

## **Scaling and Discriminability of Perceived Gloss**

Jacob R. Cheeseman<sup>1\*</sup>, James A. Ferwerda<sup>2</sup>, Frank J. Maile<sup>3</sup>,  
and Roland W. Fleming<sup>1,4</sup>

<sup>1</sup>Department of General Psychology, Justus Liebig University Giessen, Otto-Behaghel-Str. 10F,  
Giessen 35394, Germany

<sup>2</sup>Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, 54 Lomb  
Memorial Dr., Rochester, New York 14623, USA

<sup>3</sup>SCHLENK Metallic Pigments GmbH, Barnsdorfer Hauptstr. 5, Roth, 91154, Germany

<sup>4</sup>Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig  
University Giessen, Germany

\*Corresponding author: [Jacob.Cheeseman@psychol.uni-giessen.de](mailto:Jacob.Cheeseman@psychol.uni-giessen.de)

## GLOSS SCALING AND DISCRIMINABILITY

### Abstract

While much attention has been given to understanding biases in gloss perception (e.g., changes in perceived reflectance as a function of lighting, shape, viewpoint and other factors), here we investigated sensitivity to changes in surface reflectance. We tested how visual sensitivity to differences in specular reflectance varies as a function of the magnitude of specular reflectance. Stimuli consisted of renderings of glossy objects under natural illumination. Using Maximum Likelihood Difference Scaling, we created a perceptual scaling of the specular reflectance parameter of the Ward reflectance model. Then, using the Method of Constant Stimuli and a standard 2AFC procedure, we obtained psychometric functions for gloss discrimination across a range of reflectance values derived from the perceptual scale. Both methods demonstrate that discriminability is significantly diminished at high levels of specular reflectance, suggesting that gloss sensitivity depends on the magnitude of change in the image produced by different reflectance values.

## GLOSS SCALING AND DISCRIMINABILITY

### 1. Introduction

The perception of real and virtual surface gloss has been investigated with a variety of experimental and analytical techniques, including the Method of Paired Comparisons (Marlow, Kim, & Anderson, 2012), Multidimensional Scaling (Pellacini, Ferwerda, & Greenberg, 2000), Maximum Likelihood Difference Scaling (Obein, Knoblauch, & Viéot, 2004) and Maximum Likelihood Conjoint Measurement (Ho, Landy, & Maloney, 2008). These studies have focused on judgements of suprathreshold appearance differences and/or asymmetric viewing conditions to test how perceived surface reflectance varies as a function of physical surface reflectance, and other factors such as lighting and shape. Yet, for many practical purposes it is important to know not only which reflectance a given surface appears to have, but also how well observers can discriminate between surfaces that differ only in their intrinsic reflectance properties. Surface gloss discrimination is believed to involve fine-scale examination of local image features, such as specular highlights (Phillips, Ferwerda, & Nunziata, 2010). However, it is also known that observers may adopt different strategies when tasked to evaluate the “gloss” of a surface, which consists of multiple appearance dimensions (Hunter & Harold, 1987; Leloup, Pointer, Dutré, & Hanselaer, 2012). To what extent do suprathreshold judgements of surface gloss predict near-threshold discrimination of specular reflectance? How does sensitivity to gloss vary as a function of the magnitude of specular reflectance?

Mantiuk, Kim, Rempel, and Heidrich (2011) demonstrated that near-threshold image differences can predict suprathreshold differences of complex attributes such as overall image quality. However, it remains unclear whether this also applies in the domain of material appearance. Given that two images can depict surfaces that appear to be made of the same material despite visible differences (Ramanarayanan, Ferwerda, Walter, & Bala, 2007), just-noticeable changes in surface reflectance may not be relevant for judging the overall similarity of material properties such as gloss. Similarly, while suprathreshold perceptual scaling is well-suited

## GLOSS SCALING AND DISCRIMINABILITY

to assessing image similarity (Charrier, Knoblauch, Maloney, Bovik, & Moorthy, 2012), such methods are not necessarily valid for estimating the discriminability of local image features, such as specular highlights (Protonotarios, Johnston, & Griffin, 2016). Indeed, it is possible that suprathreshold and near-threshold judgements evoke non-trivial differences in sensory representation (Aguilar, Wichmann, & Maertens, 2017). Yet, where scaling methods do predict discrimination performance, this offers a highly efficient way to evaluate sensitivity without participants having to perform long and frequently frustrating experiments.

The following experiments were designed to determine whether suprathreshold scaling can predict just-noticeable differences in surface reflectance. We find, similar to previous studies which directly compared judgements of near-threshold and suprathreshold appearance differences in the watercolor effect (Devinck & Knoblauch, 2012) and visual contrast (Kingdom, 2016), that discrimination performance is well-predicted by suprathreshold scaling. At the same time, our study also measures how appearance and sensitivity vary as a function of specular reflectance. This finding has potentially important implications for future studies of material appearance across the fields of industrial manufacturing, computer graphics, and vision science.

### **2. Experiment 1: Establishing a perceptual scale for surface reflectance**

We first sought to construct and verify a perceptual scale for surface gloss. Our approach was similar to that taken by Pellacini, Ferwerda, and Greenberg (2000), who applied Multidimensional Scaling (MDS) to judgements of computer-simulated glossy spheres under artificial illumination. With this data they constructed a perceptually-scaled gloss space consisting of two dimensions (contrast and distinctness of the reflected image), which they later used to derive just-noticeable differences (JNDs) in gloss (Ferwerda, Pellacini, & Greenberg, 2001). Here we employed a newer method called Maximum Likelihood Difference Scaling (MLDS; see Maloney & Yang, 2003), more naturalistic (although still computer-generated) stimuli, and we varied only the specular reflectance of the target object, while all other scene variables were fixed.

## GLOSS SCALING AND DISCRIMINABILITY

### 2.1 Observers

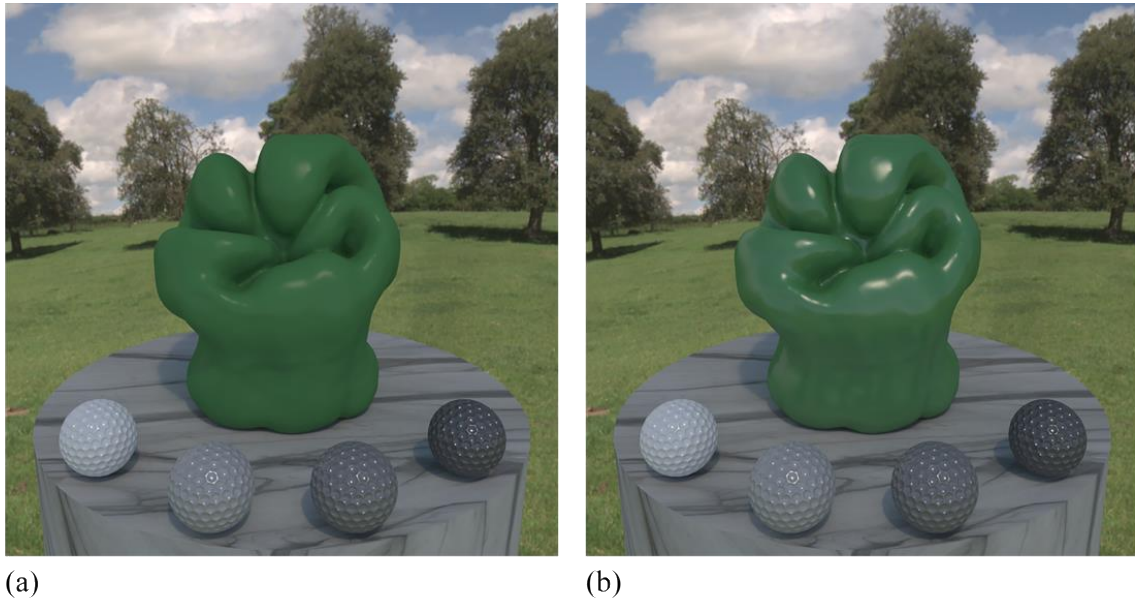
Ten adults (5 males and 5 females; age range: 19 to 40 years;  $M = 24$  years,  $SD = 6.2$  years) with normal or corrected-to-normal visual acuity participated in the experiment and were paid 8€ per hour. All participants provided informed consent prior to the following experiments, which were approved by the ethics review board at Justus Liebig University Giessen and conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

### 2.2 Stimuli

Seven stimulus images were created with the Mitsuba v0.5 physically-based renderer (Jakob, 2010). The PNG image files were rendered in sRGB pixel format and tone mapped using the method described in Reinhard, Stark, Shirley, and Ferwerda (2002) with the following parameter values: key = 0.18; burn = 0; gamma = 2.0. Global illumination calculations were performed using the photon mapping technique developed by Jensen (Jensen, 1996), with 16 samples per pixel (utilizing the Sobol Quasi-Monte Carlo generator; see Grünschloß, Raab, & Keller, 2012 and Joe & Kuo, 2008) a windowed Gaussian reconstruction filter, and interreflections limited to two bounces. At a viewing distance of about 50 cm, the  $720 \times 720$  pixel images subtended approximately 19.06 degrees of visual angle. The target object depicted in each image was a laser-scanned 3D model of a bell pepper (Norman & Phillips, 2016), which subtended approximately 4.81 degrees of visual angle. Objects of this type have been used in previous studies of shape and gloss perception (e.g., Norman et al., 2016). The rendered scene (see Fig. 1) consisted of the central target object seated on a marble-textured pedestal with four golf balls positioned in the foreground, all under natural illumination. Golf balls have been used as probe objects in previous studies where participants were required to estimate lighting conditions (e.g., Xia, Pont, & Heynderickx, 2014); therefore, we included these objects in order to provide as much information about the lighting in the scene as possible, and to anchor judgements about the varying reflectance properties of the target object (Gilchrist et al., 1999). A lens flare artifact

## GLOSS SCALING AND DISCRIMINABILITY

present in the illumination map ("Morestad Farm" from the Southampton-York Natural Scenes dataset; see Adams et al., 2016) was manually removed by applying a median-filter to the affected image region.



**Fig. 1.** Example stimulus images, in which only the specular reflectance of the green target object is varied from low (a) to high (b). The scene consists of a 3D model of a bell pepper (*Capsicum annuum*) seated on a marble pedestal under natural illumination. Golf balls are arrayed in the foreground to provide information about the illumination field.

All objects in the scene were rendered with a modified version of the Ward BRDF model that produces improved physical accuracy at grazing angles (Geisler-Moroder & Dür, 2010). The model features three parameters that separately control the specular reflectance ( $\rho_s$ ), diffuse reflectance ( $\rho_d$ ), and roughness ( $\alpha$ ) of the rendered surface. The set of stimulus images represent seven equal steps in the specular reflectance of the target object ( $\rho_s = \{0.017, 0.031, 0.044, 0.058, 0.072, 0.085, 0.099\}$ ); this range of values matches that used by Pellacini, Ferwerda, and Greenberg (2000). As in previous studies of gloss perception, a dark green color ( $R = 0.1, G =$

## GLOSS SCALING AND DISCRIMINABILITY

0.3,  $B = 0.1$ ; see Fleming, Dror, & Adelson, 2003) and low surface roughness ( $\alpha = 0.04$ ; see Pellacini et al., 2000) were used in order to accentuate specular highlights. Given that the golf balls were intended to supplement the available lighting information, each was rendered with high specular reflectance ( $\rho_s = 0.099$ ), a uniform diffuse reflectance value (RGB = 0.9, 0.45, 0.225, 0.113), and low roughness ( $\alpha = 0.04$ ). The cylindrical pedestal object was wrapped in a marble texture bitmap with 50% Lambertian reflectance.

### 2.3 Procedure

The experiment was controlled by a Dell Precision T3500 desktop computer running Windows 10 v1809 (OS Build 17763.503) and PsychoPy v3.0.7 (Peirce & Macaskill, 2018). Stimulus images were displayed using an Eizo ColorEdge CG277 self-calibrating LCD monitor (68.4 cm diagonal; 2560 x 1440 resolution), which was calibrated with the following settings: sRGB color gamut; D65 white point; 80 cd/m<sup>2</sup> brightness; 2.0 gamma. With these monitor settings and the tonemapping previously described, changes in specular reflectance correspond to proportional changes in display luminance. The participants were seated in darkroom conditions approximately 50 cm from the monitor, which displayed the images against a uniform grey background. Following the Method of Triads variant of MLDS, three images were simultaneously presented on each trial, which remained visible until the participant selected the (left or right) pair of images that depict the smallest difference in gloss relative to the central target object. After a response was entered, the images were replaced by a central white fixation cross for 750 ms, and the next trial would begin with a new combination of images. With three images presented on each trial, and seven different images in the stimulus set, each participant completed a total of 35 trials (i.e., one trial per distinct combination of three out of seven images).

### 2.4 Results

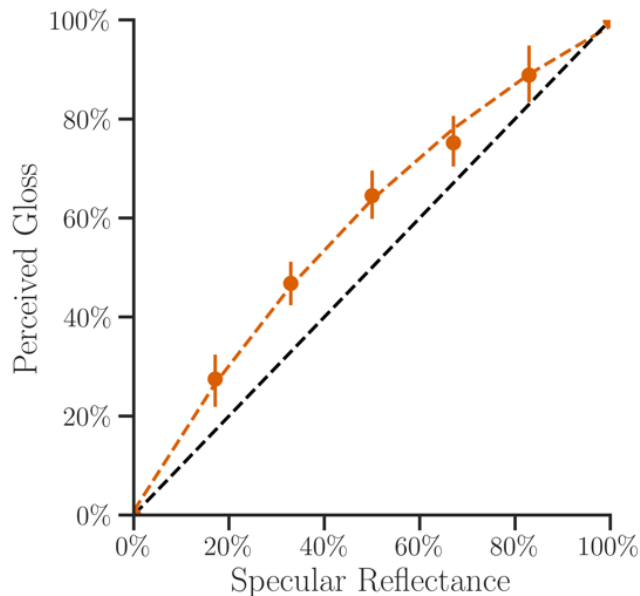
The pooled responses from all 10 participants were treated as trial repetitions and analyzed using the implementations of MLDS by Kingdom and Prins (2016) and Aguilar et al. (2017). As can be

## GLOSS SCALING AND DISCRIMINABILITY

seen in Fig. 2a (orange data), MLDS reveals that for this particular scene, linear steps in specular reflectance are nonlinearly related to differences in gloss magnitude. Previous studies have found the relationship between physical reflectance and perceived gloss to be approximately linear (Pellacini et al., 2000), or a complex nonlinear function (Obein et al., 2004), while here we observe a very mild compressive function. The assumed form of this function and its best-fitting coefficients (determined by nonlinear least squares) are shown in Eq. 1

$$\psi = -1.726 \exp(-0.1941 S) + 1.431 \quad (1)$$

where perceptual magnitude  $\psi \in \mathbb{R} : \psi \in [0,1]$  and stimulus magnitude  $S \in \mathbb{Z} : S \in [1,7]$ .



**Fig. 2.** Maximum Likelihood Difference Scaling results. The underlying MLDS data is provided in Data File 1. According to the MLDS perceptual scale (orange data), perceived magnitudes of gloss are related to physical magnitudes of specular reflectance by a compressive non-linearity, which deviates from a perfectly linear relationship (black diagonal). Error bars indicate bootstrapped 95% confidence intervals.



### **3. Experiment 2: Measuring discriminability on a perceptual scale**

With a suprathreshold perceptual gloss scale in hand, we sought to characterize discriminability at equidistant locations on this scale. However physical and perceptual magnitudes are quantitatively related for a given set of conditions, it is often assumed that the tasks employed to estimate discriminability, or to construct a perceptual scale, involve qualitatively similar kinds of judgements. In other words, the difference between suprathreshold and near-threshold judgements should be one of degree and not of kind. In the following experiment discriminability is estimated with the Method of Constant Stimuli, which unlike MLDS, requires values of specular reflectance that probe the full range of discriminability in order to determine just-noticeable differences of this parameter. This experiment therefore tests whether suprathreshold scaling (MLDS) can predict differences in discriminability that normally accompany absolute changes in stimulus magnitude.

#### **3.1 Observers**

Twenty-three adults (10 males and 13 females; age range: 18 to 29 years;  $M = 22.8$  years,  $SD = 3.2$  years) with normal or corrected-to-normal visual acuity participated in the experiment and were paid 8€ per hour. All participants provided informed consent prior to the following experiments, which were approved by the ethics review board at Justus Liebig University Giessen and conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

#### **3.2 Stimuli**

The virtual scene from the previous experiment was also used here; however, subthreshold and suprathreshold intervals of specular reflectance were used to vary the gloss of the target object. Three equidistant standard parameter values of specular reflectance were calculated using the

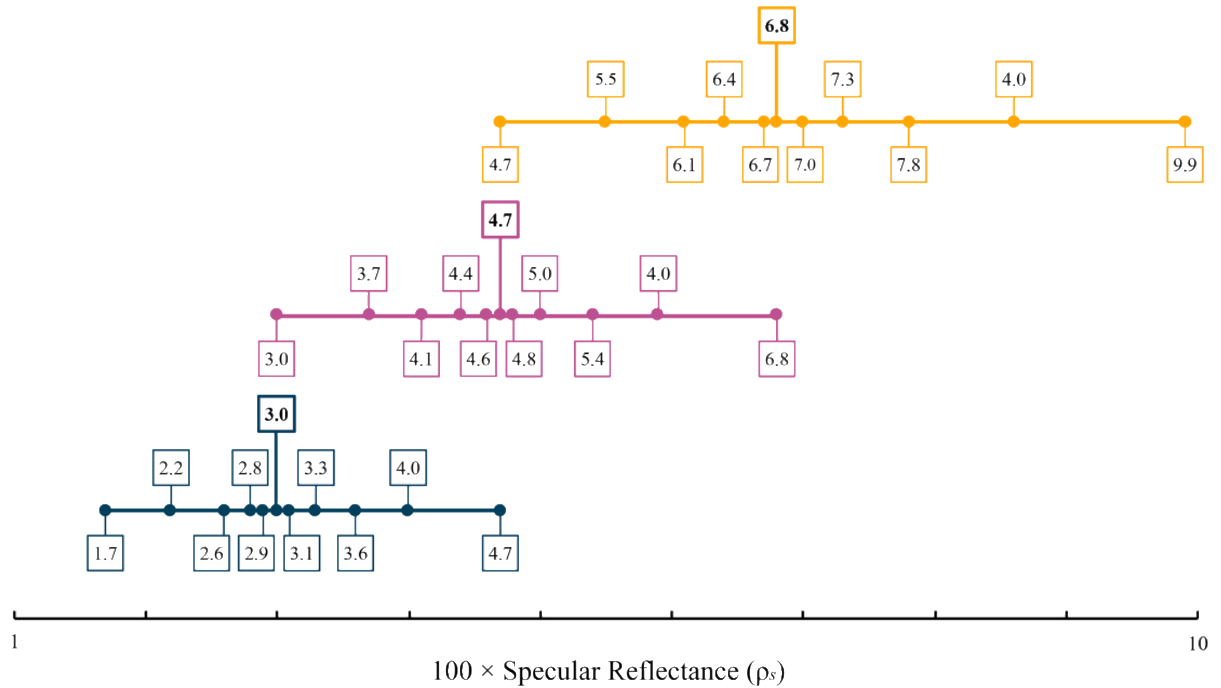
## GLOSS SCALING AND DISCRIMINABILITY

perceptually uniform scale obtained in Experiment 1. This was accomplished by inputting five linearly spaced perceptual magnitudes ( $\psi \in \mathbb{R} : \psi \in [0,1]$ ) to the inverted form of Eq. 1

$$S = \frac{\log\left(\frac{\psi - 1.431}{-1.726}\right)}{-0.1941} \quad (2)$$

transforming the resulting stimulus magnitudes to specular reflectance values using min-max normalization ( $0.017 \leq \rho_s \leq 0.099$ ), and finally retaining the middle three values ( $\rho_s = \{0.030, 0.047, 0.068\}$ ). The perceived difference in gloss between each of the three standard values of specular reflectance is therefore equivalent. Ten comparison values of specular reflectance were also calculated for each standard, with five values above and five below each corresponding standard value. In order to ensure the perceptual uniformity of each set of comparison values, the minimum and maximum comparison values for each standard were calculated using the perceptually uniform scale, while intermediate comparison values were scaled logarithmically. The complete stimulus set (3 standards + 30 comparisons = 33 images) was rendered with the values of specular reflectance listed in Fig. 3.

## GLOSS SCALING AND DISCRIMINABILITY



**Fig. 3.** Rendered values of the Ward specular reflectance parameter ( $\rho_s$ ) used to estimate discriminability with the Method of Constant Stimuli. Three standard values (shown in bold) and ten logarithmically scaled comparison values for each standard were calculated at equidistant locations on the perceptually uniform scale obtained in Experiment 1 with MLDS. On any given trial in the experiment observers visually discriminated between a low, medium, or high reflectance standard image and a randomly selected comparison image from the corresponding subset.

### 3.3 Procedure

Observers were tested under the same conditions described for Experiment 1, except for the following important differences. Here, with the goal of measuring discriminability on our perceptually uniform scale, we employ the Method of Constant Stimuli in a 2AFC task, wherein two images (i.e., a standard and comparison stimulus) are presented on each trial, and the observer selects the left or right image depicting the target object with the greater degree of gloss. The low, medium, or high reflectance standard images were only paired with images from the

## GLOSS SCALING AND DISCRIMINABILITY

corresponding subset (e.g., if the standard reflectance  $\rho_s = 0.030$ , then the comparison reflectance  $\rho_c \in \{0.017, 0.022, 0.026, 0.028, 0.029, 0.031, 0.033, 0.036, 0.040, 0.047\}$ ). The standard stimulus image appeared at random on either side of the screen. Stimulus pairs for each of the three standards were randomly interleaved, and the observers were shown 15 repetitions of the entire set (30 image pairs  $\times$  15 repetitions = 450 trials per observer). Once the observer ended the current trial by entering a response using the left or right arrow key, the screen was cleared for 1 second, and the images for the next trial were displayed. In order to limit the total duration of the experiment to approximately 1 hour, the images were displayed for a maximum of 5 seconds before disappearing from the screen, after which the observer could advance to the next trial by entering a response.

### 3.3 Results

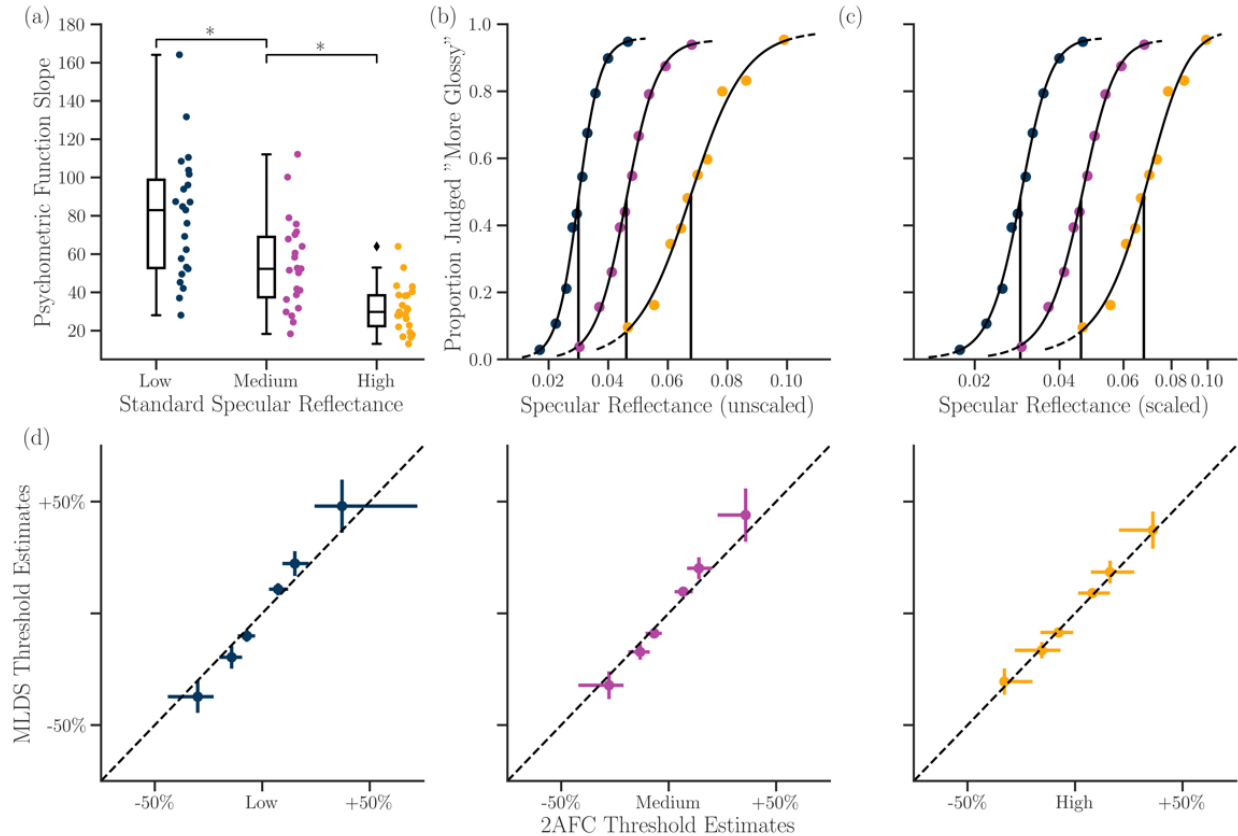
The proportion of trials in which the target object was judged to be glossier in the comparison image was calculated separately for the low, medium, and high ranges of specular reflectance. Logarithmic curves were then fit to these proportions at each value of specular reflectance via Bayesian estimation (Schütt, Harmeling, Macke, & Wichmann, 2016). The psychometric function slopes for each observer (Fig. 4a) illustrate that significant differences in discriminability were found at equidistant locations on our perceptual scale. A one-way repeated measures ANOVA confirmed that for the majority of observers, the slope of the psychometric function decreases with greater magnitudes of specular reflectance ( $F(2,44) = 46.3$ ,  $p < .001$ ,  $\eta_{2p} = .678$ ). Our observers were therefore less sensitive to increasing values of specular reflectance. Differences in discriminability can be seen when psychometric functions estimated from pooled data for each standard are plotted on the physical axis (Fig. 4b). However, these differences in slope are eliminated when the psychometric functions are plotted on the perceptual scale (Fig. 4c). This result demonstrates that the perceptual scale is responsible for the pattern of discriminability

## GLOSS SCALING AND DISCRIMINABILITY

across our range of specular reflectance, and further suggests that MLDS may be used to compensate for such differences in discrimination performance.

The perceptual scale generated by MLDS in Experiment 1 was then used to calculate discriminability estimates that could be directly compared with those obtained in the current experiment. This was accomplished by reparametrizing the perceptual scale in  $d'$  units and reading out discrimination thresholds at specified levels of performance (a detailed technical explanation is provided in Aguilar et al. (2017); analysis code available at <http://github.com/TUBvision/mlDs>). Six  $d'$  values ( $d' \in \{-2.0, -1.0, -0.5, 0.5, 1.0, 2.0\}$ ) were used to estimate thresholds from the MLDS perceptual scale at each of the three standard values of specular reflectance. In a standard 2AFC paradigm these  $d'$  values correspond to correct response rates of 8%, 24%, 36%, 64%, 76%, and 92%, respectively. Discrimination thresholds at these performance levels were then read out from the psychometric functions obtained for low, medium, and high specular reflectance standards in the current experiment (Schütt et al., 2016). According to this analysis, there is broad agreement between the thresholds predicted by MLDS and those obtained using a 2AFC task and the Method of Constant Stimuli. This can be seen in Fig. 4d, where the thresholds for both methods are plotted against each other for the low, medium, and high standards. Note that the 95% confidence intervals for all of the estimated thresholds cross the identity line, thus indicating that negligible differences exist between these methods, at least when directly compared on a common metric.

## GLOSS SCALING AND DISCRIMINABILITY



**Fig. 4.** Discriminability estimates obtained with the Method of Constant Stimuli. The underlying 2AFC data is provided in Data File 2. (a) Psychometric function slopes for individual participants (colored data points) and corresponding box plots for the three standards. Asterisks represent a significance level of  $p < .01$ . (b) Psychometric functions (pooled across participants) for each of the three standard parameter values, here plotted on the unscaled physical axis. (c) Differences in the slope of these psychometric functions are eliminated when plotted on the perceptual scale. (d) Discrimination threshold estimates for the three standard parameter values obtained from the reparametrized MLDS perceptual scale (Experiment 1) and the Method of Constant Stimuli in a 2AFC task (Experiment 2). The thresholds are expressed as differences relative to each standard. Vertical and horizontal lines indicate bootstrapped 95% confidence intervals. The confidence intervals for all estimates cross the (black diagonal) identity line, thus indicating that the estimates from each method are not significantly different.

### 4. Discussion

If surface specular reflectance signals the only difference that could be seen between two otherwise identical surfaces, how does the magnitude of this difference affect what visual information observers use to judge these surfaces? The current study set out to answer this fundamental question in two experiments. First, we established a perceptual scaling of specular reflectance using MLDS, which involves judging the similarity of suprathreshold image differences. We then characterized discriminability along this scale using the Method of Constant Stimuli in a 2AFC task, in which discrimination thresholds are estimated by presenting observers with image differences that span the full range of discriminability. Taken together, our results provide convergent evidence that MLDS can scale both small and large image differences, which allows for successful prediction of discrimination thresholds.

In the original formulation of MLDS, sensory representations are modeled as independent, Gaussian random variables with equal variance (Maloney & Yang, 2003). Simulated violations of this equal variance assumption do not affect the shape of the perceptual scale produced by MLDS [16, 30], but may affect discriminability estimates derived from the signal detection theory formulation of MLDS (Aguilar et al., 2017). The model assumptions underlying MLDS may also interact with stimulus complexity and dynamic range (Aguilar & Maertens, 2020; Protonotarios et al., 2016), both of which have been shown to affect the perception of gloss (Adams, Kucukoglu, Landy, & Mantiuk, 2018; Doerschner, Maloney, & Boyaci, 2010; Obein et al., 2004; Phillips, Ferwerda, & Luka, 2009). It is also plausible that scaling and discrimination tasks induce—or draw on—non-trivial differences in stimulus representation. In the case of surface gloss, near-threshold discrimination involves attending to local features that signal small differences in the proximal stimulus, while suprathreshold scaling involves attending to whole objects and abstracting similarity from multiple dimensions of the distal stimulus (Maloney & Knoblauch, 2020; Phillips et al., 2010). Such task-dependencies may be particularly relevant when the stimulus property in question ("gloss") consists of multiple appearance dimensions (Hunter & Harold, 1987; Toscani,

## GLOSS SCALING AND DISCRIMINABILITY

Guarnera, Guarnera, Hardeberg, & Gegenfurtner, 2020), and is thus more open to interpretation. Then again, under symmetric viewing conditions, where the only visible differences between otherwise identical images are to be found in the relative magnitudes of specular reflectance, the complexity of surface gloss is boiled down to a manipulation of local contrast. Our experiments are therefore similar to those described by Kingdom (Kingdom, 2016), who compared scaling and discrimination data from experiments in which observers judged the difference in luminance of a disk superimposed against a uniform background. Analyses of those data revealed a remarkable degree of agreement between the scaling and discrimination tasks, which was taken as evidence that the sensory representation of contrast is governed by additive noise. Given the potential limitations of MLDS described above, it is reassuring that our findings agree with previous studies that demonstrated agreement between MLDS perceptual scales and discrimination performance for other appearance characteristics (Devinck & Knoblauch, 2012; Kingdom, 2016). This suggests that, at least for comparisons of surfaces that differ only in specular reflectance, MLDS is well able to model judgements of suprathreshold and near-threshold differences in surface appearance.

The results of our experiments also indicate that gloss sensitivity cannot be captured by a single point estimate, since discriminability of gloss critically depends on the magnitude of surface specular reflectance. In this regard, gloss sensitivity would seem to follow Weber's Law, which assumes that discriminability is invariant if and only if physical magnitudes are varied in constant proportion to perceptual magnitudes (Fechner, 1966). Weber's Law has inspired considerable debate about the transducer functions that relate stimulus and sensation (e.g., Stevens, 1961), yet from its inception, Fechner acknowledged that the lawfulness of Weber's Law depends on the nature of the stimulus. For example, he comments that while the law could be demonstrated with experiments in pitch perception, a case for its existence in color perception could not then be made (Fechner, 1860/1966, p. 146). This early observation suggested that a perceptually uniform color space would be a complex mathematical entity, and these complexities were not fully



## GLOSS SCALING AND DISCRIMINABILITY

appreciated until the next century, when it was discovered that small differences in chromaticity could only be adequately specified within local regions of the CIE 1931 color space (MacAdam, 1943; Smith & Guild, 1931). Similarly, the prospect of a uniform perceptual space for surface gloss remains elusive because changes in illumination, shape, and viewpoint can drastically alter the perception of surface material properties (Fleming et al., 2003; Ho, Maloney, & Landy, 2007; Norman, Todd, & Phillips, 2020; Vangorp, Laurijssen, & Dutré, 2007; Zhang, de Ridder, Barla, & Pont, 2020), which therefore means that the validity of any gloss space will be constrained by the viewing conditions chosen for its construction (Fores, Fairchild, & Tastl, 2014). Despite these difficulties, our finding that MLDS provides a solution for both scaling and discriminability of gloss indicates that the construction of a perceptually uniform gloss space is a tractable problem. Moreover, MLDS offers considerable efficiency advantages. To evaluate sensitivity at just three reflectance values using the method of constant stimuli we used 450 trials per participant, many of which were close to threshold performance and therefore potentially frustrating for the participants. While this could be made somewhat more efficient through an adaptive sampling procedure (e.g., Kontsevich & Tyler, 1999; Lieberman & Pentland, 1982; Watson, 2017), in contrast, MLDS delivered quite accurate sensitivity estimates with just 35 trials per participant. This makes it feasible to compare sensitivity across many conditions, a pre-requisite for future studies investigating how factors such as lighting, shape and other reflectance parameters influence sensitivity to gloss.

### **5. Conclusions**

Returning to our central question: to what extent do suprathreshold judgements of surface gloss predict near-threshold discrimination of specular reflectance? It has been argued that just-noticeable differences can predict suprathreshold differences in complex visual properties (Mantiuk et al., 2011), and also that such small image differences are not necessarily relevant to the task of scaling material appearance (Ramanarayanan et al., 2007). Our results demonstrate

## GLOSS SCALING AND DISCRIMINABILITY

that MLDS, a method of perceptual scaling that works with suprathreshold appearance differences, not only predicts discriminability of specular reflectance, but also provides a means for improving the perceptual uniformity of discriminability estimates. Future work will need to characterize the extent to which estimates of gloss discriminability can generalize across asymmetric viewing conditions, in which multiple dimensions of gloss are varied in addition to changes in illumination, shape, and viewpoint. Yet in the long run, a model of surface gloss perception will only be complete if it can correctly predict variations in discriminability as well as suprathreshold appearance.

### **Funding**

H2020 Marie Skłodowska-Curie Actions (H2020-MSCA-ITN-2017, “DyViTo: Dynamics in Vision and Touch”, Project ID 765121); Deutsche Forschungsgemeinschaft (SFB-TRR-135, “Cardinal Mechanisms of Perception”, Project ID 222641018); European Research Council (ERC-2015-CoG-682859, “SHAPE”).

### **Acknowledgments**

We would like to thank Saskia L. Honneller, Britta J. Fritz, and Jasmin Kleis for administering the experimental sessions, and the observers who participated in the experiments. We especially wish to thank Guillermo Aguilar for guidance on computing and comparing the discriminability estimates obtained from the MLDS and 2AFC tasks.

### **Disclosures**

The authors declare no conflicts of interest.

### **References**

Adams, W. J., Elder, J. H., Graf, E. W., Leyland, J., Lutigheid, A. J., & Murry, A. (2016). The Southampton-York Natural Scenes (SYNS) dataset: Statistics of surface attitude. *Scientific Reports*, 6, 35805. <https://doi.org/10.1038/srep35805>

## GLOSS SCALING AND DISCRIMINABILITY

- Adams, W. J., Kucukoglu, G., Landy, M. S., & Mantiuk, R. K. (2018). Naturally glossy: Gloss perception, illumination statistics and tone mapping. *Journal of Vision*, 18(9).
- Aguilar, G., & Maertens, M. (2020). Toward reliable measurements of perceptual scales in multiple contexts. *Journal of Vision*, 20(4), 19. <https://doi.org/10.1167/jov.20.4.19>
- Aguilar, G., Wichmann, F. A., & Maertens, M. (2017). Comparing sensitivity estimates from MLDS and forced-choice methods in a slant-from-texture experiment. *Journal of Vision*, 17(1), 37. <https://doi.org/10.1167/17.1.37>
- Charrier, C., Knoblauch, K., Maloney, L. T., Bovik, A. C., & Moorthy, A. K. (2012). Optimizing multiscale SSIM for compression via MLDS. *IEEE Transactions on Image Processing*, 21(12), 4682–4694. <https://doi.org/10.1109/TIP.2012.2210723>
- Devinck, F., & Knoblauch, K. (2012). A common signal detection model accounts for both perception and discrimination of the watercolor effect. *Journal of Vision*, 12(3), 19. <https://doi.org/10.1167/12.3.19>
- Doerschner, K., Maloney, L. T., & Boyaci, H. (2010). Perceived glossiness in high dynamic range scenes. *Journal of Vision*, 10(9), 11. <https://doi.org/10.1167/10.9.11>
- Fechner, G. T. (1966). *Elements of Psychophysics Vol. I (Translated by Adler, H. E.)*. (D. H. Howes & E. G. Boring, Eds.). New York: Holt, Rinehart and Winston.
- Ferwerda, J. A., Pellacini, F., & Greenberg, D. P. (2001). Psychophysically based model of surface gloss perception. In *Human Vision and Electronic Imaging VI* (Vol. 4299, pp. 4211–4299).
- Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 3(5), 3. <https://doi.org/10.1167/3.5.3>
- Fores, A., Fairchild, M. D., & Tastl, I. (2014). Improving the perceptual uniformity of a gloss space. In *Color and Imaging Conference* (pp. 7–13). Society for Imaging Science and Technology.
- Geisler-Moroder, D., & Dür, A. (2010). A new Ward BRDF model with bounded albedo.

## GLOSS SCALING AND DISCRIMINABILITY

*Computer Graphics Forum*, 29(4), 1391–1398. <https://doi.org/10.1111/j.1467-8659.2010.01735.x>

Gilchrist, A., Kossyfidis, C., Agostini, T., Li, X., Bonato, F., Cataliotti, J., ... Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, 106(4), 795–834. <https://doi.org/10.1037/0033-295X.106.4.795>

Grünschloß, L., Raab, M., & Keller, A. (2012). Enumerating Quasi-Monte Carlo point sequences in elementary intervals. In L. Plaskota & H. Woźniakowski (Eds.), *Monte Carlo and Quasi-Monte Carlo Methods 2010* (pp. 399–408). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-27440-4\\_21](https://doi.org/10.1007/978-3-642-27440-4_21)

Ho, Y. X., Landy, M. S., & Maloney, L. T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science*, 19(2), 196–204. <https://doi.org/10.1111/j.1467-9280.2008.02067.x>

Ho, Y. X., Maloney, L. T., & Landy, M. S. (2007). The effect of viewpoint on perceived visual roughness. *Journal of Vision*, 7(1), 1. <https://doi.org/10.1167/7.1.1>

Hunter, R. S., & Harold, R. W. (1987). *The measurement of appearance* (2nd ed.). John Wiley & Sons.

Jakob, W. (2010). Mitsuba renderer. Retrieved from <https://www.mitsuba-renderer.org/>

Jensen, H. W. (1996). Global illumination using photon maps. In X. Pueyo & P. Schröder (Eds.), *Proceedings of the eurographics workshop on Rendering techniques '96* (pp. 21–30). Vienna: Springer Vienna. [https://doi.org/10.1007/978-3-7091-7484-5\\_3](https://doi.org/10.1007/978-3-7091-7484-5_3)

Joe, S., & Kuo, F. Y. (2008). Constructing sobol sequences with better two-dimensional projections. *SIAM Journal on Scientific Computing*, 30(5), 2635–2654. <https://doi.org/10.1137/070709359>

Kingdom, F. A. A. (2016). Fixed versus variable internal noise in contrast transduction: The significance of Whittle's data. *Vision Research*, 128, 1–5. <https://doi.org/10.1016/j.visres.2016.09.004>

## GLOSS SCALING AND DISCRIMINABILITY

- Kingdom, F. A. A., & Prins, N. (2016). *Psychophysics: A practical introduction* (2nd ed.). Academic Press.
- Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–2737. [https://doi.org/10.1016/S0042-6989\(98\)00285-5](https://doi.org/10.1016/S0042-6989(98)00285-5)
- Leloup, F. B., Pointer, M. R., Dutré, P., & Hanselaer, P. (2012). Overall gloss evaluation in the presence of multiple cues to surface glossiness. *Journal of the Optical Society of America A*, *29*(6), 1105–1114. <https://doi.org/10.1364/JOSAA.29.001105>
- Lieberman, H. R., & Pentland, A. P. (1982). Microcomputer-based estimation of psychophysical thresholds: The Best PEST. *Behavior Research Methods & Instrumentation*, *14*(1), 21–25. <https://doi.org/10.3758/BF03202110>
- MacAdam, D. L. (1943). Specification of small chromaticity differences. *Journal of the Optical Society of America*, *33*(1), 18–26. <https://doi.org/10.1364/JOSA.33.000018>
- Maloney, L. T., & Knoblauch, K. (2020). Measuring and modeling visual appearance. *Annual Review of Vision Science*. <https://doi.org/10.1146/annurev-vision-030320-041152>
- Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision*, *3*(8), 5. <https://doi.org/10.1167/3.8.5>
- Mantiuk, R., Kim, K. J., Rempel, A. G., & Heidrich, W. (2011). HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. In *ACM SIGGRAPH 2011 Papers* (pp. 40:1--40:14). New York, NY, USA: ACM. <https://doi.org/10.1145/1964921.1964935>
- Marlow, P. J., Kim, J., & Anderson, B. L. (2012). The perception and misperception of specular surface reflectance. *Current Biology*, *22*(20), 1909–1913. <https://doi.org/10.1016/j.cub.2012.08.009>
- Norman, J. F., & Phillips, F. (2016). Bell Peppers (v1.1) [3D Object Files]. Retrieved from <http://www.skidmore.edu/~flip>

## GLOSS SCALING AND DISCRIMINABILITY

Norman, J. F., Phillips, F., Cheeseman, J. R., Thomason, K. E., Ronning, C., Behari, K., ...

Lamirande, D. (2016). Perceiving object shape from specular highlight deformation, boundary contour deformation, and active haptic manipulation. *PLOS ONE*, *11*(2), 1–15.  
<https://doi.org/10.1371/journal.pone.0149058>

Norman, J. F., Todd, J. T., & Phillips, F. (2020). Effects of illumination on the categorization of shiny materials. *Journal of Vision*, *20*(2).

Obein, G., Knoblauch, K., & Viéot, F. (2004). Difference scaling of gloss: Nonlinearity, binocularity, and constancy. *Journal of Vision*, *4*(9), 4. Retrieved from 10.1167/4.9.4

Peirce, J., & Macaskill, M. (2018). *Building experiments in PsychoPy*. London: SAGE Publications Ltd.

Pellacini, F., Ferwerda, J. A., & Greenberg, D. P. (2000). Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 55–64). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co. <https://doi.org/10.1145/344779.344812>

Phillips, J. B., Ferwerda, J. A., & Luka, S. (2009). Effects of image dynamic range on apparent surface gloss. *Color and Imaging Conference*, 193–197.

Phillips, J. B., Ferwerda, J. A., & Nunziata, A. (2010). Gloss discrimination and eye movements. In *Proc.SPIE* (Vol. 7527). <https://doi.org/10.1117/12.845399>

Protonotarios, E. D., Johnston, A., & Griffin, L. D. (2016). Difference magnitude is not measured by discrimination steps for order of point patterns. *Journal of Vision*, *16*(9), 2.  
<https://doi.org/10.1167/16.9.2>

Ramanarayanan, G., Ferwerda, J., Walter, B., & Bala, K. (2007). Visual equivalence: Towards a new standard for image fidelity. *ACM Transactions on Graphics (TOG)*, *26*(3), 11.  
<https://doi.org/10.1145/1276377.1276472>

Reinhard, E., Stark, M., Shirley, P., & Ferwerda, J. (2002). Photographic tone reproduction for digital images. *ACM Trans. Graph.*, *21*(3), 267–276.

## GLOSS SCALING AND DISCRIMINABILITY

<https://doi.org/10.1145/566654.566575>

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, *122*, 105–123. <https://doi.org/10.1016/j.visres.2016.02.002>

Smith, T., & Guild, J. (1931). The C.I.E. colorimetric standards and their use. *Transactions of the Optical Society*. <https://doi.org/10.1088/1475-4878/33/3/301>

Stevens, S. S. (1961). To honor Fechner and repeal his law. *Science*, *133*(3446), 80–86.

Toscani, M., Guarnera, D., Guarnera, G. C., Hardeberg, J. Y., & Gegenfurtner, K. R. (2020). Three perceptual dimensions for specular and diffuse reflection. *ACM Transactions on Applied Perception*, *1*(1), 27. <https://doi.org/10.1145/3380741>

Vangorp, P., Laurijssen, J., & Dutré, P. (2007). The influence of shape on the perception of material reflectance. *ACM Transactions on Graphics (TOG)*, *26*(3).

<https://doi.org/10.1145/1276377.1276473>

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, *17*(3), 10. <https://doi.org/10.1167/17.3.10>

Xia, L., Pont, S. C., & Heynderickx, I. (2014). The visual light field in real scenes. *I-Perception*, *5*(7), 613–629. <https://doi.org/10.1068/i0654>

Zhang, F., de Ridder, H., Barla, P., & Pont, S. (2020). Effects of light map orientation and shape on the visual perception of canonical materials. *Journal of Vision*, *20*(4), 13.

<https://doi.org/10.1167/jov.20.4.13>