Linking luminance and lightness by global contrast normalization

Katharina Zeiner

Modeling of Cognitive Processes Group, Department of Software Engineering and Theoretical Computer Science, Technische Universität Berlin, Berlin, Germany

Marianne Maertens

Modeling of Cognitive Processes Group, Department of Software Engineering and Theoretical Computer Science, Technische Universität Berlin, Berlin, Germany

In the present experiment we addressed the question of how the visual system determines surface lightness from luminances in the retinal image. We measured the perceived lightness of target surfaces that were embedded in custom-made checkerboards. The checkerboards consisted of 10 by 10 checks of 10 different reflectance values that were arranged randomly across the board. They were rendered under six viewing conditions including plain view, with a shadow-casting cylinder, or with one of four different transparent media covering part of the board. For each reflectance we measured its corresponding luminance in the different viewing conditions. We then assessed the lightness matches of four observers for each of the reflectances in the different viewing conditions. We derived predictions of perceived lightness based on local luminance, Michelson contrast, edge integration, anchoring theory, and a normalized Michelson contrast measure. The normalized contrast measure was the best predictor of surface lightness and was almost as good as the actual reflectance values. The normalized contrast measure combines a local computation of Michelson contrast with a region-based normalization of contrast ranges with respect to the contrast range in plain view. How the segregation of image regions is accomplished remains to be elucidated.

Introduction

Human observers are proficient at judging the apparent lightness of an object’s surface. This is a remarkable accomplishment of the visual system, because the sensory information on which this perceptual judgment is based is ambiguous. The retinal image can be thought of as a two-dimensional (2-D) array of intensity values that represent the luminances that are reflected from all locations in the observed scene. The problem with that image is that it is locally ambiguous with respect to its sources, because the amount of light that is reflected from surfaces depends on both the surface’s reflectance and the illumination incident on the surface. How the human visual system computes the lightness (perceived reflectance) of objects from intensity variations between retinal image regions is still elusive. It is undisputed that the perceived lightness of a retinal image region is modulated by its context, but whether surface lightness is computed early in visual processing, based on signals at luminance borders (Blakeslee & McCourt, 2004; Land & McCann, 1971; Rudd, 2010), or whether the lightness computation involves a more elaborate analysis of the scene with respect to other properties such as depth relationships, and/or frameworks of illumination (Adelson, 1993; Allred & Brainard, 2013; Anderson & Winawer, 2005; Gilchrist et al., 1999; Knill & Kersten, 1991) is still under debate.

The theoretical disagreement is accompanied, if not partially caused, by a diversity of stimulus arrangements that have been used in the study of lightness perception. On the one end of the spectrum are geometrically simple stimuli such as the classical simultaneous brightness contrast display. Here, two equiluminant regions are embedded in two background regions of different luminance, and the region on the darker background appears to be relatively lighter than the region on the lighter background (e.g., Wallach, 1948, or see Gilchrist, 2006 for a historical overview).
The stimulus is photometrically simple, because it consists of only three luminance values, and it is geometrically simple, because all regions are in the same depth plane and oriented fronto-parallel to the observer. On the other extreme, stimuli comprised real scenes (Gilchrist, 1977, 1980; Ripamonti et al., 2004) or rendered versions of them (e.g., Boyaci, Doerschner, Snyder, & Maloney, 2006; Kitazaki, Kobiki, & Maloney, 2008). These stimuli are photometrically and geometrically complex, because they span a large range of luminance values and the depicted surfaces were at different depths and different orientations with respect to each other and with respect to a hypothetical light source and the observer. Evidently, it is difficult to relate results that were obtained under such diverse test conditions to each other, or decide about their consistency.

Here, we follow the approach adopted by Allred, Radonjic, Gilchrist, and Brainard (2012) who used checkerboard patterns, which were of intermediate complexity with respect to the luminance relations in the stimulus, to study the effect of different contexts on the perceived lightness of a test patch. Allred et al. (2012) measured what they called “context transfer functions” to describe the perceived lightness of patches in different contexts. They reported the typical simultaneous contrast effect, that is, an increase or decrease in surround luminance caused a change in perceived target lightness that was opposite in sign to that of the surround. They also found a larger effect of the near context on perceived target lightness than of the far context. In order to model the full range of observed context transfer functions, the authors implemented an extended version of a gain-offset model that included an additional exponent parameter that varied with context (Allred et al., 2012). An important feature of the stimuli used by Allred et al. (2012) was that the tested luminances spanned a rather large range (from 0.24 to 211 cd/m²). The authors suggested that it was this feature of their stimuli which allowed the rejection of contrast-coding and gain–offset models, and which in general provided a fuller picture of how lightness varies with luminance across contexts.

In the present study we also used checkerboard stimuli, but we rendered them so that they were depictions of three-dimensional (3-D) checkerboards. Thus, in addition to photometric variations of the contexts our stimuli also contained geometric cues to depth and to differences in illumination. Our aim was to characterize the luminance to lightness mapping across contexts in the presence of photometric and geometric cues to scene segmentation, which are present also in natural viewing situations. We rendered 3-D checkerboard stimuli that consisted of randomly arranged checks with 10 different reflectance values. To manipulate the context in which the checks were viewed, a partial region of the checkerboard was either obscured by a shadow or by a transparent medium that varied in surface reflectance and transmittance (Figure 1). Since with rendering the reflectances of the checks’ surfaces and not their luminances are specified, we first had to measure what Adelson (2000) called the atmospheric transfer functions (ATFs), which describe the mapping from surface reflectance to luminance in different contexts. We also measured these functions for a real checkerboard that was crafted out of gray
papers for a number of transparent media and a shadow scenario in order to compare these to the rendering results. For the matching data of the human observers we determined in a first step what Adelson called the lightness transfer functions (LTF), which describe the luminance to lightness mapping in different contexts, and which, if the system were perfectly lightness constant, would be the inverse of the ATFs. In a second step we tested a number of different algorithms that were proposed for the computation of lightness from retinal luminance and compared their performance in predicting lightness matches across contexts.

**Methods**

**Observers**

Four naive observers (two males; age range between 24 and 34) participated in the study. All observers had normal or corrected-to-normal visual ability. Observers participated voluntarily and were reimbursed for their attendance.

**Stimuli and apparatus**

Stimuli were presented on a linearized 21-in. Siemens SMM21106LS monitor (400 × 300 mm, 1024 × 766 px, 130 Hz) controlled by a DataPixx toolbox (Vpixx Technologies, Inc., Saint-Bruno, QC, Canada) and custom presentation software (https://github.com/TUBvision/hrl). The maximum luminance this setup can produce is about 550 cd/m².

Stimuli were 2-D perspective projections of a custom checkerboard rendered with Povray (Persistence of Vision Raytracer Pty. Ltd., Williamstown, Victoria, Australia, 2004). The checkerboard consisted of 100 checks (10 × 10) with 10 different surface reflectance values that were randomly arranged across the board except for the target location which was specified by the design. With rendered images the experimenter does not directly control the pixel intensities (gray values) in the rendered images, but only specifies the desired reflectance values in the description file. The 10 different reflectance values for the checks were chosen so that they would result roughly in 10 perceptually equidistant luminances on the experimental monitor.

The checkerboard was rendered in six viewing contexts: in plain view (Figure 1A), with a shadow casting object placed on top of it (Figure 1B), and with one of four different types of transparent media overlapping part of the checkerboard (Figure 1C through F). The checkerboard, light source, and camera positions were identical across contexts. In the shadow condition a cylinder was located on the right side along the horizontal diagonal of the checkerboard, and the shadowed region covered about 17 out of the 100 checks. In the transparency conditions a square-shaped transparent layer was placed between the checkerboard and the camera. It covered an area of about 48 of the 100 checks. Four different transparent media were used resulting from a combination of two reflectance values of the transparent medium (light and dark) and two transmittance values (high and low). The luminance values of the checks at the transparent region were predicted according to the equation $L_i = \frac{a}{1 - \alpha} \times L_{C_i} + \left(1 - \alpha \right) \times L_T$, whereby $L_i$ is the luminance of the $i$-th check seen through the medium, $\alpha$ is the transmittance of that medium, $L_T$ is the luminance of the medium when rendered as an opaque surface, and $L_{C_i}$ is the luminance of the checks when seen in plain view. We used values of $\alpha = 0.2$ for the low and $\alpha = 0.4$ for the high transmittance manipulation. The luminance values of the transparent medium when rendered as opaque ($\alpha = 0$) were 158.4 cd/m² for the light and 22.6 cd/m² for the dark transparency.

Povray description files for each stimulus were generated automatically with Python (Python Software Foundation, Delaware), and Povray was called from within Python to create the images. The images were subsequently converted to a gray-scale matrix and normalized to contain values between 0 and 1. All stimuli were created prior to the experiments and loaded later for presentation. Images were 19.4º wide and 4.6º visual angle high. The checkerboard was rotated by 45º so that one of its major diagonals was parallel to the $x$-axis and the other extended in an imaginary z-axis (depth). Individual checks had an edge length of about 1º visual angle. A comparison field (1.3 × 1.3") was presented on a local checkerboard background (3.6º × 3.6º) that was randomly generated from trial to trial but constant during the adjustment within a trial. The intensities of the single checks in the comparison checkerboard were drawn from 20 equally spaced luminance values between 11 and 473 cd/m². Observers were seated 80 cm away from the screen, and their head position was fixed with a chin-rest. Responses were recorded with a ResponsePixx button-box (VPixx Technologies, Inc., Saint-Bruno, QC, Canada). The experimental setup was located in an experimental cabin that was dark except for the light emitted by the monitor.

**Design**

The main variable of interest was the viewing context in which the checkerboard was presented. Six different viewing contexts were realized: plain view, with a...
shadow casting cylinder, and with four types of transparent media (Figure 1), a dark transparency with high or low transmittance, and a light transparency with high or low transmittance. Each observer judged each of the 10 check reflectance values in all six different contexts.

Procedure

Adjustments

Observers made 12 adjustments for each of the 10 reflectance values in each of the six viewing conditions. The six viewing contexts were randomized across trials. If the checkerboard rows were denoted a–j and the columns were denoted 1–10, then the three checks of interest were checks E2, F2, and F3. The respective rows and columns of interest were labeled with black letters and numbers during the experiment as indicated in Figure 2. These checks were chosen because they were located within the region of the checkerboard that was covered by the shadow or the transparency. The judgments for each of the three checks of interest, E2, F2, and F3 were performed consecutively for each stimulus, so that one adjustment was made for each of the three checks of interest within the same checkerboard. We chose to do it that way in order to encourage observers to consider the lightness of the checks in their respective viewing context. The labels of the relevant check were colored white so that the observer knew which check was currently to be judged.

The labels were shown outside the checkerboard, and hence, it is unlikely that they exerted any effect on perceived check lightness. Observers adjusted the perceived lightness of a comparison field, presented above the checkerboard on its own local checkerboard background, by pressing one of four buttons. Two of the buttons decreased or increased the comparison intensity by 5%, and the other two buttons by 0.5%. Observers indicated a satisfactory match and initiated the next trial by pressing a fifth button at the center of the response box. Observers performed all trials within one session. Their progress was indicated at the bottom right of the display.

Luminance measurements

The stimulus luminances were measured using a Konica Minolta LS-100 spot luminance photometer (Osaka, Japan), which in combination with a 122 close-up lens allows measurements of spot sizes down to 3.2 mm. The photometer was mounted on a tripod (Manfrotto, Cassola, Italy) and was connected via a serial port (RS-232C) to the presentation software. The measurements were performed in a semiautomatic way. The photometer was manually focused on each of the three checks of interest (E3, F2, F3). Then the 10 different reflectance values were presented for each of the six contexts while the photometer automatically read the emitted luminance values that were saved in a file. Each measurement was repeated 12 times. The highest standard error of the mean was 2.8 cd/m² for the highest luminance in the plain condition (398 cd/ m²). In plain view the luminance range we studied corresponded with the range of luminances measured in natural scenes (Laughlin, 1981).

Real checkerboard

To validate the reflectance to lightness mappings (ATFs) in different contexts as they were produced by the rendering software, we also crafted a real checkerboard (Figure 3). This checkerboard was built of a metal plate to which eight by eight squares (side length) were flexibly attached with magnets. The surfaces of the 64 checks were Color-aid papers of nine different reflectances that were randomly arranged on the board. The board was placed on a black table and it was illuminated by two point light sources. One lamp was located above the checkerboard, the other one on the side, in order to generate the shadow. The light sources were adjusted so as to produce approximately the same range of luminance values in plain view that was measured for the rendered board on the monitor. One lamp was a standard desktop lamp and the shadow-
casting lamp was a spotlight (Source Four® jr 50°, ETC, Middleton, WI). We measured the ATF for plain view, for a shadow, and for three transparencies (light transp: Rosco E-Color 1/16th white; dark transp low and dark transp high: Rosco Cinegel N.6, N.3, respectively; Rosco, Stamford, CT).

Results

Atmospheric transfer functions

Figure 4 depicts the ATFs that were measured for the rendered and for the real checkerboard. The ATF in plain view shows the largest luminance range that was realized in the present set of stimuli and can be regarded as a reference for the luminance ranges in the other conditions. All context manipulations caused a reduction in the range of luminances that were reflected from the checks as indicated by the shallower slopes of all transfer functions relative to plain view. The shadow and the plain view condition differed only in the slope of the ATF, whereas the transparent media differed in both intercept and slope from each other and from plain view (Table 1).

The ATFs that were measured for the real checkerboard show similar effects for introducing a shadow or a transparent medium. The ATFs in the shadow and the dark transparency with low transmittance were almost identical and the dark transparencies with high and low transmittance differed only in slope (3 vs. 1 cd/m²). The dark and light transparencies differed most markedly in intercept (7 vs. 61 cd/m²). The numerical values are of different order of magnitude because the range of reflectance values for the real surfaces was 3% to 73% whereas the range of reflectances specified in the rendering script was 0 to 2.2.

Lightness transfer functions

The LTF depicts the relationship between luminance and perceived lightness, and should theoretically be an inverse mapping of the ATF to allow a veridical representation of surface reflectances in the perceptual domain. Figure 5 shows the luminance matches of the four observers from all viewing contexts. The pattern of results is similar across observers and the variability of individual data points (as indicated by the error bars) is generally smaller than the difference between neighboring data points. The lower panel of Figure 5 shows the linear regressions to the LTF data in different contexts together with the linear predictions derived from inverting the intercept and slope parameter from the ATFs (dashed lines in Figure 5).
for the plain view conditions, the slope of the predicted LTFs was always steeper than the slope for the measured LTF. The opposite was true in the plain view condition, except for observer “if” where the functions lay on top of each other. This means that observers’ perceived lightness values roughly corresponded to the actual underlying reflectances, but in particular for the higher luminances the perceived lightness of the checks was systematically smaller for checks presented in contexts other than plain view. This can be seen more clearly in a direct reflectance versus lightness plot which summarizes the ATF and LTF (Figure 6).

for the plain view conditions, the slope of the predicted LTFs was always steeper than the slope for the measured LTF. The opposite was true in the plain view condition, except for observer “if” where the functions lay on top of each other. This means that observers’ perceived lightness values roughly corresponded to the actual underlying reflectances, but in particular for the higher luminances the perceived lightness of the checks was systematically smaller for checks presented in contexts other than plain view. This can be seen more clearly in a direct reflectance versus lightness plot which summarizes the ATF and LTF (Figure 6).
If one assumes that the goal of the visual system is to accurately represent perceptually the differences in properties of external stimuli than one could think of the reflectance values as the best possible predictor of perceived surface lightness. Since observers matched the lightness of the same 10 checks, and thus the same 10 reflectances in all six contexts, a veridical representation of reflectance would mean that the reflectance-to-lightness mapping functions from different contexts lie all on top of each other. In order to quantify the extent to which the data are consistent with a single underlying function we calculated what we called the global $R^2$ (Maertens & Shapley, 2013). For the global $R^2$ we fit a single linear function to the data from all different viewing contexts (global fit) and then computed the standard coefficient of determination. The global $R^2$ for the reflectance-to-lightness mapping functions was on average 0.91 (range from 0.89 to 0.94), whereas it was only 0.63 on average for the LTFs (range from 0.59 to 0.66). If we think of reflectance as the best predictor of perceived lightness, and of the local luminance as a poor predictor of it, then the question is whether we can improve the prediction of perceived lightness based on the proximal stimulus, i.e., retinal luminance, by incorporating other variables in the computation.

In the following we will present different hypothetical computations that have been proposed for the computation of lightness. They involve a consideration of luminances in the surround of the target surface whereby the surround can be of different spatial extent or geometrical complexity. They will be implemented by transforming the $x$-axis according to the proposed computation and by comparing the resulting global $R^2$ values. In the final section of the results we computed the condition-wise $R^2$ in order to evaluate the models’ performance separately in our different stimulus conditions. Parts of the data involving the predictions based on luminance and on Michelson contrast for the plain, shadow, and one of the transparent contexts (light transparency with high transmission) have been reported in a previous article (Maertens & Shapley, 2013).

**Lightness based on Michelson contrast**

One computation that has been suggested to stabilize the perception of surface reflectance against changes in illumination is the contrast (Shapley & Enroth-Cugell, 1984). Here we computed a version of the Rayleigh or Michelson contrast which is defined as $(L_{\text{target}} - L_{\text{surround}})/(L_{\text{target}} + L_{\text{surround}})$, whereby the surround luminance was the mean luminance of the eight checks surrounding the target check. We have previously shown that a surround computation that was based on all eight adjacent checks, instead of the four checks that shared an edge with the target check, resulted in a better prediction of target lightness (Maertens & Shapley, 2013). The transfer functions based on Michelson contrast are depicted in Figure 7.

In comparison with the luminance transfer functions (Figure 5) the contrast-to-lightness mapping functions in different contexts are more similar to each other. The contrast computation can be interpreted as expressing the intensity of the target check as a signed deviation from the average luminance in the surround. Since the luminances were roughly equally spaced around some mean, the computed contrast values were centered at zero. It is evident from Figure 7 that the effect of different viewing contexts was a reduction in contrast range. The smallest reduction was observed for the shadow, where the contrast range was 1.52 (min = −0.95, max = 0.57) compared to 1.75 in plain view (min = −0.9, max = 0.85). For the transparency conditions the contrast range reductions were markedly larger. The smallest contrast ranges were realized in the light transparency condition with a range of 0.35 (min =
Lightness based on edge integration

A more sophisticated algorithm for the computation of lightness has been proposed by Rudd (2013). In his edge integration theory, lightness is computed by first computing local differences in log luminance across borders and then integrating these differences across space. The model was shown to predict lightness judgments for disc-annulus stimuli in which the lightness of a target disc is computed based on the weighted integration of the log difference between the target and its local surrounding annulus, and the log difference of the surround annulus and the global background:

\[ \Phi_D = w_N \times \log \frac{D}{A} + w_F \times \log \frac{A}{B}, \]

whereby \( D, A, \) and \( B \) denote the disc, annulus, and background luminances, respectively, and \( \Phi_D \) is the predicted disc lightness. Edge weights fall off with distance to the target edge, so the weight that is applied to the far (annulus-background) border, \( w_F, \) is smaller than the weight that is applied to the near (disc-annulus) border, \( w_N. \) In addition, the magnitude of the weights depends on the sign of the luminance difference (Rudd, 2013). For decremental targets the weight associated with the far edge is about 30% as large as the weight associated with the near edge. For increments the weight associated the far edge is about 79% as large as the weight associated with the near edge.

The present checkerboards are geometrically and photometrically more complex than the disc-annulus display so it is not obvious how to apply the proposed model equation to the present context. Presumably the path of edge integration originates from the background and terminates at the target surface, but how a particular path is determined or how all possible paths are integrated was not clear to us. Furthermore according to the model, an ideal edge-integrating observer would exclude illumination or atmospheric edges from the edge integration in order to relate the surface reflectances within an image to one another. But feeding directly the reflectance values into the model would sidestep the actual problem of how reflectance edges are distinguished from illumination edges when all that is given are the retinal luminances.

In order to apply the edge integration algorithm to the present stimuli we assigned the transition from the shadow, or the transparent medium, to the checkerboard in plain view as the far edge, and the edge between the target check and its eight adjacent checks as the near edge. The distance of the near edge was identical for all viewing contexts, and we made the simplifying assumption that the distance of the far edge was also identical (which is exactly true only for the four transparent media). We also assumed identity of the contrast polarity of the far edge across all viewing contexts. This was true for the dark transparencies and the shadow, because their contrast polarity relative to plain view was decremental (mean luminances for plain, shadow, dark transparency with high and low transmittance were 150, 47, 73, and 47, respectively). The mean luminances for the light transparencies, however, were almost identical to, or in fact slight increments relative to plain view (mean luminances for light transparency with high and low transmittance were 157 and 159, respectively). The contrast polarity of the near edge was determined by the log luminance ratio of the target and its surround. To compute lightness for the present stimuli we substituted the terms in Equation 1 in the following way: In \( \log (D/A) \) and \( \log(A/B), \) \( D \) was the target luminance, \( A \) was the mean luminance of the eight adjacent checks, and \( B \) was the mean luminance of the checkerboard in plain view, as this was the identical background for all viewing contexts. \( w_F \) was set to 1 and \( w_N \) was set to 1/0.3 = 3.33 for targets that were decrements relative to the mean luminance of the eight surrounding checks, and to 1/0.79 = 1.26 for targets that were increments relative to their near surround. The functions relating match luminance to lightness predicted by edge integration theory are depicted in Figure 8. Match luminances were log scaled in order to compute the global \( R^2 \) for the linear prediction of match luminance from predicted lightness. The average global \( R^2 \) was 0.70 (range 0.54 to 0.85).

Lightness based on anchoring theory

According to the anchoring theory of lightness perception (Gilchrist et al., 1999) the brightest region in an image appears white and the lightness of every other image region is determined as the ratio of its own luminance \( l_i \) relative to the highest luminance \( l_{\text{max}}. \) So, in simple images the lightness, \( R_i \), of an image region is specified as \( R_i = [(l_i)/(l_{\text{max}})] \times 0.9. \) In more complex
images the correct prediction of lightness becomes more difficult, because different image regions might be seen at different levels of illumination. In order to take these illumination differences into account, it was suggested that the retinal image is segmented into frameworks, that is, into regions of common illumination (Gilchrist et al., 1999; Radonjic & Gilchrist, 2010). In complex images, each image region then belongs to at least two frameworks, a global framework that comprises the entire scene, and a local framework that results from the local viewing context of the image region. The predicted lightness value for the image region of interest is a weighted average of the relative luminance computed within the local and the global framework,

\[ R_i = w_{local} \times \frac{l_i}{l_{max,local}} \times 0.9 \]
\[ + (1 - w_{local}) \times \frac{l_i}{l_{max,global}} \times 0.9. \] (2)

The weight \( w_{local} \) depends on the strength of framework segregation, which depends on the presence of grouping factors, but how exactly grouping works is not well understood. There is an additional computational step suggested by anchoring theory, which is a scale normalization in order to expand (or compress) the range of resulting lightness values to the canonical range of 30:1 (Gilchrist, 2006, p. 295). The range adjustment is done according to the following formula

\[ L_{i,adj} = (1 + (0.56 \times (\log(30) - \log(range(L_i)))) \times L_i \] (3)

whereby \( L_{i,adj} \) are the so-adjusted lightness values, \( \log(30) \) corresponds to the canonical range of reflectances, \( range(L_i) \) is the range of lightnesses resulting from the anchoring computation described above, and \( L_i \) is one of those lightnesses that needs to be range-adjusted. In order to test the prediction of anchoring theory for the present data set, we transformed the target luminances in each viewing context according to the above-mentioned formulas. Figure 9 depicts match luminance as a function of lightness values predicted based on anchoring theory. The data are identical to the data in Figure 5 except that now the local check luminances were transformed according to the formulas proposed by anchoring theory (see text). The solid lines are linear regressions that were fitted to the mean matches of each observer in each viewing context.
luminances as a function of lightness values predicted by anchoring theory.

The data in Figure 9 depict the prediction for an equal weighting of the local and global frameworks \((w_{local} = 1 - w_{local} = 0.5)\). This resulted in a mean \(R^2\) value of 0.72 (range from 0.68 to 0.77). The highest value for the global \(R^2\) resulted from a unique weighting of the local framework \((w_{local} = 1)\) with 0.74 (range from 0.73 to 0.77). However, we decided in favor of the equal weighting, because the difference in \(R^2\) is small, and as will be evident from the model comparison section, with an equal weighting the model yielded reasonable lightness predictions for surfaces in the shadow. Since the model was originally conceived of to deal with illumination differences and not with the presence of transparent media, we found it fair to use parameters so as to allow optimal model performance with respect to that goal. The anchoring prediction became worse for local weights that were smaller than 0.5 and a unique weight of the global framework resulted in the worst performance (average \(R^2 = 0.52\)). The present predictions for anchoring theory and edge integration theory are seriously limited, because both models were designed in the realm of simpler scenes, and the present predictions had to rely on parameters that were empirically derived in different experiments. Yet, we consider the attempt to apply the different models to the present stimulus fruitful, because it might provide useful hints for future model development.

**Lightness based on normalized contrasts**

So far, the Michelson contrast, which has no parameters or assumptions, resulted in the best global prediction of lightness across different viewing contexts. However, due to the reduced range of contrast values in different contexts, a prediction on contrast alone has limited power. To improve the contrast-based lightness prediction we considered the following findings. The Michelson contrast has been demonstrated to be used by the visual system to initiate scission and to construct perceptual transparency (Singh & Anderson, 2002). Furthermore, according to the transmittance anchoring principle, the visual system exploits contrast to determine which parts of an image are unobscured by transparent layers. Image regions of highest contrast are seen in plain view (Anderson, 1999). Taking these different pieces of evidence together we devised the normalized contrast measure, which was computed in the following way. First, the
Michelson contrasts are computed across the image. Second, the image is divided into regions of different contrast ranges. This image segmentation step might be based on photometric and/or geometric cues, such as regional darkening (Singh & Anderson, 2006), or differences in depth (Gilchrist, 1977). Finally, the Michelson contrasts within each region are normalized relative to the region with the highest contrast range, namely the image region that is seen in plain view. The prediction of match luminance based on this normalized contrast measure is shown in Figure 10. The global $R^2$ was on average 0.88 (range from 0.86 to 0.91). These values are of the same order of magnitude as the $R^2$ values for the match luminance predictions based on reflectance. Also the pattern of functions relating normalized contrast and match lightness is similar to the one observed for reflectance (see Figure 6). The lightness predictions in all contexts underestimate the luminance matches at higher contrast values relative to plain view.

**Model comparison**

In order to evaluate the respective model performance separately for the different viewing contexts, we computed a condition-wise $R^2$ equivalent. Figure 11 depicts this measure, which is defined as the proportion of residual variance to the total variance in each condition. The residual variance is the sum of squared differences between the actual data in each condition and the predicted data based on the global fit. Here, small values are desirable, because then most of the variability is already explained by the global fit, and hence the residuals are small. The scaled contrast model resulted in the smallest residuals across all conditions. In particular, it outperformed the simple Michelson contrast prediction for the light transparency conditions, for which the largest reduction in contrast range was observed. Anchoring theory performed comparably well for the shadow and the dark transparency condition with the high transmittance. These two conditions were rather similar in contrast and luminance range. The edge integration prediction for the shadow was worse than that of the other models whereas the prediction for the dark transparency with high transmittance was comparable to that of the others. As for anchoring theory the largest prediction errors were observed for the light transparencies. While Figure 11 provides a reasonable summary for the present analyses, it is just one means to assess how well a certain computation predicts perceived lightness. This measure was based on the idea that a putative lightness computation is good when it transforms the input data (luminances) so that the predicted lightness values are consistent with a single function. An alternative approach could be to compare the models with respect to the degree to which the predicted lightness values resemble the lightness predictions based on reflectance, because reflectance could be regarded as the best predictor of lightness.

**Discussion**

In the present experiment we measured the perceived lightness of image regions in relatively naturalistic checkerboard stimuli. We tried to relate perceived lightness to retinal luminance, which is the local input signal to the eye. We measured the perceived lightness of target surfaces in different viewing contexts including plain view, a shadowed region and through four different transparent layers (Figure 1). Although we used rendered versions of the stimuli we made an attempt to ensure that the photometric effects of introducing different viewing contexts were comparable to those in the real world (Figure 4; Koenderink, 1999). We tested a number of models that have been suggested for the computation of lightness and compared their respective ability to unify the variability of lightness matchings in different contexts by a common underlying model (Figures 5 through 10).

The computations differed with respect to the spatial extent to which they considered luminance variations in the surround of the target surface. The most local measure was the prediction of lightness based on local luminance alone, that is, the so-called LTF (Figure 5; Adelson, 2000). In terms of explaining the data by a single underlying model, this measure was the worst as it explained only about 63% of the variability across contexts. All other computations considered to different degrees the luminance values in the surround of the target. For the Michelson contrast measure (Figure 7), the luminances of the eight checks directly adjacent to the target were averaged and the target luminance was expressed as a signed deviation relative to the surround. Being a rather local image-based computation, this measure worked rather well as it explained 81% of the between-context variability in lightness matches. Admittedly, the measure was not purely local, because with eight out of 10 possible luminance values sampled for the surround term, it is a rather good estimate of the average luminance in that viewing context. We also used Rudd’s (2013) edge integration model to predict lightness based on an extended surround. We interpreted the model so as to include the checks adjacent to the target in the surround term, and the checks seen in plain view, outside the shadow or transparent layer, as the background term. The so-derived prediction was not much better than the local luminance in terms of reducing variability between contexts (Figure 8, 64%...
variability explained). However, we had to make many simplifying assumptions in order to derive a prediction for the present stimuli, and this might not have done justice to the model. We also derived predictions for target lightness based on Gilchrist et al.’s (1999) anchoring theory. Lightness predictions based on that model were slightly better as they explained 74% of the variability across contexts (Figure 9). Interestingly, the best prediction of lightness based on anchoring theory was accomplished when the weight for the global framework was set to zero, and hence only the local framework was considered. Again, an interpretation of the predictions based on anchoring and edge integration theory is difficult because we simply utilized numerical parameters that were derived with very different experimental stimuli. It remains a task for the future to specify how a respective algorithm can be applied to a wider range of stimuli.

The model that accounted best for the variability in lightness matches across viewing contexts was a normalized version of the Michelson contrast (Figure 10). The justification for the normalization was based on the following reasoning. The luminance contrast of surfaces in an image is reduced when the surfaces are seen through a transparent layer relative to when they are seen in plain view. This contrast reduction has been shown to be used as a cue to initiate a perceptual layer separation at the corresponding image locations (Singh & Anderson, 2002). It might thus not be too far-fetched to assume that the visual system has a way of telling the input regions apart from each other based on their respective between-region differences in contrast ranges. In addition, the visual system seems capable of identifying the image region with the highest contrast range and assigning the perceptual attribute of “seen in plain view” (Anderson, 1999). For the normalized contrast measure we made one additional assumption, which is that in order to assign lightness values, the visual system scales the contrast range within each nonplain region so as to match the contrast range in plain view. This measure explained 88% of the variability in lightness matches across contexts. This predictive power was of similar magnitude as that of the reflectance-based prediction, which accounted for 91% of the variability across contexts. The reflectance-based prediction (Figure 6) served as an upper limit for the expected goodness of the prediction of the tested models. This is because we assume that the perceptual representation of surface reflectance (lightness) should resemble actual surface reflectances in order to provide the human visual system with reliable information about the environment. The pattern of results for the predictions based on normalized contrasts and reflectances were also qualitatively similar, because there was a systematic underestimation of lightness for the higher reflectances/contrasts in all contexts relative to plain view. It remains an open question whether this underestimation reflects a true deviation from lightness constancy or whether it was due to the 2-D depiction of the surfaces, and hence a stronger perceptual influence of their actual luminance values.

The importance of luminance contrast for the perception of lightness has often been suggested (Shapley & Enroth-Cugell, 1984; Wallach, 1948; Whittle, 2009; Whittle & Challands, 1969). Here, we computed the Michelson contrast based on all eight surrounding checks because we have shown that this contrast measure accounts better for the variability introduced by different viewing contexts than a contrast measure based on only the four surround checks that share an edge with the target (Maertens & Shapley, 2013). A so-defined surround measure represents a good sample of all possible surround intensities and, assuming a random distribution of surface reflectances across the checkerboard, it can be regarded as an estimate of the mean illumination of a particular surround region. This is because in the absence of a reflectance change all changes in luminance should be attributed to a change in illumination. Given the simplicity of the contrast computation the measure worked reasonably well in predicting lightness from luminances across the image. However, in particular for the transparency situations this local normalization was not sufficient to predict perceived surface lightness. Based on results on transmittance anchoring and contrast-based layer separation, we argued that an additional global normalization step is required, and this step improved the prediction of lightness values so that it was close to a prediction based on reflectance itself.

We think that our proposed contrast-based algorithm for the perception of lightness relates well to the descriptive model that was suggested by Allred et al. (2012) in order to account for their data. They studied the effect of photometric context variations on the perceived lightness of a target region using coplanar checkerboards. In contrast to the present study, Allred et al. (2012) manipulated the luminance range of the surround in the absence of concomitant changes in geometric cues to shadows or transparent layers. They analyzed their data by plotting so-called context-transfer functions in which the luminance values that evoked identical lightness percepts in different contexts are plotted against each other. To account for the matching data in different contexts they had to extend a pure gain-offset model by an exponent that was allowed to vary with context (Allred et al., 2012; equations 6 and 7). We think that the gain-offset part of their model can be identified with the local computation of Michelson contrast, as that comprises a divisive (gain) and a subtractive component (offset). The region-based normalization of contrast ranges
As Whittle (2009) has pointed out, this kind of ambiguity in a stimulus poses a problem, because it requires the observer to take one or the other attitude towards the perceptual interpretation of the stimulus. He intentionally used the word “attitude” to express “subject’s bewilderment confronted with the impoverished stimuli of psychophysical experiments” (Whittle, 2009, p. 132). In the absence of disambiguating cues different observers might favor one or the other perceptual interpretation, or even a single observer might oscillate between the alternatives. The advocates of simple stimuli have argued that they were particularly apt to discover regularities in the workings of the visual system, because they were void of potential confounds that are introduced when scenes become increasingly complex. On the other hand, it is possible that in the absence of cues that are indicative of a 3-D scene interpretation, the visual system is forced into a state where it must rely solely on luminance-related cues to lightness (e.g., luminance ratios or contrasts). Then the restrictions that were imposed on the stimulus also restricted the potential insights into mechanisms of lightness perception. Another problem of simple stimuli is that they were often generated as 2-D intensity arrays, and hence, the distal stimulus that might have elicited a certain retinal image was not known. It was thus impossible to determine to what extent the perceptual representation resembled the external stimulus, because the proximal stimulus might have been ambiguous with respect to its real world sources. The problem has been vividly demonstrated in the domain of transparency perception. Singh and Anderson (2002) noted that in the contrast-contrast effect (Chubb & Solomon, 1989) an additional transparency dimension emerged from a 2-D gray scale image in addition to the intended apparent contrast effect. Similarly, Ekroll and Faul (2013) reported that observers resorted to a fourth, transparency dimension in order to match a specified target color in an asymmetric color matching experiment although the stimuli were conceived in a 3-D color space. Thus, unintended by the experimenters, the dimensionality of the perceptual space exceeded that of the stimulus space. For these reasons we think that the use of realistic scenes is preferable, but to understand the relative contributions of lower- and higher-level scene variables on lightness perception, stimulus complexity needs to be increased incrementally in order to be able to relate different empirical findings.

In the present experiment we used stimuli of moderate photometric and geometric complexity that covered a relatively naturalistic luminance range (Laughlin, 1981), and we tried to quantitatively compare the predictions of a number of models of lightness perception. We found a model that combined a local computation of Michelson contrast with a global contrast range normalization to be most
successful in predicting lightness matches. However, these model comparisons were explorative in nature as the variables that we systematically manipulated were the viewing context and the check reflectances. The normalized contrast, as well as all the other measures that were used as predictors of lightness, was derived in a post hoc fashion from the target luminances and the luminances of randomly assigned surround check reflectances. It remains a task for future experiments to put the normalized contrast computation under scrutiny by systematically manipulating the contrast range in different regions of illumination and testing the effects of varying the luminance (image) or the reflectance contrast (real world surfaces).

**Conclusion**

Allred et al. (2012) began their discussion by saying that we are currently far from having a complete model of the perception of surface lightness “that would allow (the) prediction of the lightness of any image region, given the luminance of each location in the image” (p. 12). We still do not have a complete model, but with the normalized contrast model we provide an algorithm that allows to compute lightness from retinal luminances only. The important piece of information that is still missing, and which we secretly inserted, was the knowledge about regions of different contrast range. Here we simply used the values that we knew to originate from the checks within the regions corresponding to plain view, shadow, or transparent media, but for a model to be applicable to any image this segmentation step still needs to be elucidated.

*Keywords: lightness/brightness perception, segmentation, contrast, surface reflectance*

**Acknowledgments**

This work has been supported by an Emmy-Noether research grant of the German Research Foundation to Marianne Maertens (DFG MA5127/1-1). We are grateful to Robert Shapley and Felix Wichmann for critical discussion of the experiment and for insightful comments on earlier versions of the manuscripts. We would like to thank two anonymous reviewers for insightful comments on an earlier version of the manuscript.

Commercial relationships: none.

Corresponding author: Marianne Maertens.

Email: marianne.maertens@tu-berlin.de.

Address: Modeling of Cognitive Processes Group, Department of Software Engineering and Theoretical Computer Science, Technische Universität Berlin, Berlin, Germany.

**References**


