightness. This second class of effect arises because the visual processing of the light reflected from an object depends on the entire retinal image. The classic example of such a context effect is simultaneous contrast, where the lightness of a test patch varies with its surround (see Adelson, 2000; Gilchrist, 2006 for discussion). To understand the perception of object lightness, and more generally the perception of object color, we must understand such contextual effects.

Ultimately, a successful theory should allow us to predict perceived lightness from the left and right eye retinal images. Achieving this goal is challenging because the number of possible retinal image pairs is astronomical. Not all may be studied directly. Thus, it is necessary to specify a subclass of scenes for study and then identify principles that allow prediction of lightness for all scenes within the class, on the basis of a feasible number of measurements. For example, early work considered scenes whose images consisted of a spatially uniform test region presented against a spatially uniform background. Within this class, the question becomes how the luminance of the test and background interact to yield lightness, and for this class of scenes, the luminance ratio between test and background explains much of the variance (Wallach, 1948).

Despite a great deal of research, however, successful theories of lightness have remained elusive even for moderately complex scenes, such as those where all objects are flat and coplanar. Once multiple objects are introduced in the regions surrounding the test, there is disagreement about what scenes to study, the nature of the key empirical phenomena, and the most appropriate

doi: 10.1167/12.2.7 Received August 12, 2011; published February 8, 2012 ISSN 1534-7362 © ARVO

theoretical approach (Gilchrist, 2006). Some investigators have split the uniform background surrounding the test into two parts, so that the test is now bordered by two rather than one spatially uniform region (Gilchrist, 2006). Others have split the background in a different way, using, for example, two annuli, one immediately adjacent to the test and one remote (Hong & Shevell, 2004; Rudd & Zemach, 2004). Still others have studied checkerboard-like patterns, often referred to as Mondrians, in which the question becomes how to predict the lightness of each check given its luminance and that of all the others (e.g., Arend & Spehar, 1993a, 1993b; Blakeslee & McCourt, 2001; Land & McCann, 1971; Schirillo, 1999). For each such choice, the hope is both that it will be possible to discover simplifying regularities and that the regularities will generalize to allow predictions for more complex scene classes (see Brainard & Maloney, 2011).

Here, we follow in the tradition that uses checkerboard scenes as a model system for studying perceived lightness. We report data from experiments that measure how varying the checkerboard context of a test patch affects its perceived lightness. The present experiments extend previous work in two key ways.

First, we incorporated a pervasive feature of natural scenes, namely, that the images of such scenes can contain large (>10,000:1) variations in luminance from one location to another (Heckaman & Fairchild, 2009; Mury, Pont, & Koenderink, 2009; Xiao, DiCarlo, Catrysse, & Wandell, 2002). Although classic studies using spatially uniform surrounds incorporated large luminance changes (Heinemann, 1955), this manipulation has been little explored even for spatial patterns as simple as checkerboards. This is in part because typical CRT displays do not provide high dynamic range. To conduct the present studies, we employed a custom high dynamic range display.

Second, we systematically manipulated the luminance of checks grouped near to and remote from the central target patch, similar in spirit to manipulations of local and remote annuli used in some experiments (Rudd & Zemach, 2004). These manipulations simulate to a limited extent the spatial changes in illumination that occur in natural scenes. We manipulated checkerboard luminance to vary the degree to which lighter checks are segregated from darker checks in the checkerboard. Our stimuli, however, are missing the geometric factors that, in natural scenes, are associated with a strong impression of different fields of illumination. These include corners, occlusion boundaries, penumbrae at cast shadows, edges, and ratio-invariant X-junctions. Thus, the present studies allow us to investigate the extent to which photometric manipulation in the absence of such geometric cues affects perceived lightness. Across the set of experiments, we used a large range of luminances both for the test patches whose lightnesses were being judged and for the overall luminances of the local and remote groups of contextual checks.

The general characterization of lightness can be split into two pieces (Wallach, 1976). First, for a standard context, we need to understand the mapping between target luminance and perceived lightness. Addressing this question requires specifying a scale for lightness and measuring how luminance maps onto this scale for one single choice of context. Adelson (2000) refers to the resultant mapping as the lightness transfer function (LTF). Elsewhere, we describe some initial measurements of such functions for high dynamic range images without photometric segregation (Radonjić, Allred, Gilchrist, & Brainard, 2011). The second part of the characterization is to describe the mapping between lightness as perceived in any test context and lightness perceived in the standard context. Once an LTF is in hand for a single standard context and the mapping between lightness in any other context and the standard context is characterized, it is possible to map the luminance of a test in any context onto the lightness scale.

In this paper, we address the second aspect of Wallach's program by considering how lightness is matched across contexts. We describe our experimental methods, characterize the main empirical findings, and discuss the broad implications of the results for models of lightness.

Methods

We were interested in characterizing the effect of context on perceived lightness. To that end, we asked observers to make lightness judgments for 24 target patches of different luminance embedded in each of nine different checkerboard contexts.

Observers

Observers were 7 adults between the ages of 20 and 35. Observers CH, MG, OT, PK, and WW were paid volunteers who were naive to the purposes of the experiment and had little experience in psychophysical observation. The other observers (SRA and JL) were laboratory members (one is the first author). All observers had normal color vision as assessed by the Ishihara Color Plates and had normal depth perception and normal or corrected-to-normal visual acuity as assessed by the Keystone VS-II vision screener.

Apparatus and task

Observers looked through an aperture into a rectangular enclosure, at the end of which they viewed an achromatic 25 square checkerboard presented on a custom-built high dynamic range (HDR) display (Figure 1). The checkerboard itself subtended 19.4° and each check subtended

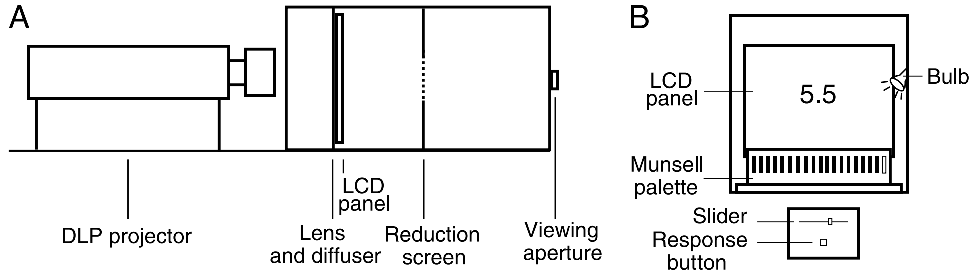

Figure 1. Diagram of apparatus components. (A) Schematic of the HDR display. The DLP projects an image onto the LCD display assembly (consisting of Fresnel lens, diffuser, and the LCD panel itself) through an aperture at the rear of the enclosing box. The box was lined with black cloth to minimize reflection of stray light. The observer viewed the resulting image through a reduction screen and viewing aperture at the other end of the enclosing box. The dotted portion of the reduction screen diagram shows the vertical extent of the square aperture in that screen. Dimensions of the HDR display are provided in the text. (B) Front view of the matching chamber. The chamber was constructed from plywood and painted a matte gray. Inside, the chamber was 40 cm wide, 40.5 cm high, and 40.5 cm deep. The matching palette was located at the bottom of the chamber, 22 cm from the edge of the chamber closest to the observer. The chamber was illuminated by a halogen bulb mounted 27 cm above the bottom of the chamber, on the edge of the chamber to the observers' right. Observers indicated their response using a slider on a custom response box (shown below chamber in diagram). This varied the number on an LCD panel mounted at the back of the viewing chamber. Out-of-range responses were displayed as text on the same monitor. Observers indicated selection of the desired response by pressing a button on the same box.

3.9°. Observers were asked to judge the lightness of the center check in the checkerboard context (the target patch). To do so, observers looked into a separately illuminated matching booth immediately to their left (Figure 1). This matching chamber (positioned at 89-cm height) contained a palette of Munsell papers mounted on white matte cardboard (reflectance = 0.84) approximately 61 cm from the observer. Observers were instructed to choose the Munsell paper that was most similar in lightness to the target patch. Each palette paper was 1.1 cm horizontal by 3.0 cm vertical (1.0° by 2.8° of visual angle). These papers were matte and ranged from Munsell 2.0 to 9.5 in value steps of 0.5. Under the halogen illumination, as measured with a PhotoResearch PR-650 spectral radiometer, the CIE *xy* chromaticity of the light reflected from Munsell palette papers was in the range x = [0.442–0.447] and y = [0.409–0.411].

We measured the reflectance of each palette paper. This was accomplished as follows. First, we used the PR-650 spectral radiometer to measure the luminance of the light reflected from each paper. Denote the reflected luminance from the ith paper by l_p^i. We then measured the luminance of the white palette background directly adjacent to each paper. The background measures, which were at regularly spaced locations, were smoothed with a fourth-order polynomial fit to the measured luminances. Denote the estimated background luminance adjacent to the ith paper by l_b^i. Finally, we estimated the reflectance of the background by measuring the luminance of a white reflectance standard (LabSphere Teflon reflectance standard). This was done at a subset of the locations where we had measurements of the background luminance. Denote the luminance of the standard by l_s^i. Given these measures, we estimated the reflectance of the white background, r_b, by averaging the quantity l_b^i/l_s^i at the locations where we had both measurements. We then estimated the reflectance of each palette paper as $r^i = r_b(l_p^i/l_b^i)$. The Munsell values, nominal reflectance of each palette paper obtained from the Munsell standard (Newhall, Nickerson, & Judd, 1943), the reflectance that we measured in situ (as described above), and the measured luminance reflected from each palette paper to the observer are provided in the Supplementary materials. The only measured reflectance that differed reliably from the nominal reflectance was Munsell 2.0; its measured reflectance was higher than expected and relatively close to the nominal measurement of Munsell 2.5. Repeated measures confirmed the reliability of this measurement. Data reported here are based on the measured rather than nominal reflectance values.

Observers indicated their choice of matching Munsell paper via a slider response box. By adjusting the slider, observers could change a numerical value presented on an LCD flat panel display mounted above the Munsell palette. The displayed numerical values were in increments of 0.5 and corresponded to the Munsell values of the papers. These values were also indicated on the palette immediately below each paper. In addition, observers were given two other response options. These were "darker than 2.0" (displayed when the observer moved the slider all the way to the left) and "lighter than 9.5" (displayed when the observer moved the slider all the way to the right).

Observers were instructed to press a button on the slider box when the number displayed on the screen represented the Munsell value of the paper that was most similar in lightness to the target patch or when they had chosen one of the out-of-range options. Observers could look back and forth between the checkerboard display and the matching palette and were instructed to take as long as they needed to make their judgments. After observers

Requested	Measured (cd/m^2)
0.0000	0.096*
0.0013	0.24
0.0018	0.33
0.0024	0.44
0.0032	0.59
0.0042	0.78
0.0056	1.05
0.0075	1.41
0.0100	1.90
0.0133	2.54
0.0178	3.42
0.0237	4.58
0.0316	6.15
0.0422	8.27
0.0562	11.09
0.0750	14.89
0.1000	19.99
0.1334	26.84
0.1778	36.01
0.2371	48.35
0.3162	64.90
0.5623	116.95
0.7499	157.01
1.0000	210.77

Table 1. Mapping between requested luminance [0–1] and displayed luminance (in cd/m^2) for each of the 24 target patches. *This minimum luminance is estimated (see Methods section) because it was below the measurement range of our instruments.

made their response, a new target patch appeared on the HDR display. The checkerboard surround remained on display during this transition and was fixed throughout each block of trials.

Observers ran in 2–3 sessions that lasted between 45 and 60 min. In each session, observers completed blocks of trials. Each block consisted of 3 repetitions each of 24 different target patches in a single checkerboard context. Blocks were constructed such that all 24 target patches were matched before any patch was repeated. A new random order for the 24 target patches was chosen for each repetition. Between blocks, the entire display was set to the minimum luminance while the experimenter initiated a new block of trials with a different checkerboard context. During this time, the room lights were turned on and observers were given the opportunity to take a break. The main experiment consisted of measurements for 9 different checkerboard contexts.

Stimuli

The target patch in each checkerboard context took 24 different luminance values, ranging from 0.096 cd/m^2 to 211 cd/m^2. The smallest value was the minimum luminance value of the HDR display as configured for these experiments. The remaining target patches were selected in equal log steps between 0.24 cd/m^2 and the maximum luminance of the display (211 cd/m^2 at the chosen *xy* chromaticity of 0.43, 0.40). The full list of target patch luminances is shown in Table 1. The same 24 target patch values were used in all nine checkerboard contexts.

A standard checkerboard context was created by taking 24 luminance values between 0.11 cd/m^2 and 211 cd/m^2 (contrast ratio 1878:1) that were equally spaced in $\log_{10}$ luminance. These 24 luminance values were randomly assigned to a 5 × 5 checkerboard surrounding the center target patch. Random draws were taken until neither the brightest nor the darkest check was immediately adjacent to the center target patch. The first appropriate configuration drawn was used as the standard context in all experiments. The center panel of Figure 2 shows a pictorial representation of the standard checkerboard, and the luminance values are reported in Table 2. Although in principle any context could serve as a standard, using a context that spanned a large luminance range and checks sampled that range evenly seemed a reasonable starting point. To create the remaining 8 checkerboards, we divided the 24 checks into an inner ring (which consisted of the 8 locations immediately adjacent to the center target patch) and an outer ring (which consisted of the 16 locations surrounding the inner ring). We chose values for low luminance and high luminance inner and outer rings in the following fashion. First, the low (high) luminance inner ring values were the 8 lowest (highest) luminance values in the standard context (for low ring: minimum luminance = 0.11 cd/m^2, maximum luminance = 1.11 cd/m^2, contrast

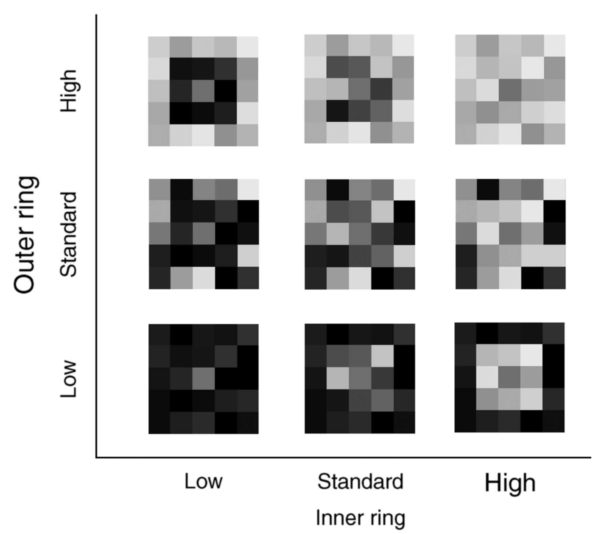

Figure 2. Illustration of the nine checkerboard contexts. Average luminances of inner ring and outer ring were divided into low, standard, and high conditions.

21.26	0.22	15.32	7.95	210.77
40.94	2.98	4.13	78.85	0.16
11.04	56.82	Target	1.54	0.30
0.58	0.42	2.14	5.73	109.43
0.82	29.50	151.87	0.11	1.11

Table 2. Luminance (in cd/m^2) values for the standard checkerboard context. Each value indicates the luminance of one check. The center check is the target patch, which varied on each trial.

ratio = 9.9:1; for high ring: minimum luminance = 21.3 cd/m^2, maximum luminance = 211 cd/m^2, contrast ratio = 9.9:1). To create the luminance values for the low (high) luminance outer ring, we took the minimum and maximum values of the low (high) luminance inner rings, and the remaining luminance values were placed at equal log steps between the minimum and maximum. The remaining 8 checkerboards were the various permutations of these low, standard, and high luminance inner and outer rings (e.g., low–inner, low–outer checkerboard; low–inner, standard–outer checkerboard; low–inner, high–outer checkerboard; *mutatis mutandis*). Table 3 shows the minimum and maximum luminances and contrast ratio for each checkerboard. Spatial locations of the low and high inner and outer rings in each checkerboard preserved the rank order of luminance values in the standard context. Luminance values for all nine checkerboard contexts can be found at http://color.psych.upenn.edu/supplements/hdrlocalremote. This method of creating contexts has several implications. First, the contrast between the brightest and darkest checks varies between contexts. Both the low–low and high–high luminance checkerboards have a lower contrast than the standard context (see Table 3). Second, even within one spatial ring, there is not a single multiplicative factor that scales the luminances in the standard checkerboard to those in the test checkerboards. This lack of contrast invariance means that the test checkerboards cannot be obviously characterized as differing from the standard checkerboard by a single (or double) change in a simulated illuminant.

Display system and stimulus characterization

Grayscale checkerboard images were presented on a custom computer-controlled high dynamic range (HDR) display (see Figure 1). The design of the HDR display was adopted from Seetzen et al. (2004). The output from a DLP video projector (Panasonic #PT-D7600U) was projected onto a 19″ LCD display panel (ViewSonic), through a Fresnel lens and diffuser placed directly against the backside of the panel (where its backlight would normally go). The projector was equipped with a short-throw lens (Panasonic DLP Projection Fixed Lens SXGA 0.8/XGA 1.0); the front edge of the lens was 20 cm from the LCD panel assembly. Because the LCD panel is a transmissive display, it provides a multiplicative attenuation of the projector image, resulting in an overall dynamic range that is the product of the native dynamic ranges of the projector and panel. Both display devices were driven at a pixel resolution of 1280 by 1024 and at a refresh rate of 60 Hz by a dual-port video card (NVIDIA GeForce GT 120). The host computer was an Apple Macintosh G5.

The displays were arranged so that the LCD panel was enclosed in a box that prevented stray light within the experimental room from reaching the front of the panel and reflecting back to the observer. Visible surfaces within this box were lined with light absorbing black cloth. The observer viewed the LCD panel monocularly from a distance of 73 cm through a circular aperture 6.1 cm in diameter at the end of the enclosing box. The observer's head was stabilized with a chin rest, which could be adjusted so that the eye was centered in the circular aperture.

To display calibrated high-resolution images on the HDR display, it is necessary both to align the projector image with the LCD panel and to map desired stimulus values to appropriate RGB input settings for the two video cards. These tasks were completed using custom software developed in the laboratory, following general methods outlined by Seetzen et al. (2004). The experimental software consisted primarily of MATLAB routines. To control the display, we also relied on routines from the Psychtoolbox

	Inner min	Inner max	Inner CR	Outer min	Outer max	Outer CR	Overall CR
Low–low	0.11	1.11	9.92	0.11	1.11	9.92	9.92
Low–standard	0.11	1.11	9.92	0.11	210.77	1878	1878
Low–high	0.11	1.11	9.92	21.26	210.77	9.92	1878
Standard–low	0.42	78.85	189.36	0.11	1.11	9.92	702.4
Standard	0.42	78.85	189.36	0.11	210.77	1878	1878
Standard–high	0.42	78.85	189.36	21.26	210.77	9.92	506.1
High–low	21.26	210.77	9.92	0.11	1.11	9.92	1878
High–standard	21.26	210.77	9.92	0.11	210.77	1878	1878
High–high	21.26	210.77	9.92	21.26	210.77	9.92	9.92

Table 3. Minimum and maximum luminance values (in cd/m^2) and contrast ratios for the inner and outer rings of all 9 checkerboard contexts. The left column denotes the luminance profile (low, standard, or high) of the inner (first word) and outer (second word) rings. The contrast ratio (CR) was calculated by dividing the maximum luminance of the ring by the minimum luminance of the ring.

(Brainard, 1997), MGL toolbox (http://gru.brain.riken.jp/mgl), and custom C routines that were called from MATLAB and that accessed the OpenGL API directly.

To align the two displays, an observer adjusted a 20 by 20 grid presented by the projector so that it aligned with a corresponding fixed grid displayed on the LCD panel. The alignment coordinates were used to create a warping map for the images displayed on the projector so that they were in spatial register with the images on the LCD panel. The warping was performed at the frame rate by processing on the video card. Because the Fresnel lens/diffuser/LCD panel onto which the image was projected had a significant thickness, this alignment was specific to the observer's eye position. The use of an aperture in the display enclosure ensured consistency of this position across sessions and observers. Accidental movement of the Fresnel lens/diffuser/LCD panel could cause the projector and LCD displays to become misaligned. Experimenters periodically inspected alignment of the experimental stimuli before data collection and repeated the alignment procedure as necessary.

We used the PR-650 spectral radiometer to characterize the properties of the projector and LCD panel separately. This was done in situ, with the radiometer placed at the observer's eye position. First, we characterized the projector, which we used as a grayscale device so that its RGB values were always set with R = G = B. We set the RGB input values of the LCD panel to their maximum level (corresponding to maximum transmission through the panel) and measured the relation between the R = G = B values to the projector and luminance output for a series of 30 input values. We then splined these to produce a full gamma curve for the projector. Second, we set all projector pixels to their maximum input values (full light output) and measured separately the gamma curves of the R, G, and B channels of the LCD panel, as well as the transmitted spectrum for each channel. For any desired display luminance and chromaticity, the characterization data were used to compute an R = G = B value for the projector and R, G, B values for the LCD panel that produced the desired output. The algorithm followed that of Seetzen et al. For any given desired luminance and chromaticity, we first used the calibration data for the LCD panel, obtained when the projector was set to maximum output, to determine the linear proportion of each red, green, and blue primary required to produce the desired output. We then took the maximum of these three values and found the square root of this maximum. The projector R = G = B input value was set to produce this proportion of maximum luminance. We then compensated for this decrease by appropriate choice of LCD panel R, G, and B input values. The procedure apportions the luminance attenuation at each pixel across the projector and LCD panel.

The HDR display is a new device in our laboratory, and more generally such displays are only recently coming into use in visual psychophysics. Thus, we are still gaining experience with precise stimulus control for such displays. Here, we note some limitations in the precision to which our procedures characterized the stimuli. Development of improved stimulus characterization and control procedures is ongoing in the laboratory. A key limitation is that the lowest luminances displayable by the device are below the minimum luminance that could be reliably measured by the radiometer. Visual inspection, however, revealed discriminable differences between these low luminance stimuli. To estimate the lowest luminances, we measured the luminance of the center patch on the checkerboard for 97 test values that spanned the requested stimulus range in the experiments. A line (in log–log space: slope = 1.02; y-intercept = 2.32) provided an excellent fit between measured luminance values and input luminance values across the range of test stimuli that we could measure. The full list of 97 input luminance values and measured luminance and chromaticity values are available in the Supplementary materials. Checkerboard and target luminances reported here are those obtained from the fit to these data, with luminance values below the instrument's measurement range obtained by extrapolation. We somewhat arbitrarily assigned the luminance corresponding to nominal luminance of 0 to be 0.096 cd/m^2. This was obtained by subtracting the difference between the lowest two extrapolated values from the lowest extrapolated value. We thus view the data obtained for the low end of the luminance range as less precise than that for higher luminances.

A second limitation is that we calibrated the central patch of the display only; there is likely to be some location-to-location variation in the stimuli at the other display locations. Values reported for checkerboard luminances are those obtained for the center location with the same input settings.

Finally, direct measurements revealed some variation in target square chromaticity from the desired target values. Across test luminance range, x chromaticities were in the range [0.43–045] and y chromaticities were in the range [0.38–0.40]. The measured variability was more pronounced at lower test luminances. The chromaticity variations were not visually salient; the stimuli appeared to vary primarily in luminance.

Data analysis

The goal of this paper is to understand how context affects the map between luminance and perceived lightness. To do so, we computed context transfer functions (CTFs) by analyzing the data to establish target patch luminances that were matched to common palette papers. For each target patch, observers indicated the best matching Munsell paper. The Munsell palette papers, which remained unchanged as checkerboard contexts and target patches varied, provided the reference for computing CTFs. If observers matched two target patches in different checkerboard contexts [L_x, L_y] to the same Munsell paper (where the subscript indicates checkerboard context), then we took those luminance values to be perceptually equivalent. That is, $L_x \simeq L_y$, where $\simeq$ denotes perceptual equivalence.

To compute CTFs for human observers, we calculated perceptually equivalent luminance values for each Munsell paper in each context. We write

$$\mathrm{MP}_a \simeq \left(\frac{\sum_{i=1}^{N_1} \log_{10}(L_1^i)}{N_1} \right) \simeq \left(\frac{\sum_{i=1}^{N_2} \log_{10}(L_2^i)}{N_2} \right) \ldots$$

$$\simeq \left(\frac{\sum_{i=2}^{N_9} \log_{10}(L_9^i)}{N_9} \right), \quad (1)$$

where MP_a represents a given Munsell paper (and a ranges from 2.0 to 9.5) and $L_x^{\ i}$ represent the test luminances that were matched to that Munsell paper (MP_a). The value of N_x varies with both Munsell paper and context and represents the total number of test luminances that were matched to that Munsell paper.

Equation 1 gives us estimates of perceptually equivalent luminances across contexts. If there existed a Munsell paper to which no luminance value was mapped by an observer in a given context, that observer's data point was excluded from the average calculation. For each context, we discarded data for Munsell papers where fewer than half of the observers matched any stimulus to that paper. The data are tabulated in the Supplementary materials.

In principle, one can describe the CTF between any two contexts; here, we always use the average of all observers in the standard checkerboard context as a reference so that the data set consists of the CTFs between the standard context and each of the 8 other contexts.

Observers could also respond "darker than 2.0" or "lighter than 9.5." Because these judgments did not correspond to a palette paper, they were dropped from further analysis. Of all 504 judgments across 7 observers in the standard checkerboard context, 50 were "darker than 2.0" and 14 were "lighter than 9.5." The number of out-of-range judgments in other contexts can be found in the Supplementary materials.

We fit the CTFs using two models. Both models were derived using the classic framework that perceptual matches occur when two stimuli produce the same response in an internal "lightness" mechanism (Fechner, 1966; Heinemann, 1961; Hillis & Brainard, 2007a, 2007b). We assume that the relation between target luminance and mechanism response varies with context. For the first model, we assumed that the only adaptation parameters controlling the variation are multiplicative gain and subtractive offset. A wide variety of experimental evidence provides support for both multiplicative and subtractive adaptation (Chubb, Sperling, & Solomon, 1989; Hood & Finkelstein, 1986). Note that in their simplest forms, models that postulate Weber contrast coding of the stimulus followed by multiplicative contrast gain control (Chubb et al., 1989; D'Zmura, 1999) or divisive contrast normalization (Blakeslee & McCourt, 2004; Heeger, 1992) are algebraically equivalent to models that postulate a combination of multiplicative and subtractive adaptation. In addition, Gilchrist's (2006) notions of anchoring combined with scaling and Adelson's (2000) elaboration of an affine atmospheric transfer function represent models that incorporate two adaptation parameters.

For the first model (gain–offset), the mechanism response (y) in any context is given by

$$y = f[g * (L - o)], \quad (2)$$

where g is a multiplicative gain parameter, o is a subtractive offset, and $f()$ is a fixed and invertible nonlinear function. Thus, the model yields

$$f(g_{\mathrm{st}} * (L_{\mathrm{st}} - o_{\mathrm{st}})) = f(g_x * (L_x - o_x)), \quad (3)$$

where L_{st} and L_x are perceptually equated luminances between the standard context and context x. Since $f()$ is invertible,

$$g_{\mathrm{st}} * (L_{\mathrm{st}} - o_{\mathrm{st}}) = g_x * (L_x - o_x). \quad (4)$$

This provides a predicted relation between the luminances of the stimuli that match in lightness across contexts:

$$L_x = g * (L_{\mathrm{st}} - l_0), \quad (5)$$

where $g = \frac{g_{\mathrm{st}}}{g_x}$ and $l_0 = o_{\mathrm{st}} - \frac{o_x}{g}$. Using numerical parameter search (Matlab fmincon), we solved for the values of g and l_0 that minimized the sum squared error between the observed and predicted $\log_{10} L_x$.

We found that the fits produced by the first model deviated systematically from the data (see Figure 3 for model fits). For this reason, we developed a second model (gain–offset–exponent) with additional adaptation parameters. To do so, we used the specific form of the Naka–Rushton function for $f()$:

$$f(y) = R_{\max} * [y^n / (y^n + y_0^n)]. \quad (6)$$

We fixed $R_{\max} = 1$ and $y_0 = 1$ and allowed the exponent n to vary with context. This exponent controls the steepness of the sigmoidal function described by Equation 6. Under this model, we predict the luminance of stimuli that match in lightness across context as

$$L_x = f_x^{-1}(f_{\mathrm{st}}((g_{\mathrm{st}} * (L_{\mathrm{st}} - l_{0_{\mathrm{st}}}))) / g_x + l_{0_x}. \quad (7)$$

We set the exponent of f_{st} as $n_{st} = 3$ and found g_{st} and $l_{0_{st}}$ so that the luminances of the standard context were spread across the non-saturating regime of f_{st}. We then used parameter search to find the parameters g_x, l_{0_x}, and n_x that provided the best fit to each CTF.

Results

The measured context transfer functions (CTFs) for our eight test contexts, relative to the standard context, are shown in Figure 3. Each plotted point shows data for one palette paper. The average luminance of all the target patches matched to that palette paper in the test checkerboard context is shown on the x-axis, while the y-axis represents the average luminance of all the target patches matched to that palette paper in the standard checkerboard context. The extent to which each plotted point deviates from the diagonal indicates the magnitude of the effect of changing from standard to test context.

The CTFs have several salient features. First, in cases where the luminance of the checkerboard context unambiguously decreased relative to the standard context (top left, top right, and bottom left panels of Figure 3, red lines), target patches appeared lighter (red points shifted leftward from the diagonal). Conversely, when the luminance of the checkerboard context unambiguously increased (top left, top right, and bottom left panels of

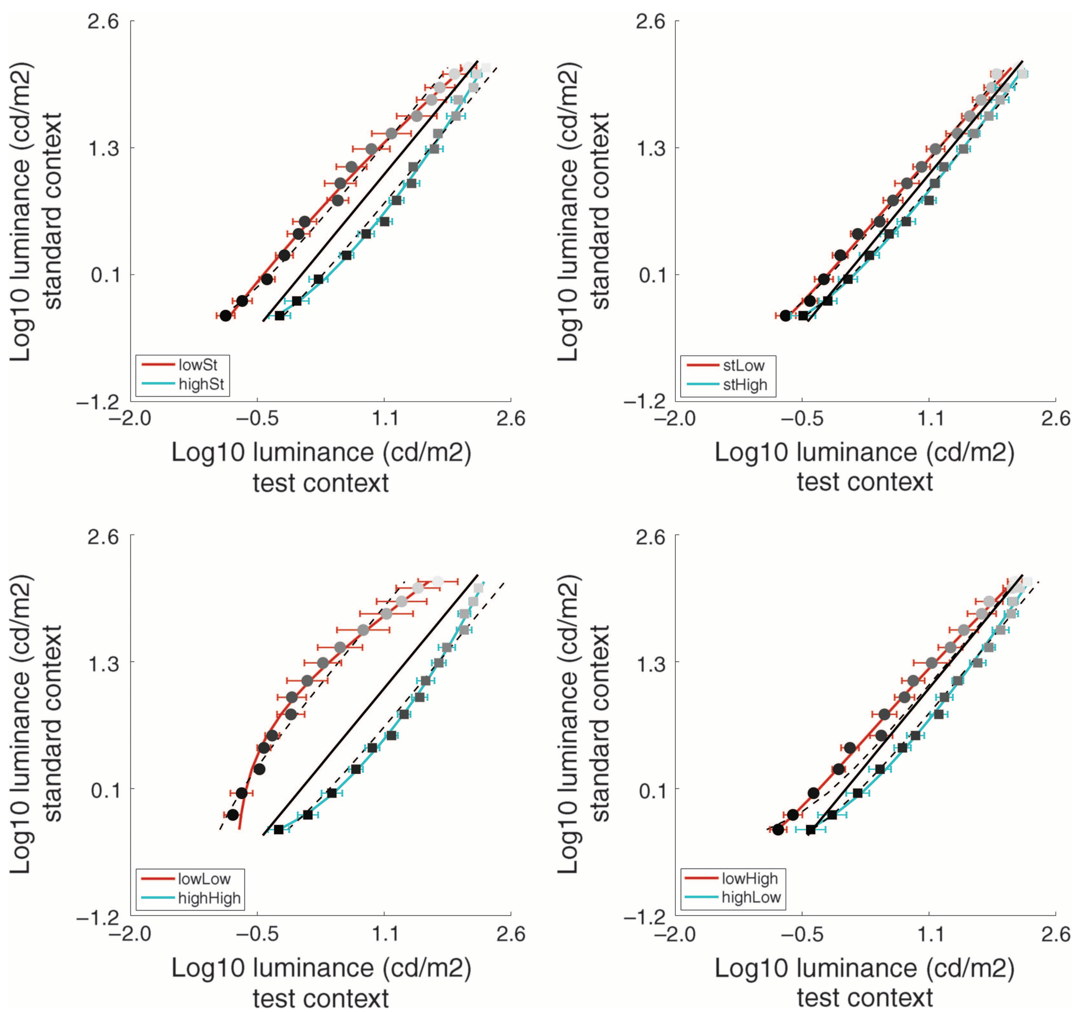

Figure 3. Effect of context on lightness. Each data point represents the average of the target luminance values matched to a different Munsell paper in a test checkerboard context (x-axis) and the standard context (y-axis). The top left panel shows data for the low inner, standard outer (red/circles) and high inner, standard outer (cyan/squares) test contexts; the top right panel shows data for the standard inner, low outer (red circles) and standard inner, high outer (cyan squares) contexts; the bottom left panel shows data for the low inner, low outer (red/circles) and high inner, high outer (cyan/squares) test contexts; the bottom right panel shows data for the low inner, high outer (red/circles) and high inner, low outer (cyan/squares) contexts. Error bars are *SEM* across observers. Lines are fits to the gain–offset model (black, dashed) and the gain–offset–exponent model (colored, solid). The standard context is defined as the average match across all observers in the standard condition.

Figure 3, blue lines), target patches appeared darker. Qualitatively, both of these effects are consistent with the idea that perceived lightness is determined by a comparison of the target's luminance to some aggregate of the luminance surrounding the target. In addition, if these aggregate changes in surround luminance are interpreted as illuminant changes of the same sign, then both effects are qualitatively consistent with lightness constancy. Note also that both effects can be quite large. For example, going from the darkest to the brightest context changed the average luminance matched to Munsell 5.0 (a mid-gray) by a factor of 23 (average match: 0.87 cd/m^2 in low–low, 20 cd/m^2 in high–high). For one observer, CH, a target patch perceived as white (Munsell 9.5) in one context (2.54 cd/m^2, low–low) was perceived as black (Munsell 2.0) in the brightest context (high–high). Use of an HDR display provides the capability of measuring lightness across a large luminance range, without substantial floor or ceiling effects.

Second, and also unsurprisingly, manipulating the luminance of the inner ring alone appears to have a greater effect on the CTF than does manipulating the luminance of the outer ring alone. For each context change, the locus of the points in the CTF provides a measure of the effect of context, and in the top left panel of Figure 3, the CTFs are shifted more than in the top right panel of Figure 3. We do note that our comparison of inner and outer ring effects is specific to the particular choice of luminance distribution in the inner and outer rings.

Third, changing both rings together has a greater effect than either alone (bottom left panel of Figure 3) and it appears that for our stimuli this effect is larger for decreasing luminance (red line).

Fourth, in addition to the magnitude of the contextual effect changing between contexts, the shape of the CTF changes between contexts. For example, when only the outer ring is changed, the CTF is fairly linear (top right panel of Figure 3), but when the luminance of both rings is changed, the CTFs are curved (bottom left panel of Figure 3).

To quantify these four effects, and to understand these broad features of the data more completely, we asked whether the locus of the points has an interpretable parametric form. The thin dashed lines in Figure 3 show fits of our first model, one that incorporates both multiplicative and subtractive adaptation (see Methods section). This model captures overall shift of the data relative to the diagonal, although it clearly fails in detail. This overall shift in the model fits is driven primarily by the multiplicative gain parameter. Indeed, if the subtractive offset term were forced to be zero, the fits would be lines with unit slope and the gain parameters would allow shifts of these lines to pass in aggregate through each CTF.

Although the gain–offset model captures the overall trends in the data, it fails in detail. In particular, most of the CTFs show significant curvature that is not fit by the model. This curvature is most noticeable for the low–low and high–high contexts (bottom left panel of Figure 3). The colored solid lines in the figure show the fit of our second model, where the exponent of the underlying sigmoid is allowed to vary with context. Allowing this variation captures the curvature of the CTFs and provides a good fit to the data. We have also found that the same parametric form can account for measurements of perceived lightness in high dynamic range contexts that do not segregate inner and outer rings (Radonjić et al., 2011). Table 4 provides the fitted model parameters for each context.

Because the model fits the data, we can summarize the effect of context on lightness by characterizing how context affects the model parameters. For this purpose, we have found it most intuitive to reparameterize the model. Rather than examining the gain, offset, and

	$\text{Log}_{10}(g)$	I_0	Exp	Log_{10}(black point)	Log_{10}(white point)	Log_{10}(gray point)
Low–low	−1.00	5.90	1.78	−0.66	1.50	−0.02
Low–standard	−0.36	2.79	2.19	−0.65	1.95	0.58
Low–high	−0.21	1.88	2.41	−0.57	2.02	0.74
Standard–low	−0.16	1.09	2.65	−0.45	1.99	0.81
Standard	0*	0*	3*	−0.20	2.07	0.98
Standard–high	0.17	−1.46	3.38	−0.19	2.16	1.14
High–low	0.33	−3.25	3.96	−0.08	2.22	1.28
High–standard	0.40	−4.81	4.57	0.04	2.22	1.34
High–high	0.55	−7.72	5.90	0.12	2.23	1.45

Table 4. Best fit parameters of the multiplicative gain–offset–exponent model of the CTFs for each checkerboard context. Values are parameters fit to the average data. The left column denotes the luminance profile (low, standard, or high) of the inner (first word) and outer (second word) rings. Column values are given as follows: (2) log of the gain parameter; (3) the subtractive offset parameter; (4) the exponent parameter; (5) the $\log_{10}$ of the black point, calculated by evaluating the model at the black point (the average luminance matched to 2.5) in the standard context (0.64 cd/m^2); (6) the $\log_{10}$ of the white point, calculated by evaluating the model at the white point (the average luminance matched to 9.0) in the standard context (117 cd/m^2); (7) the $\log_{10}$ of the gray point, calculated by evaluating the model at the gray point (the average luminance matched to 5.5) in the standard context (9.49 cd/m^2). *Notes*: *Indicates fixed values in the reference condition.

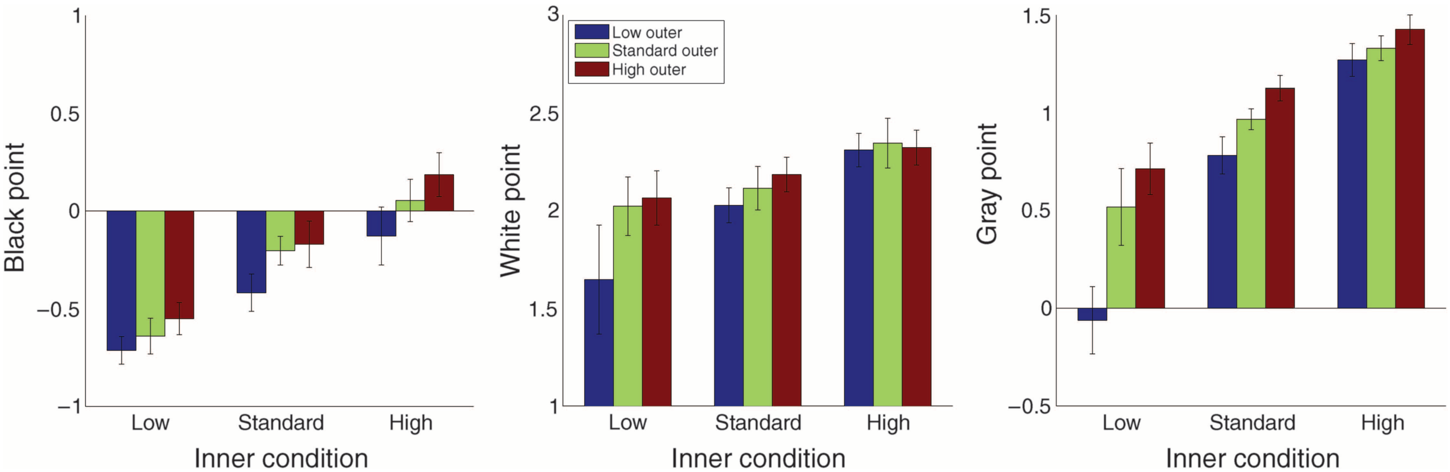

Figure 4. Effect of checkerboard context on black point (left panel), white point (center panel), and gray point (right panel). Black points (taken as model predictions for match to N 2.5), white points (taken as model predictions for match to N 9), and gray points (taken as model predictions for match to N 5.5) were grouped by inner ring conditions for each of our nine conditions. The values shown were obtained by averaging the values (expressed as $\log_{10}$ cd/m^2) obtained from fits to individual observer data; error bars show standard error of the mean.

exponent, which do not directly and explicitly describe the measured CTFs, we examine the model predictions for what we refer to as the black, white, and gray points. In concept, these are the luminance values that appear black, white, and gray in each context; here, they are defined operationally as the predictions for luminance in each test context that match luminances of 0.64 cd/m^2, 117 cd/m^2, and 9.49 cd/m^2 in the standard context. These luminances correspond to the matches made in the standard context to the Munsell 2.5, 9.0, and 5.5 palette papers. Because the data are well fit by a 3-parameter model, most three-point reparameterizations should be equivalent; we verified that specification of the black, white, and gray points is sufficient to recover the model parameters for each context.

Checkerboard context affects the black point, as shown in Figure 4. Decreasing the luminance of the context decreases the black point (left bars, left panel), and increasing the luminance of the context increases the black point (right bars, right panel). More generally, the black point varies systematically from left to right in Figure 4, indicating a regular dependence on the overall luminance of both the inner and outer checkerboard rings. As one might expect, the inner ring has a stronger effect than the outer ring. For example, decreasing (increasing) the inner ring changed the black point by an average of 0.42 (0.66) cd/m^2, while decreasing (increasing) the outer ring changed the black point by an average of 0.24 (0.15) cd/m^2. Note also that the effect of context on the black point is not due solely to the lowest contextual luminance. As shown in Figure 5, in all but two of the contexts (standard inner–high outer and high inner–high outer) the lowest luminance in the overall checkerboard context is the same (0.11 cd/m^2).

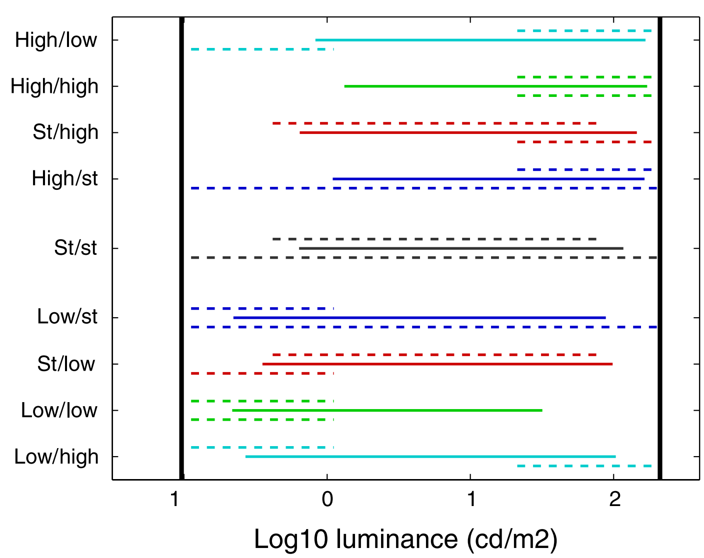

Figure 5. Effect of checkerboard context on the black point and white point. Each solid line begins at the black point and ends at the white point for that context. Black and white points are values from Table 4. Note that since these values are computed from the parameters fit to the average data, they may differ slightly from the values in the left and center panels of Figure 4, which are the average of values computed from parameters fit to individual subjects' data. The dashed lines show the luminance range of the checkerboard context (inner ring above the solid line; outer ring below the solid line). The vertical solid black lines indicate the minimum (left solid line) and maximum (right solid line) target luminances.

The increase in the black point could arise in two different ways. The same 24 target patches were used in every context and observers were given the option of reporting darker than 2.0. An increase in the black point could occur because observers are calling more low luminance stimuli "darker than 2.0," thus excluding them from contributing to the average luminance matched to palette papers, or because observers are matching progressively more stimuli to the low end of the palette. Both effects occur in our data. The total number of judgments that are out of range in each context (out of 504) varies from 10 in the low–low context to 91 in the high–high context. The total number of 2.0 and 2.5 judgments in each context (again out of 504) varies from 16 in the low–low condition to 93 in the high–high condition.

Checkerboard context also affects the white point, as shown in the center panel of Figure 4. Decreasing the luminance of the context decreases the white point (left bars, center panel), and increasing the luminance of the context increases the white point (right bars, center panel). As with the black point, the inner ring has a greater effect than the outer ring, although this effect is not as pronounced as for the black point. For our contexts, the white point is not determined solely by the highest contextual luminance. All but two of our contexts have the same highest luminance (Figure 5), and the white point varies across these contexts. At the same time, we note that the deviations from the "highest luminance appears white" anchoring rule for these seven contexts are modest, about 0.3 log units.

To examine in more detail the effect of context on the black and white points, we tested whether the effects of inner and outer rings were additive. To do so, we conducted 3-way ANOVAs on both the black and white points. Both inner and outer rings have a significant additive effect on both the black and white points, but there was not a significant interaction (black point, Table 5; white point, Table 6). In other words, the effect on black and white points of changing the outer ring does not depend on the luminance profile of the inner ring, and vice versa. For the white point, this result is qualitatively consistent with Gilchrist's (2006) notion that the visual system segments the image into "frameworks" and that the white point within each framework represents a weighted combination of influence between frameworks.

Source	SumSq	*df*	MeanSq	*F*	p
Observer	5.45	6	0.91	20.91	<0.001
Inner	1.81	2	0.90	20.79	<0.001
Outer	0.47	2	0.23	5.38	<0.01
Interaction (inner outer)	0.36	4	0.09	2.10	0.10 (n.s.)
Error	2.09	48	0.04		
Total	10.18	62			

Table 6. ANOVA for white points. The table shows the results of a 3-way ANOVA on the individual observer white points. Observer was coded as a random effects variable, while inner and outer rings were coded as fixed effects. The ANOVA modeled all main effects and the inner by outer interaction.

The gray points also vary systematically with context. In addition, the variation of the gray points is not described as the result of additive effects of inner and outer rings (ANOVA, Table 7). Consideration of Figures 3 and 4 suggests that the interaction may be driven by the data from the low–low and high–high conditions: In Figure 3, this is indicated by the increased curvature for these conditions and in Figure 4 by the fact that the gray point is lower than expected for low–low condition.

The low–low and high–high checkerboards differ from the other contexts in several different ways. First, the overall range of luminance in the context is much lower (9.92:1 in both) than in any other checkerboard. Because of this, and because we have the same 24 target patches in every context, this means that for these two conditions, a significant portion of our target patches are out of the range of the checkerboard context. In the low–low checkerboard, 17 of the 24 target patches were brighter than the brightest square in the checkerboard; in the high–high checkerboard, 17 of the 24 target patches were of lower luminance than the lowest luminance in the checkerboard.

Source	SumSq	*df*	MeanSq	*F*	p
Observer	3.20	6	0.53	32.45	<0.001
Inner	4.76	2	2.38	144.67	<0.001
Outer	0.63	2	0.32	19.17	<0.001
Interaction (inner outer)	0.06	4	0.02	0.97	0.43 (n.s.)
Error	0.79	48	0.02		
Total	9.45	62			

Table 5. ANOVA for black points. The table shows the results of a 3-way ANOVA on the individual observer black points. Observer was coded as a random effects variable, while inner and outer rings were coded as fixed effects. The ANOVA modeled all main effects and the inner by outer interaction.

Source	SumSq	*df*	MeanSq	*F*	p
Observer	2.85	6	0.47	10.63	<0.001
Inner	9.65	2	4.82	108.02	<0.001
Outer	1.95	2	0.98	21.84	<0.001
Interaction (inner outer)	0.83	4	0.21	4.62	<0.005
Error	2.14	48	0.04		
Total	17.42	62			

Table 7. ANOVA for gray points. The table shows the results of a 3-way ANOVA on the individual observer gray points. Observer was coded as a random effects variable, while inner and outer rings were coded as fixed effects. The ANOVA modeled all main effects and the inner by outer interaction.

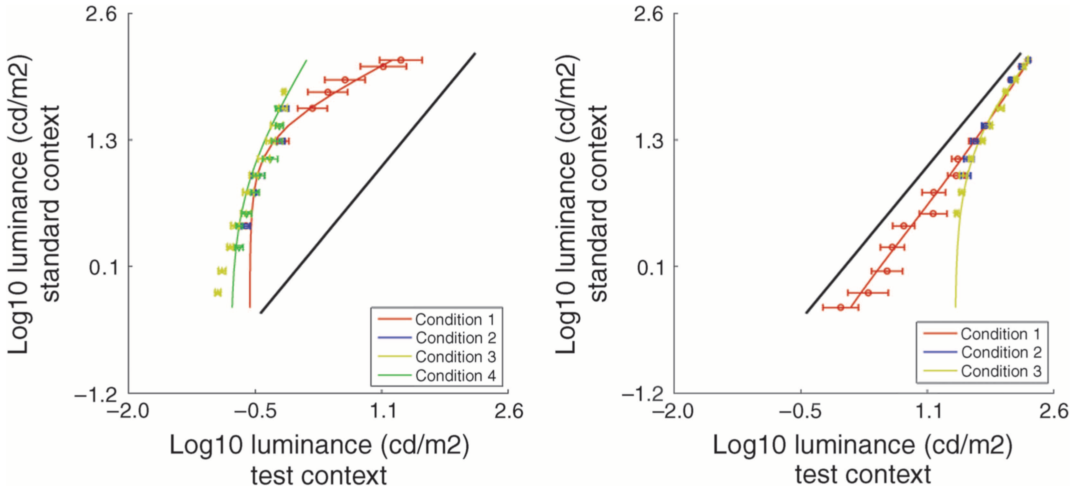

Figure 6. Effect of target stimulus range on CTF. For 5 observers, we plot the CTFs for 4 sets of target patches in the low–low checkerboard context (left panel) and 3 sets of target patches in the high–high context (right panel): (1) the standard target patches from the main experiment (both panels, red data points); (2) the subset of target patches from Condition 1 that fell within the luminance range of the checkerboard context (both panels, blue points, 7/24 target patches); (3) a low range set of 24 target patches that fell within the luminance range of the checkerboard context (both panels, yellow points). For the low–low checkerboard, the 24 target patches were equal log steps between 0.11 cd/m^2 and 1.11 cd/m^2 luminance; for the high–high checkerboard context, the 24 target patches were equal log steps between 21.2 cd/m^2 and 211 cd/m^2; and (4) for the low–low checkerboard context (left panel), a medium luminance set (green data points) consisted of 24 patches in equal log steps between 0.18 and 20.0 cd/m^2, from which we plot the 9 target patches within the luminance range of the checkerboard. Error bars are *SEM*. Yellow lines represent the best fit gain–offset–exponent model calculated with target stimuli from all conditions that were within checkerboard luminance range (yellow, green, and blue data points in the left panel and yellow and blue data points in the right panel). Red lines are the best fit gain–offset–exponent model for the standard target patches (red points in both panels). Differences between the CTFs calculated with the standard target stimuli (red lines) and the low–low and high–high CTFs in Figure 3 likely reflect differences between observers in the main and control experiments.

To investigate whether the test stimuli that were outside the range of the checkerboard context exerted an effect on the appearance of target patches that were within the range of the checkerboard context, we performed a control experiment with 5 observers where we varied the range and values of the target patches presented in the low–low and high–high checkerboard contexts and fit the model to the resulting CTFs. In Condition 1, the five observers made matches for the same set of target patches as in the main experiment. In Condition 2, we used the same data set as in Condition 1, but we computed the CTF using only the subset of target patches that fell within the luminance range of the context (7/24 target patches). In Condition 3, 24/24 target patches were within the range of the checkerboard context. Finally, in Condition 4 (low–low context only), 9/24 target patches were within the range of the checkerboard context. In Figure 6 (left panel), we compare the CTFs for the different sets of target patches presented in the low–low context. If presenting out-of-range target patches within an experimental session affected the lightness of those target patches that were within the range of the checkerboard context, then the data from Conditions 2–4 (left panel) would differ from each other; however, the green, blue, and yellow data points are very similar to each other in the left panel of Figure 6. The right panel of Figure 6 shows that the same broad trend is true for the high–high context, that is, presenting target patches outside the range of the checkerboard context in an experimental session does not affect judgments of target patches made within the range of the checkerboard context.

Although the data in Figure 6 show that judgments of in-range target patches are affected very little by within-session presentation of target patches outside of the luminance range of the checkerboard context, they also show that extrapolations of appearance for target patches outside of the range are very inaccurate. For comparison, the CTF calculated with the full target patch range for the control observers is shown in Figure 6. The white point extrapolated from the within-range CTF for the low–low condition is 1.30 cd/m^2, compared to 11.3 cd/m^2 for the CTF from the full target patch range. Similarly, the black point extrapolated from the within-range CTF for the high–high condition is 1.83 cd/m^2, compared to 25.9 cd/m^2 for the CTF from the full target patch range.

Discussion

A complete model of the perception of surface lightness would allow prediction of the lightness of any image region, given the luminance of each location in the image.

We are currently far from that goal, particularly for images that are typical of natural viewing. The present results serve to provide one step toward a complete model, in that they characterize lightness for simple images that contain some salient features of natural images. In particular, our images exhibit a high range of luminance variation, and across images the spatial grouping of this variation was varied by the introduction of photometric variation between the inner and outer checkerboard rings. We close by summarizing our data and emphasizing key features of the results.

We measured the effect of context on the mapping between luminance and lightness. To do so, we asked observers to match the luminance of a test patch embedded in a checkerboard to a standard series of Munsell papers. The Munsell papers served as a convenient place holder to infer perceptually equivalent luminances across contexts; we assumed that when two different luminance values were matched to the same Munsell paper, then to within the resolution of the Munsell papers, those two luminance values appeared the same. In principle, we could have used any matching palette that spanned the appropriate reflectance range, since rather than being viewed as meaningful in itself the palette was used only to link the appearance of luminance values between contexts. As long as the perception of lightness in the palette is monotonic with perception of lightness in the checkerboards, the derived CTFs will be invariant to palette choice. Although this assumption seems reasonably secure, we do note that recent research (Logvinenko & Maloney, 2006) suggests that under some circumstances, two perceptual dimensions are required to capture the full richness of the phenomenology of lightness. To the extent that a second dimension is required, the CTFs may fail to capture the full effect of context on perceived lightness.

In broad strokes, several features of our results are consistent with previous work. First, the CTFs shift up from the identity line when overall context luminance is increased and down from the identity line when overall context luminance is decreased. It would be surprising if this were not the case: A long tradition of lightness research suggests that changing the luminance surrounding a target patch causes a change of opposite sign in perceived lightness. Second, context changes near the target patch have a larger effect than context changes further from the target patch. This was also expected and is consistent with many previous studies in lightness and color demonstrating that the size of contextual effects is dependent on the proximity of contextual changes (Gilchrist, 2006; Hong & Shevell, 2004; Rudd & Zemach, 2004).

Other features of our data are novel. First is the manner in which the shape of the CTFs varies with context. Early proposals about how context affects lightness focused on the notion that lightness is computed via a ratio to some reference luminance (Land, 1986; Wallach, 1948, see Brainard & Wandell, 1986) or as a fixed function of contrast. These models predict that the CTFs will plot as lines of slope 1 in the type of log–log representation we employ and are clearly contradicted by the data. Deviations from a line of unit slope are also predicted by the parametric form required to account for a variety of visual phenomena, however. These include brightness induction (Spehar, Debonet, & Zaidi, 1996), chromatic induction (Jameson & Hurvich, 1972), subtractive adaptation in sensitivity regulation (Adelson, 1982; Geisler, 1978), background discounting (Shevell, 1978; Walraven, 1976), contrast adaptation (Chubb et al., 1989), and scale normalization (Gilchrist, 2006). Our gain–offset model implements the core feature of these models as they apply to our stimulus configuration. Indeed, the gain–offset model may be thought of as incorporating what are often referred to as first-site (multiplicative gain control) and second-site (subtractive) adaptation. Although the gain–offset model captures the broad trends of the data, however, it clearly misses in detail (Figure 3). To account for our data, we must also allow an additional parameter to vary with context. In particular, we were able to fit the measured CTFs well with our gain–offset–exponent model (Figure 3). As discussed in much more detail elsewhere, the response functions inferred from our model provide a representation of context effects that is in the same form as typical neural measurements (Radonjić et al., 2011), so that a potential use of the model is to infer how we might expect neural mechanisms subserving lightness perception to adapt to different visual contexts.

Our ability to reject contrast-coding and gain–offset models derives from the fact that we varied test luminance over a large range, providing us with a much fuller picture of how lightness varies with luminance across contexts than can be obtained when the test stimuli are varied over a more restricted range. In his classic study, for example, Wallach varied test luminance over a range of about 10:1, much smaller than we employ here. More generally, we have not found many studies in the literature that study lightness parametrically across a large range of test luminances (but see Bartleson & Breneman, 1967; McCann, 2006.)

The CTFs also indicate that even without the geometric cues to scene segregation that occur under natural viewing conditions, observers' lightness matches are consistent with the visual system treating the photometric variation in checkerboard context as spatial variation in the illumination. For example, the changes in observers' lightness between the standard and high–low checkerboard contexts are qualitatively consistent with the direction of the effect one would expect for the performance of a lightness constant visual system that treated the stimuli as a set of surfaces whose central region was illuminated by a small spotlight (bottom right panel in Figure 2). However, inferences about how the visual system parsed the stimuli into separately illuminated regions must remain speculative, since our stimuli were not constructed as simulations of illuminated surfaces nor did we measure either the observers' estimates of the illumination or the perceived lightness at locations in the checkerboard context other than the central target patch. Indeed, the impracticality of making the vast number of measurements required to characterize the lightness of every check illustrates the need for models that specify lightness and illumination at all locations in a scene.

In considering the relationship between context and CTF, an important conceptual issue arises. Ideally, one would study context effects using experimental manipulations that fix the visual system's state of adaptation at a target location, and then probe this state by studying the response to a set of targets presented at this location. In this ideal situation, the target itself would not perturb the state of adaptation, so that the psychophysical measurements across a set of targets reveal performance for a fixed adaptation state set entirely by the context (Stiles, 1967). A limitation of this approach is that there is no guarantee that the target in fact has no effect on the adaptation state. Indeed, the analysis in Figure 6, which shows that the shape of the predicted CTF varies with the test range for some of our contexts, suggests that there is a test effect. Thus, although we fit our data with a model derived from mechanistic ideas about adaptation, it is important to note that the model parameters probably reflect the combined effect of the checkerboard context and the target itself.

We do not regard this as problematic for our fundamental purpose, which is to obtain a functional characterization of how context affects the map between luminance and perceived lightness. For this purpose, what matters is that for a given context, the model fits the measured CTF. However, it is important to remember that as we ultimately try to relate the data directly to neural mechanisms of adaptation, some caution in the mapping between model and mechanism parameters will be warranted. Indeed, for the first author, the phenomenal experience of increasing target patch luminance much beyond the highest luminance in the low–low checkerboard context is that the lightness of the target patch remains constant, while the entire checkerboard context appears to get darker. This should be taken as an introspective note rather than an empirical claim since, as noted previously, we measured the perceived lightness of central target patch and not the checkerboard context. However, this observation in conjunction with the control data suggests that it is worth considering the target itself as an important component of the context as we move to develop models that predict the CTF from a description of the image. Some recent models of context effects in lightness (Blakeslee & McCourt, 2001) and color (Brainard et al., 2006) indeed embody this notion.

The caveat above noted, the fact that we can account for the functional form of the CTFs with the gain–offset–exponent model means that we can simplify the overall problem of understanding how context affects lightness by asking how context affects the model parameters. We presented some initial results of this sort by asking how context affects the black point, the white point, and the gray point. Parameterizing the model in this manner has the advantage that it refers the model parameters directly to the data. One reason we chose to do so is that we do not think that the underlying gain–offset–exponent form of the model is uniquely determined by the data. If we had chosen a different parametric form for the static non-linearity and allowed one of its parameters to vary, we might have found different parameters for the gain and offset terms of the model. However, any three-parameter model that accounted for the data would lead to the same values for the black/white/gray point parameterization.

In examining how the model parameters depended on context, we did not find simple relations. Indeed, although an additive model provided a reasonable description of how the black and white points depended on the inner and outer ring characterization, the gray point dependence exhibited a significant interaction. These effects are qualitatively consistent with Gilchrist's and Adelson's notions that in spatially complex scenes, the visual system parses the image into spatially distinct frameworks/atmospheres and that lightness processing within any one of them results from an interaction of the image statistics within each region. Our results add to this the observation that the nature of this interaction cannot be purely additive. Here, the division of the contextual image into separate regions was accomplished through variation in the luminance distribution within the inner and outer rings. Further richness may arise when geometric cues that support segmentation are considered. Our data set, which we have carefully tabulated in the Supplementary materials, provides the opportunity to further refine these and other (Blakeslee & McCourt, 2001; Rudd & Zemach, 2004) models of how context affects lightness.

Acknowledgments

The authors would like to acknowledge Li Jiang, Vanessa Troiani and Elizabeth Meegas for help collecting data for versions of this experiment.

Commercial relationships: none.
Corresponding author: Sarah R. Allred.
Email: srallred@camden.rutgers.edu.
Address: Rutgers, 311 N. Fifth Street, Camden, NJ 08102, USA.

References

Adelson, E. H. (1982). The delayed rod afterimage. *Vision Research, 22,* 1313–1328.

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339–352). Cambridge, MA: MIT Press.

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

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

Bartleson, C. J., & Breneman, E. J. (1967). Brightness perception in complex fields. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 57,* 953–956.

Blakeslee, B., & McCourt, M. E. (2001). A multiscale spatial filtering account of the Wertheimer–Benary effect and the corrugated Mondrian. *Vision Research, 41,* 2487–2502.

Blakeslee, B., & McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Research, 44,* 2483–2503.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Brainard, D. H., Longere, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M., & Xiao, B. (2006). Bayesian model of human color constancy. *Journal of Vision, 6*(11):10, 1267–1281, http://www.journalofvision.org/content/6/11/10, doi:10.1167/6.11.10. [PubMed] [Article]

Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illuminant models. *Journal of Vision, 11*(5):1, 1–18, http://www.journalofvision.org/content/11/5/1, doi:10.1167/11.5.1. [PubMed] [Article]

Brainard, D. H., & Wandell, B. A. (1986, Oct). Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 3,* 1651–1661.

Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived. *Proceedings of the National Academy of Sciences of the United States of America, 86,* 9631–9635.

D'Zmura, S. B. M. (1999). Contrast gain control. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), *Color vision: From genes to perception* (pp. 369–385). Cambridge, UK: Cambridge University Press.

Fechner, G. T. (1966). *Elements of psychophysics, Henry Holt edition in psychology*. New York: Holt, Rinehart and Winston.

Geisler, W. S. (1978). Adaptation, afterimages and cone saturation. *Vision Research, 18,* 279–289.

Gilchrist, A. (2006). *Seeing black and white*. Oxford, UK: Oxford University Press.

Heckaman, R. L., & Fairchild, M. D. (2009). Jones and Condit redux in high dynamic range and color. In *Seventeenth color imaging conference: Color science and engineering systems, technologies and applications* (pp. 8–14). Albequerque, NM: Society for Imaging Science and Technology.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9,* 181–197.

Heinemann, E. G. (1955). Simultaneous brightness induction as a function of inducing- and test-field luminances. *Journal of Experimental Psychology, 50,* 89–96.

Heinemann, E. G. (1961). The relation of apparent brightness to the threshold for differences in luminance. *Journal Experimental Psychology: Human Perception and Performance, 61,* 389–399.

Hillis, J. M., & Brainard, D. H. (2007a). Distinct mechanisms mediate visual detection and identification. *Current Biology, 17,* 1714–1719.

Hillis, J. M., & Brainard, D. H. (2007b). Do common mechanisms of adaptation mediate color discrimination and appearance? Contrast adaptation. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 24,* 2122–2133.

Hong, S. W., & Shevell, S. K. (2004). Brightness contrast and assimilation from patterned inducing backgrounds. *Vision Research, 44,* 35–43.

Hood, D. C., & Finkelstein, M. A. (1986). Sensitivity to light. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Sensory processes and perception* (pp. 5-1–5-66). New York, NY: Wiley.

Jameson, D., & Hurvich, L. M. (1972). Color adaptation: Sensitivity, contrast, and afterimages. In D. Jameson & L. M. Hurvich (Eds.), *Handbook of sensory physiology* (pp. 568–581). New York: Springer.

Land, E. H. (1986). Recent advances in retinex theory. *Vision Research, 26,* 7–21.

Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 61,* 1–11.

Logvinenko, A. D., & Maloney, L. T. (2006). The proximity structure of achromatic surface colors and the impossibility of asymmetric lightness matching. *Perception & Psychophysics, 68,* 76–83.

McCann, J. J. (2006). Measuring constancy of contrast targets in different luminances in complex scenes. In *Proceedings of IS&T/SID Colour Imaging Conference, IS&T/SID* (pp. 297–303). Scottsdale, AZ.

Mury, A. A., Pont, S. C., & Koenderink, J. J. (2009). Structure of light fields in natural scenes. *Applied Optical, 48,* 5386–5395.

Newhall, S. M., Nickerson, D., & Judd, D. B. (1943, Jul). Final report of the O.S.A. Subcommittee on the spacing of the Munsell colors. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 33,* 385–411.

Radonjić, A., Allred, S. R., Gilchrist, A., & Brainard, D. H. (2011). The dynamic range of human lightness perception. *Current Biology, 21,* 1931–1936.

Rudd, M. E., & Zemach, I. K. (2004). Quantitative properties of achromatic color induction: An edge integration analysis. *Vision Research, 44,* 971–981.

Schirillo, J. A. (1999). Surround articulation: II. Lightness judgments. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 16,* 804–811.

Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., et al. (2004, August). High dynamic range display systems. *ACM Transactions on Graphics, 23,* 760–768.

Shevell, S. K. (1978). The dual role of chromatic backgrounds in color perception. *Vision Research, 18,* 1649–1661.

Spehar, B., Debonet, J. S., & Zaidi, Q. (1996). Brightness induction from uniform and complex surrounds: A general model. *Vision Research, 36,* 1893–906.

Stiles, W. S. (1967). Mechanism concepts in colour theory. *Journal of the Colour Group, 11,* 106–123.

Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal Experimental Psychology: Human Perception and Performance, 38,* 310–324.

Wallach, H. (1976). *On perception*. New York: Quadrangle/New York Times Book Co.

Walraven, J. (1976). Discounting the background—The missing link in the explanation of chromatic induction. *Vision Research, 16,* 289–295.

Xiao, F., DiCarlo, J., Catrysse, P., & Wandell, B. (2002). High dynamic range imaging of natural scenes. In *Tenth color imaging conference: Color science, systems, and applications* (pp. 337–342). Springfield, VA: IS&T: The Society for Imaging Science and Technology.