When luminance increment thresholds depend on apparent lightness

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

Felix A. Wichmann
AG Neuronale Informationsverarbeitung, Mathematisch-Naturwissenschaftliche Fakultät, Eberhard Karls Universität, Tübingen, Germany
Bernstein Center für Computational Neuroscience, Tübingen, Germany
Max-Planck-Institut für Intelligente Systeme, Abteilung Empirische Inferenz, Tübingen, Germany

A fundamental question in visual perception research is whether the sensitivity to stimulus differences is limited by the sensory representation of the external stimulus, that is, the proximal stimulus, or by its perceptual representation, i.e., stimulus appearance. In the domain of lightness perception, the question translates into whether discrimination thresholds depend on the local luminance in the retinal image or on the apparent lightness of the corresponding image region. The majority of findings seem to indicate that sensitivity is limited by the sensory stimulus representation, which would imply different mechanisms for stimulus discrimination and appearance. We think this conclusion needs to be qualified. We report data suggesting that the relationship between discrimination and appearance judgments depends on how exactly they are being measured. We propose a theoretical account that provides a common mechanism for appearance and sensitivity. An interesting corollary of this model is that it also accounts for the perceptual phenomenon of assimilation.

Introduction

Visual psychophysics is a useful tool in the repertoire of neuroscientific methods as its results allow inferences about underlying neuronal or perceptual mechanisms. The goal of psychophysical examination is to characterize the relationship between some physical stimulus variable (x) and the corresponding psychological sensation ($\Psi[x]$, Fechner, 1860). One of the few lawful relationships between external stimulation and internal phenomenal experience is expressed in Weber’s law, which states that the just noticeable difference (JND) between two quantities increases in proportion to their overall intensity (Fechner, 1860, p. 300). Unfortunately, “intensity” might mean different things, and therefore it is not always self-evident which is the right dimension to look at. Consequently, depending on the dimension considered, the observed relationship between (x) and ($\Psi[x]$) might look more or less lawful (Maertens & Shapley, in press).

The ambiguity of the “appropriate” intensity dimension is well illustrated in the phenomenon of simultaneous brightness contrast (Hering, 1920), in which two equiluminant targets are presented on backgrounds of different intensity. The target on the darker background appears lighter than the target on the lighter background. The validness of Weber’s law could now be considered for two intensity dimensions: One is the local luminance (measured with a photometer in $\text{cd/m}^2$) of the target regions in the proximal stimulus, and the other is the perceived lightness of the target regions as measured, e.g., in an asymmetric matching procedure on a “neutral” background (also in $\text{cd/m}^2$ but of the matching field). The question that arises from that ambiguity is whether JNDs increase in proportion to the proximal stimulus intensity (luminance) or in proportion to the perceived lightness.

Assuming that perception should allow us a meaningful interaction with the environment, one would presume that discriminability between two surfaces should be limited by how accurately their lightness is...
perceived. Retinal luminance is only the vehicle by which the visual sensor makes contact with the outside world, and it can vary tremendously due to changes in illumination. For Weber’s law to be a meaningful characterization of perceptual behavior, one could argue it should apply to a stage of image encoding at which the stimulus is stabilized against illuminant changes (Hillis & Brainard, 2007). This may be the stage at which surface lightness is derived.

Even though it is intuitively appealing, at least to us, this idea has found mixed empirical support. While some studies found that discriminability was—at least to some extent—affected by apparent lightness (Henning, Millar, & Hill, 2000; McCourt & Kingdom, 1995; Krauskopf & Zaidi, 1986), others reported Weber fractions to depend solely on local luminance (Burr, 1966; Cornsweet & Teller, 1965).

Most previous studies employed rather simplified two-dimensional stimulus configurations that were lacking an obvious three-dimensional interpretation. We believe that in order to assign surface lightness to retinal image regions, it is crucial to provide the visual system with cues that are indicative of the depth and/or illumination structure of the entire scene (e.g., Anderson & Winawer, 2008; Arend, 2009; Gilchrist, 1980; Logvinenko, 2005; Purves, Shimp, & Lotto, 1999; Ripamonti et al., 2004; Snyder, Doerschner, & Maloney, 2005). In the present study, we thus tested the question of whether sensitivity is limited by apparent lightness or retinal luminance, again measuring both sensitivity and appearance in more realistic stimuli, namely a customized version of the Adelson checkerboard.

The Adelson checkerboard is designed so that the luminance values of checks A2 and B1 are identical even though check A2 looks markedly lighter than check B1 (Figure 1A). The experimental logic is the following: If discrimination performance is limited by retinal luminance, then discrimination performance measured at positions A2 and B1 should be identical because the checks are equiluminant. If, alternatively, discrimination performance is limited by the apparent lightness of checks A2 and B1, then the discrimination threshold should be higher on check A2 than on check B1 because the JND would increase in proportion to their apparent lightness. To anticipate, we found that whether or not discrimination thresholds depend on local luminance or perceived lightness depends on the type of probe with which thresholds are being measured.

Measuring discrimination performance in these more naturalistic patterns is a challenge. One question concerns the choice of the proper increment and the other the choice of the task. One could argue that, in order to test the sensitivity to differences in the check’s surface intensity, one would, of course, have to vary the intensity of the entire check’s surface (Figure 1B). The benefit of that type of increment is that a single check’s surface is an integral part of the entire checkerboard pattern, so it would probe supposedly natural conditions. However, for that type of increment, the two equiluminant test regions do not only differ in appearance, but also in their border contrast because the check in the shadow (A2) is flanked by darker and the check outside the shadow (B1) by lighter checks. Alternatively, one could superimpose an appropriate probe, e.g., an elliptical shape, on the check’s surfaces (Figure 1C). The benefit of the elliptical increment is that its local luminance difference (contrast) is identical for both checks of interest because the ellipses as well as the underlying check surfaces are equiluminant. Hence, putative differences in thresholds cannot be attributed to differences in luminance contrast. However, if the border contrast, instead of the ellipses’ lightness, is used as a cue to detect the increment, then performance differences are not to be expected, exactly because the contrast is identical. The probed mechanism would not be lightness discrimination but contrast detection. Whittle (2009) has pointed out that the guiding principle for discrimination measurements to be comparable to measurements of appearance is that both tasks must rely on the same subjective stimulus dimension and thus likely the same neuronal events (Whittle, 2009, p. 72). A third type of increment is ellipses with a blurred border or so-called “blob” increments (Figure 1D). They are often used as increments instead of proper ellipses (Hillis & Brainard, 2007) in order to avoid ceiling performance due to the human visual system’s high sensitivity to sharp contrast borders. However, blurred ellipses underlie the same restriction as proper ellipses. Because there was no a priori reason to favor one increment over another, experiments were performed with all three types of increments.

With respect to the experimental task, the choice was between a forced-choice and a so-called “yes-no” task. From a signal detection point of view, the forced-choice task has the benefit of (theoretically) being criterion-free (e.g., Green & Swets, 1966). The yes-no task, on the other hand, comes closer to natural viewing because we are not usually presented with two slightly different versions of the same scene. All experimental conditions were thus run with both versions of the task.

**Methods**

**Observers**

One of the authors (MM) and nine naive observers participated in the experiments; three of them were
male. Observers’ ages ranged from 20 to 34. All observers had normal or corrected-to-normal visual ability. All naive observers participated voluntarily and were reimbursed for their attendance.

Stimuli and apparatus

Stimuli were presented on a linearized 21-in Siemens SMM 21106 LS monitor (400 × 300 mm, 1024 × 766 pixel, 130 Hz) controlled by a Cambridge Research Systems 10-bit graphics card. The maximum luminance that the monitor can produce is 570 cd/m². In the present experiments, we used a look-up table (LUT) of 256 linearly spaced luminance values spanning a range between 28.1 cd/m² and 196.2 cd/m².

The stimulus scene was rendered once using Povray (Persistence of Vision Pty. Ltd., 2004). The resulting image was subsequently converted to a gray scale matrix and normalized to contain values between 0 and 255 using the imaging library in Python. Povray is a general-purpose rendering software that makes the depicted objects look as realistic as possible. The user specifies the desired reflectance values in the description file but obviously has no control over the resulting pixel intensities in the rendered images because they are ray-traced in order to mimic physical optics. We thus derived mapping functions between surface reflectance values (input) and pixel intensities (output) for the checks inside (Figure 1, A2) and outside the shadow (Figure 1, B1). Here we will use “in” and “out” as short labels to refer to the checks inside and outside the shadow. The mapping function was then inverted to determine the input reflectance values such that the corresponding pixel intensities were “equiluminant” at the check locations. However, although the resulting
pixel intensities were, on average, identical between the checks of interest, they were—as a result of realistic rendering—not homogeneous across a single checks’ surface. To exclude the possibility that any of our reported results were influenced by this very slight inhomogeneity, we replaced all pixels belonging to one check’s surface manually by the intensity of the central pixel. To ensure that the so-manipulated checks of interest did not noticeably differ from the original, we tested their indistinguishability in a two-interval forced choice (2IFC) procedure. The discrimination performance of one observer was at chance for the two check positions (out: 26/50, \( p = 0.89 \); in: 21/50, \( p = 0.32 \)).

The two equiluminant checks had a luminance value of 58.2 cd/m\(^2\). Increments varied between 0.01 and 0.4 (in units of Weber fractions) and were adjusted if necessary based on observers’ performance in previous sessions to allow a reasonable sampling of the psychometric function (Wichmann & Hill, 2001). All stimuli were created prior to the experiments and loaded later for presentation. Images were 520 pixels wide and 520 pixels high. The checkerboard was rotated by 45° so that one of its major diagonals was parallel to the x-axis and the other extended in an imaginary z-axis (depth). The lateral edges of the checkerboard subtended 6.5° visual angle, and individual checks had an edge length of about 1.1°.

Observers were seated 100 cm away from the screen in an experimental cabin that was dark except for the light emitted by the monitor. The background luminance was 105.7 cd/m\(^2\).

**Threshold measurements**

Discrimination thresholds were measured at two different check positions within the Adelson checkerboard and with three types of increments. The in and out checks were equiluminant but differed in surface lightness as one was the light (Figure 1, A2) check inside the shadow and the other was the dark check outside the shadow (Figure 1, B1), respectively. For the check increments, a constant value was added to the entire check area (Figure 1B). The ellipse increment had a horizontal to vertical aspect ratio of 2:1, and its horizontal main axis was 38 pixels wide (Figure 1C). The blob increment had a comparable shape but with a horizontal diameter of 44 pixels and a Gaussian border profile \((scale \text{ factor} \times \exp\{[(0.5 \times x)^2 + y^2]/\text{diameter}\}\) (Figure 1D). The blob increment was scaled by a factor of two in order to equate the volume under the blob and ellipse increments as much as possible (the volume of the blob ended up 2% larger than that of the ellipse increment). Increments corresponding to different Weber fractions were realized by varying the height of the blob or the ellipse. While the blob and check experiments were repeated with 9 of the 10 observers, the ellipse experiment was repeated with only four of the original nine and one new observer.

The fixation cross was positioned equidistant to the centers of the two checks of interest. In order to encourage global processing of the entire checkerboard pattern, the position of the checkerboard was slightly jittered relative to the fixation cross from one presentation to the next. The position of the center of the checkerboard was randomly drawn from a circle with a radius of 2° visual angle centered at the fixation position.

A 2IFC trial started with a 770-ms fixation screen, followed by two stimulus presentations, 200 ms each, which were separated by a 200 ms presentation of the fixation screen. The increment had a 50% chance of being in either of the two intervals, and observers were asked to indicate the temporal position of the increment with a button press. In each session, five different increment intensities were tested at each check position. Increment intensities were repeated 36 times and were randomly intermixed between trials. Breaks were introduced every 36 trials.

A yes-no trial was identical in timing to a 2IFC trial except that only one stimulus was presented. Trials were grouped in blocks of 50 trials that were followed by breaks. In half of the trials within a block the increment was present. Within a block, the increment intensity was constant. In each session, five different increment intensities were tested at each check position. A run of five blocks for each increment at the two check positions always started with the highest intensity, but the sequence of intensities that followed in the subsequent blocks was random.

Each observer performed between 11 and 19 sessions with the check and the blob increments and between three and five sessions with the ellipse increment. Sessions with the blob and check increments were performed in succession. Some observers started with the blob sessions and some with the check sessions. Observers performed at least three sessions a week. The ellipse experiment was performed about 8 months later when it turned out to be an important control. Within each 1-hr session, observers completed the 2IFC experiment and the yes-no experiment. It was randomly chosen whether they started with the 2IFC or the yes-no task. Each check position (Figure 1, A2 or B1) was probed in a separate block, and instructions were shown at the beginning to indicate the current check of interest.

**Appearance measurements**

In order to assess the appearance of the two checks of interest, observers were asked to adjust the intensity of a
test patch such that it seemed to have the same surface lightness as one of the respective checks. The comparison patch was 1.1° × 1.1° visual angle wide and positioned centrally above the checkerboard. Because the apparent lightness of the white check in the shadow was very similar to that of the background, the comparison patch was presented on its own local background: a square that was 3.4° × 3.4° wide and positioned 1.1° above the checkerboard. Four different luminances were employed as test surrounds: 86.7, 95.8, 116.9, or 128.9 cd/m². The values were chosen such as to be equidistant decrements and increments relative to the background (c = −0.1, −0.05, 0.05, 0.1; contrast was calculated as the difference between the square and background luminances divided by their sum). The adjusted intensities differed between these backgrounds in accordance with simultaneous contrast, but we averaged the adjustments across backgrounds because that variation was introduced only for technical reasons. Five adjustments were made for each surround, resulting in 20 trials for each block at one check position. Blocks were repeated three times. The sequence of surround luminances was randomized within a block. At the beginning of each block, an instruction screen indicated which of the two checks had to be matched. The initial luminance value of the comparison patch was chosen randomly from the 256 different LUT entries. Observers used a five-button response box to increase or decrease the patch intensity and to switch between small and large adjustment steps (1 or 10 LUT steps corresponding to 0.65 and 6.6 cd/m²). The fifth button served two purposes: Upon a single click, it triggered the presentation of the checkerboard, which was limited to 200 ms. With a double click, the observer indicated that she was satisfied with her luminance setting and triggered the next trial.

In order to assess the appearance of the elliptical increments, observers adjusted the intensity of one of the blobs or ellipses to make it look identical to the other. The test ellipse or test blob was either inside or outside the shadow, and it had to be matched by a target that was outside or inside the shadow, correspondingly. Blobs were compared with blobs and ellipses with ellipses. We used this asymmetric matching procedure instead of comparing the ellipses’ intensities to a third test patch because the apparent lightness of the Gaussian ellipses was easier to judge relative to each other. Adjustments for the blobs and ellipses inside and outside the shadow were made in separate blocks of trials. Instruction displays at the beginning of each block indicated which ellipse had to be adjusted. Observers made five adjustments for each of seven ellipse intensities (64.5, 71.0, 77.5, 84.0, 90.6, 97.1, 103.8 cd/m²), which were presented randomly within each block. Here we will consider only the second smallest ellipse intensity of 71.0 cd/m², which corresponds to a 22% increment (the other values were realized for a different study). It was thus clearly visible but still comparable to the checks’ intensity. Observers adjusted the intensity of the ellipses in the same way as in the previous experiment except that the stimulus was now constantly visible. This experiment was performed as a control experiment subsequent to the first data collection, and by the time we had confirmed that the results were identical to those from short presentations, one of the nine observers, a female, could not be recruited for the experiment.

**Psychometric functions**

To characterize discrimination behavior, psychometric functions were fitted to individual data, which relate the percentage of correct responses to increment luminance increases in Weber fractions. Weibull functions were fitted using the maximum-likelihood procedure of Wichmann and Hill (2001) implemented in psignifit. The parameter of interest was the performance level, or discrimination threshold, at which observers were correct in 87.5% of the cases (corresponding to $\Psi^{-1}(x) = 0.75$ for values that vary between 0.5 and 1.0).

**Analysis of variance**

Due to the sequence of data collection, we performed the following analysis on the discrimination data: We computed a global analysis of variance (ANOVA) on the data from the check and blob increments with the factors of increment type (blob vs. check), task (2IFC vs. yes-no), and check position (in, A2 and out, B1). In addition, we performed three separate ANOVAs for data collected with each increment with only the two factors of task and check position (see Table 1). Although statistical differences between the increments cannot completely be tested this way, the pattern of results for each increment type together with information about the consistency of observers conveyed by the figures can be compared qualitatively.

**Result**

**Appearance measurements**

Figure 2 depicts the matching data for all three types of increments. A substantial difference in appearance is evident for the checks inside and outside the shadow. The perceived lightness of the dark check outside the shadow is matched to its actual local luminance value: a one-sample t test for comparison with a luminance
value of \( 58.2 \text{ cd/m}^2 \), \( t(8) = -1.89, p = 0.10 \). The perceived lightness of the light check inside the shadow was significantly higher than the actual luminance, \( t(8) = 14.09, p < 0.001 \), and than the apparent lightness of the dark check, \( t(8) = -13.35, p < 0.001 \). An analogue appearance difference was evident for the proper ellipses. The effect seemed slightly larger, but one has to keep in mind that the ellipses were mutually matched, and hence the observed difference reflects not only the appearance of the matching target, but also that of the test. A paired \( t \) test comparing the matches between ellipses was significant, \( t(7) = -9.36, p < 0.001 \). The effect was much smaller for the blurred ellipses. The difference was significant probably due to the small variability, \( t(7) = -3.93, p = 0.006 \), but again, the difference contains both the appearance effect on the matching target and on the test.

Given the data from the matching experiments and based on the hypothesis that Weber’s law applies to the apparent lightness and not the local luminance of a region, one would expect almost no sensitivity differences for discrimination performance with blob increments and comparable sensitivity differences with the check and the ellipse increments. In particular, thresholds should be higher for increments presented inside (A2) than for increments presented outside the shadow (B1) because if discrimination is based on apparent lightness then higher discrimination thresholds should result for the lighter-looking check (A2).

The discrimination data are plotted in Figures 3 through 5, which show the psychometric functions of one observer in the yes-no (panel A) and 2IFC (panel B) tasks and the individual thresholds of all observers together with the averaged thresholds at the two check positions and in the two tasks (panel C).

A three-factorial ANOVA on the check and blob data showed that all three main effects were significant: Thresholds were higher on the check inside compared to the check outside the shadow (0.107 vs. 0.094), with the check than with the blob increment (0.121 vs. 0.080), and with the yes-no than with the 2IFC task (0.116 vs. 0.085). The effects can best be read from the group data depicted in the rightmost x-axis category in the lower panels of Figures 3 and 4.

Thresholds obtained with the blob increment differed significantly only between the two types of tasks because the yes-no task yielded, on average, higher thresholds than the 2IFC task (0.088 vs. 0.072, Figure 3). Sensitivity did not differ much between the two checks of interest and was even slightly higher for increments presented on the lighter than on the darker-looking check (4% difference in the 2IFC and 8% difference in the yes-no task).

An ANOVA performed on check increments also showed higher thresholds for the yes-no compared to the data.

Table 1. Significant effects of different ANOVAs computed for the data. Notes: Degrees of freedom were \( F(1, 8) \) for the first three ANOVAs and \( F(1, 4) \) for the last. “Inctype” is short for “increment type” and had the two levels: check and blob. The “task” had the two levels, yes-no and 2IFC, and “check” is short for “check position” and had the two levels: inside and outside the shadow.

<table>
<thead>
<tr>
<th>ANOVA effect</th>
<th>( F )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2(inctype) ( \times ) 2(task) ( \times ) 2(check)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inctype</td>
<td>38.53</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>task</td>
<td>81.74</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>check</td>
<td>5.5</td>
<td>( p = 0.047 )</td>
</tr>
<tr>
<td>inctype ( \times ) task</td>
<td>59.67</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>inctype ( \times ) check</td>
<td>15.6</td>
<td>( p = 0.004 )</td>
</tr>
<tr>
<td>task ( \times ) check</td>
<td>11.31</td>
<td>( p = 0.010 )</td>
</tr>
<tr>
<td>inctype ( \times ) task ( \times ) check</td>
<td>8.83</td>
<td>( p = 0.018 )</td>
</tr>
<tr>
<td>blob: 2(task) ( \times ) 2(check)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>task</td>
<td>42.81</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>check: 2(task) ( \times ) 2(check)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>task</td>
<td>84.41</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>check</td>
<td>10.12</td>
<td>( p = 0.013 )</td>
</tr>
<tr>
<td>task ( \times ) check</td>
<td>12.45</td>
<td>( p = 0.008 )</td>
</tr>
<tr>
<td>ellipse: 2(task) ( \times ) 2(check)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>check</td>
<td>12.61</td>
<td>( p = 0.024 )</td>
</tr>
</tbody>
</table>
the 2IFC task (0.144 vs. 0.097), but here, in accordance with the appearance data, higher thresholds were measured on the lighter compared to the darker-looking check (0.136 vs. 0.106, Figure 4). This sensitivity difference between the checks inside and outside the shadow was larger in the yes-no (37%) than in the 2IFC task (16%, see Table 1 for the test statistics).

An ANOVA computed for the ellipse increments also showed a main effect for the position of the increment, however, in the direction opposite to that observed for the check increments: Thresholds were, on average, 27% lower on the lighter-looking check inside the shadow than on the equiluminant darker-looking check outside the shadow (Figure 5).

The observed differences in sensitivity between the forced-choice and yes-no tasks were mainly quantitative and in the expected direction. According to signal detection theory, sensitivity in the forced-choice task should be higher by a factor of about $\sqrt{2}$ relative to the yes-no task, assuming the standard equal-variance Gaussian model (Green & Swets, 1966). We did not test for that quantitative difference, but we did find that sensitivity in the forced-choice task was higher than or sometimes equal to sensitivity in the yes-no task but never lower than that. The magnitude of the differences between the tasks depended on the type of increment that was used.

### Discussion

The objective of the experiments was to test whether sensitivity to intensity differences is limited by the proximal or the perceptual representation of the stimulus, e.g., the local luminance of an image region or its perceived lightness. In order to test that question, we measured apparent lightness and luminance increment thresholds at two equiluminant image regions, which were the equiluminant dark and light checks in a version of the Adelson checkerboard.

The important result concerns the relationship between appearance and sensitivity, which was fundamentally different for the different types of increments. For elliptical increments with a Gaussian border profile, the appearance and sensitivity differences between blobs presented inside and outside the shadow were negligibly small. For checks and proper ellipses, the apparent lightness inside the shadow was judged to be markedly higher than outside the shadow, and the effects were of similar magnitude. That was not the case for the thresholds. Check increments yielded higher thresholds inside than outside the shadow whereas for elliptical increments the opposite was true. Here,
thresholds were lower inside than outside the shadow. Hillis and Brainard (2007), who conducted a very similar experiment, also observed threshold differences between the checks inside and outside the shadow. They measured thresholds at different pedestal intensities and found that the sign of the threshold difference depended on the pedestal intensity: For low pedestals, thresholds were lower outside the shadow as with our check increment whereas at high pedestals thresholds were lower inside the shadow as with our ellipse increment. The exact crossover point might thus depend on the spatial structure of the test, and so an exact comparison between the thresholds in both studies is difficult.

Despite the differences between increment types, the present results are not compatible with the assumption that sensitivity is solely determined by the local luminance of an image region because we found that Weber fractions for lightness increments were not (only) a function of the (physical) local luminance on the retina (proximal). The only increment for which identical thresholds were observed on equiluminant backgrounds was the blurred ellipse. However, blurred ellipses placed on equiluminant checks were also judged to be almost identical in appearance. Thus, the blurred ellipses cannot be considered for answering the present research question. Both other increments, ellipses and checks, despite being identical in local luminance and even in local luminance contrast (ellipses) yielded significant differences in sensitivity. Before we address the question of why the threshold differences with checks and ellipses were in opposite directions, we will consider the reasoning that was underlying the hypothesis that the sensitivity to stimulus differences at an image region is determined by its apparent lightness (Burkhardt, 1966; Cornsweet & Teller, 1965; Heine mann, 1961; Henning et al., 2000; McCourt & Kingdom, 1995; Krauskopf & Zaidi, 1986).

Implicit in that logic was the assumption that the Weber-type discrimination mechanism responds equally to an enhanced internal response that was elicited by a more intense external stimulus and to an enhanced internal response that was elicited by an indirect effect of the context (Fechner, 1860, p. 301). In other words, such a discrimination mechanism would not distinguish between an internal response to a true luminance of $x \frac{cd}{m^2}$ and an internal response to an actually lower luminance that only looks like $x \frac{cd}{m^2}$. Figure 6A shows again the magnitude of the appearance difference that we observed between the two equiluminant checks inside and outside the shadow. Figure 6B illustrates the above logic according to which the checks—which look different because of their context (inside vs. outside the shadow)—are treated as actually being different in luminance and hence elicit different internal responses. We refer to that logic as the single-transfer function scenario. Now, with constant JNDs on the internal response axis (black arrows in Figure 6B), the higher luminance value will result in a higher increment value (red arrow) than the lower luminance value (blue arrow). Thus, in the single-transfer function scenario, the dependence of the discrimination mechanism on stimulus appearance would be reflected in higher thresholds for stimuli that appear lighter.

However, there is a problem with that logic, which has been vividly expressed by Guth (1973), who encountered it studying the inhibitory effects of an annulus on the apparent brightness of the center and its influence on thresholds. He argued “that visual psychophysics as a science has uncritically accepted the logically untenable notion that thresholds can be used to assess the inhibitory effects of interretinal or intraretinal illumination. To the extent that our measuring instrument (the $\Delta I$ or threshold light) is affected by the same inhibition that we are trying to evaluate, our efforts may prove useless. There are many examples of the problem ... an attempt to use incremental thresholds to evaluate the inhibitory activity across a Mach band is only permissible if the inhibition does not also affect the test flash” (Guth, 1973, p. 952). Thus, according to Guth, the same mechanism that causes the light check to look lighter than the equiluminant dark check would also cause any luminance increment that is added to the light check to be expanded relative to identical increments that are added to the dark check. His idea is consistent with a different scenario in which the luminance-to-lightness
mapping involves separate functions at the check positions inside and outside the shadow (Figure 6C). With two different functions, it is possible that identical distances on the internal axis (black arrows in Figure 6) translate into similar or even smaller distances in units of stimulus intensity (colored arrows of Figure 6C). According to this scenario, even if discrimination were based on appearance (in a Weber-type fashion), that would not necessarily be expressed in differences in thresholds because, as Guth pointed out, the increments would undergo the same change as their respective background luminances, and hence JNDS would be identical. Adelson (2000) suggested the existence of different lightness transfer functions to describe the computations the visual system has to perform in order to inverse the mappings between surface reflectance and luminance that occur with different intervening media, such as shadows and transparencies. The idea has also been supported experimentally (Allred, Radonjic, Gilchrist, & Brainard, 2012). However, their implication for studying the relationship between sensitivity and lightness has not been so clear, at least not to the authors of the present paper.

The different transfer function scenario thus provides an account of the present results for the ellipses that is still consistent with the idea that visual discrimination performance is limited by the perceptual and not by the proximal representation of external stimuli. That is because, depending on the shapes of the transfer functions that are being probed, discrimination thresholds might increase, decrease, or be identical for retinal image regions that differ in appearance. In order to critically test the question of whether discrimination performance and appearance are related, one has to first determine the shapes of the transfer functions because they will allow predictions about discrimination performance (Maertens & Shapley, in press).

We would like to note that the proposition of different transfer functions might also provide an account for the phenomenon of assimilation. Assimilation describes an effect by which two equiluminant regions differ in apparent lightness. However, unlike the equiluminant target regions in simultaneous contrast, the target regions in assimilation displays are identical also in luminance contrast relative to their local backgrounds (e.g., Reid & Shapley, 1988). When viewed under suprathreshold conditions like in the adjustment task of the present experiment, the proper ellipses provide a vivid demonstration of assimilation (Figure 1C). Whereas assimilation effects have usually been reported to be smaller in magnitude than simultaneous contrast effects, we found the appearance difference between proper ellipses inside and outside the shadow to be at least as great as the corresponding difference between the checks (Figure 2). Within the two–transfer function scenario, and according to Guth’s (1973) logic, the appearance difference between the ellipses (assimilation) should be of the same size as the difference between the checks because both effects result from a common underlying mechanism. This is because two different transfer functions not only map identical luminance values to different perceived lightness values, but they also map identical luminance differences to unequal differences in perceived lightness (Figure 6D). The different transfer function scenario would explain why the appearance effects for ellipses and checks were of comparable extent and why the
thresholds for elliptical increments were smaller inside the shadow than outside.

The assumption of different transfer functions does not explain why for check increments the sensitivity was lower inside than outside the shadow. That result would be better explained by a single transfer function scenario; however, given the above reasoning, we do not find that possibility convincing. In our opinion, the check increment was the most realistic one because it probed observers’ sensitivity to differences in surface lightness for surfaces that were naturally embedded in the test stimulus. However, it is also the most complicated increment because the checks differed in two ways in their local surround: They were exposed to different illumination (shadow vs. full illumination), and they were surrounded by checks of different reflectance (dark surrounding checks inside and light surrounding checks outside the shadow). Discrimination of check surface lightness was, in fact, more difficult than that for the other two increment types. Therefore, it is not unlikely that different and probably rather complicated mechanisms that require an analysis of the scene have been involved in check lightness discrimination.

Our study is very similar to that of Hillis and Brainard (2007), who also measured discrimination performance and appearance in a version of the Adelson checkerboard using blurred ellipses as increments. However, they tested what they called the “common mechanism hypothesis,” which, in contrast to our model, does not imply that discrimination for two stimuli that appear identical should match. Instead, it is supposed that separate intensity-response functions exist inside and outside the shadow, and it was asked whether a pair of such functions could simultaneously account for both thresholds and appearance. Appearance is thereby predicted by the response (equal response → equal appearance), and thresholds are predicted by the slopes of the intensity-response function (steeper slope → lower threshold). This model failed to account for threshold and appearance data inside and outside the shadow, and hence, the common mechanism hypothesis was rejected (Hillis & Brainard, 2007).

Conclusions

In the present paper, we measured discrimination thresholds and appearance in a naturalistic scene context in which surface reflectance and scene illumination are independently defined and rendered such as to result in a luminance image. We found that for equiluminant and equal-contrast increments (proper ellipses), a common mechanism could account for both stimulus appearance and sensitivity to stimulus differences. That common mechanism involves the proposition of different mapping or transfer functions that relate luminance and apparent lightness (e.g., Adelson, 2000). We also found that for increments that were an integral part of the tested stimulus pattern (checks) appearance and sensitivity data could not be explained by the same model. This more realistic increment was also more complicated to judge, and thus, it is likely that different processes were involved in that type of judgment. We report two interesting corollaries of the present experiments: (a) The proposed transfer functions provide a novel explanation of the phenomenon of assimilation, and (b) blurring an ellipse’s border can profoundly change its appearance, at least when its appearance is determined by the context.

Keywords: lightness, luminance, transfer function, discrimination, appearance, Weber’s law, assimilation

Acknowledgments

This work has been supported by an Emmy-Noether research grant of the German Research Foundation to Marianne Maertens (DFG MA5127/1-1) and, in part, by the German Federal Ministry of Education and Research (BMBF) through the Bernstein Computational Neuroscience Program Tübingen (FKZ: 01GQ1002). We are grateful to Robert Shapley for critical discussion of the experiments and to Torsten Betz, Thomas Otto, and two anonymous reviewers for insightful comments on earlier versions of the manuscript. In particular, we would like to thank our editor David Brainard, who spared no effort to clarify the proposed model relating appearance to discriminability; obviously, all remaining errors and obscurities are our own.

Commercial relationships: none.
Corresponding author: Marianne Maertens.
Email: marianne.maertens@tu-berlin.de.
Address: Modelling of Cognitive Processes Group, Department of Software Engineering and Theoretical Computer Science, Technische Universität Berlin, Berlin, Germany.

References


