Journal of the Optical Society of America A
OPTICS, IMAGE SCIENCE, AND VISION

Scaling and discriminability of perceived gloss

Jacob R. Cheeseman,[1,*] **James A. Ferwerda,**[2] **Frank J. Maile,**[3] **and Roland W. Fleming**[1,4]

[1]Department of General Psychology, Justus Liebig University Giessen, Otto-Behaghel-Str. 10F, Giessen 35394, Germany
[2]Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Dr., Rochester, New York 14623, USA
[3]SCHLENK Metallic Pigments GmbH, Barnsdorfer Hauptstr. 5, Roth 91154, Germany
[4]Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University, Giessen, Germany
*Corresponding author: Jacob.Cheeseman@psychol.uni-giessen.de

Received 9 September 2020; revised 4 December 2020; accepted 8 December 2020; posted 10 December 2020 (Doc. ID 409454); published 18 January 2021

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 (MLDS), 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, thus indicating that gloss sensitivity depends on the magnitude of change in the image produced by different reflectance values. Taken together, these experiments also suggest that internal sensory noise remains constant for suprathreshold and near-threshold intervals of specular reflectance, which supports the use of MLDS as a highly efficient method for evaluating gloss sensitivity.

https://doi.org/10.1364/JOSAA.409454

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 [1], multidimensional scaling [2], maximum likelihood difference scaling [3], and maximum likelihood conjoint measurement [4]. 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 [5]. 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 [6,7]. To what extent do suprathreshold judgments 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 [8] 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 [9], 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 to assessing image similarity [10], such methods are not necessarily valid for estimating the discriminability of local image features, such as specular highlights [11]. Indeed, it is possible that suprathreshold and near-threshold judgements evoke non-trivial differences in sensory representation. For example, maximum likelihood difference scaling (MLDS) is a popular suprathreshold perceptual scaling method in which sensory representations are modeled as independent, Gaussian random variables with equal variance. If this internal sensory noise were multiplicative rather than additive, this would not be evident from the shape of the perceptual scale produced by MLDS [12,13], but this could affect discriminability estimates derived from the same perceptual scale [14]. When scaling does predict discrimination, however, performance in both tasks can be modeled under the assumption that suprathreshold and near-threshold judgements share a common transducer function, thus indicating how internal sensory noise grows with stimulus magnitude [15].

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 [16] and visual contrast [15], that discrimination performance is well-predicted by suprathreshold scaling. Perhaps most notably, our study furnishes evidence that internal sensory noise remains constant for suprathreshold and near-threshold intervals of specular reflectance, which supports the use of MLDS as a highly efficient method for evaluating sensitivity without participants having to perform tedious discrimination experiments. These findings have 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 [2], who applied multidimensional scaling to judgments 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 in gloss [17]. Here we employed MLDS (see [12]), 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.

A. Observers

Ten adults (five males and five 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).

B. Stimuli

Seven stimulus images were created with the Mitsuba v0.5 physically based renderer [18]. The rendered scene (see Fig. 1) consisted of a central target object (a laser-scanned 3D model of a bell pepper [19]), on a marble-textured pedestal with four golf balls positioned in the foreground, all lit by a high dynamic range illumination map of an outdoor scene [20]. Global illumination calculations were performed using photon mapping [21], with 16 samples per pixel, and two-bounce interreflections. Surface reflectance properties were represented using the Ward-Dur light reflection model [22]. The model has three parameters that specify the specular reflectance (ρ_s), diffuse reflectance (ρ_d), and microscale roughness (α) of a surface. To produce the seven stimulus images, the specular reflectance of the target object was varied in seven equal steps ($\rho_s = \{0.017, 0.031, 0.044, 0.058, 0.072, 0.085, 0.099\}$); this matches the range of values used by Pellacini, Ferwerda, and Greenberg [2]. As in previous studies of gloss perception, a dark green diffuse color (RGB$\rho_d = \{0.1, 0.3, 0.1\}$ and low

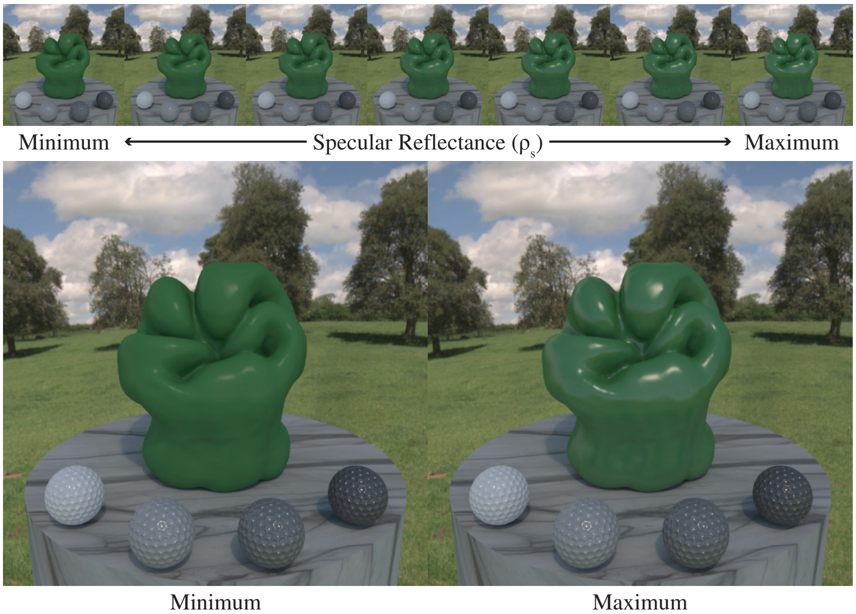

Fig. 1. Stimulus images shown in Experiment 1. The specular reflectance of the green target object is varied in seven equal steps from low ($\rho_s = 0.017$) to high ($\rho_s = 0.099$). The scene consists of a 3D model of a bell pepper (*Capsicum annuum*) seated on a marble pedestal under natural illumination. Golf balls are arrayed in the foreground to provide information about the illumination field.

Table 1. Stimulus Properties

Target Object (Pepper)

	Surface Reflectance Properties					Displayed Image Properties			
Obj. ID	ρ_d	ρ_s	α	**Chromaticity**		**Min. Lum.**	**Max. Lum.**	**Diff. Cont.**	**Spec. Cont.**
	(0.1,0.3,0.1)	0.000	0.04	0.2979	0.4370	1.62	6.61	0.61	0.00
1	"	0.017	"	0.2978	0.4309	1.66	9.39	"	0.09
2	"	0.031	"	0.2977	0.4267	1.70	11.79	"	0.14
3	"	0.044	"	0.2977	0.4232	1.75	13.99	"	0.17
4	"	0.058	"	0.2976	0.4197	1.80	16.32	"	0.20
5	"	0.072	"	0.2975	0.4166	1.85	18.60	"	0.21
6	"	0.085	"	0.2974	0.4138	1.89	20.68	"	0.23
7	"	0.099	"	0.2973	0.4111	1.93	22.90	"	0.24

Anchor Objects (Golf Balls)

	Surface Reflectance Properties					Displayed Image Properties			
Obj. ID	ρ_d	ρ_s	α	**Chromaticity**		**Min. Lum.**	**Max. Lum.**	**Diff. Cont.**	**Spec. Cont.**
1	(0.9,0.9,0.9)	0.099	0.04	0.2942	0.3121	10.06	35.48	0.42	0.14
2	(0.45,0.45,0.45)	"	"	0.2956	0.3150	6.53	28.65	0.32	0.30
3	(0.225,0.225,0.225)	"	"	0.2952	0.3155	3.27	21.08	0.37	0.36
4	(0.113,0.113,0.113)	"	"	0.2937	0.3135	1.63	19.08	0.44	0.40

surface roughness ($\alpha = 0.04$) were used to ensure that surfaces had visible specular highlights (see [2,23]). In previous studies where participants were required to estimate lighting conditions [24], golf balls have been used as probe objects; therefore, we included these objects in our scene to provide supplementary information about scene lighting, and to anchor judgments about surface reflectance properties (e.g., [25]). Each golf ball had high specular reflectance ($\rho_s = 0.099$), achromatic diffuse reflectance ($\rho_d = 0.9, 0.45, 0.225, 0.113$ left-to-right), and low surface roughness ($\alpha = 0.04$). The matte gray pedestal object had reflectance parameters ($\rho_d = 0.5$, $\rho_s = 0.00$) that were modulated by a marble-patterned texture map. The stimulus parameters are summarized in Table 1. The 720 × 720 rendered images were converted to the sRGB color space, tone mapped using the method described in Reinhard, Stark, Shirley, and Ferwerda [26] with parameter values (key = 0.18; burn = 0; gamma = 2.0), and stored in the PNG image format. The complete set of stimulus images for Experiment 1 is contained in Dataset 1, Ref. [27].

The images were displayed on an Eizo ColorEdge CG277 LCD monitor (27 in. diagonal; 2560 × 1440 resolution). At a viewing distance of 50 cm, each image subtended approximately 19 deg of visual angle. The display was calibrated to have an sRGB color gamut, 80 cd/m^2 D65 white point, and a gamma of 2.0. With these settings, changes in the specular reflectance of the target object produced proportional changes in displayed image luminance. Of particular interest for the purposes of this study is the specular contrast of the target object, which is the increase in the contrast of the image of the target object above the base contrast in the image of a diffusely reflecting target object. The base contrast in the displayed image of the target object was (61%) and the specular contrasts produced by our chosen ρ_s values were (9%, 14%, 17%, 20%, 21%, 23%, and 24%). The displayed stimulus image luminances, chromaticities, and contrasts are summarized in Table 1. The display was viewed in a dark room, and the images were presented against a uniform middle-gray background.

C. 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 [28]. 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). Observers typically completed the experiment in less than 5 min.

D. Results

The pooled responses from all 10 participants were treated as trial repetitions and analyzed using the implementations of MLDS by Kingdom and Prins [13] and Aguilar *et al.* [14]. As can be seen in Fig. 2(a) (orange data), MLDS reveals that for this particular scene, linear steps in stimulus magnitude (specular reflectance) are non-linearly related to differences in perceptual magnitude (perceived gloss). Previous studies have found the relationship between physical reflectance and perceived gloss to be approximately linear [2], or a complex non-linear function [3], while here we observe a very mild compressive function. The assumed form of this function and its best-fitting coefficients (determined by non-linear least squares) are shown in Eq. (1):

$$\psi = -1.8472 \exp(-14.27\, S) + 1.4495, \tag{1}$$

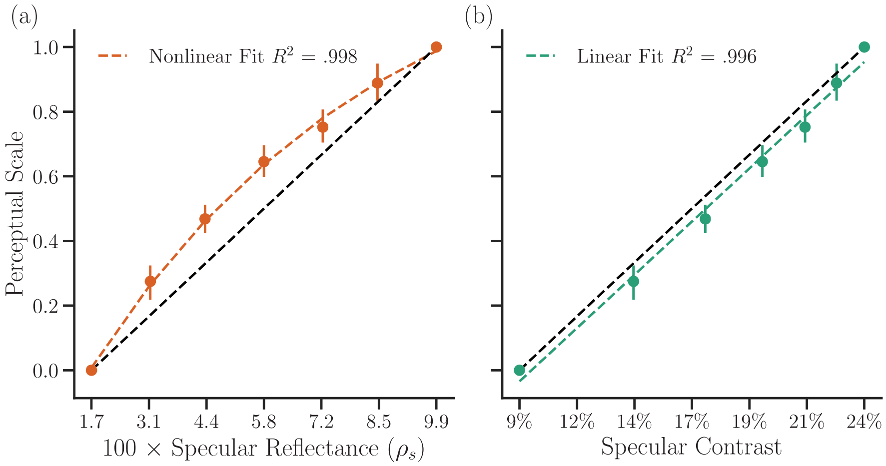

Fig. 2. Maximum likelihood difference scaling results. The underlying MLDS data is provided in Data File 1. (a) According to the MLDS perceptual scale, perceived magnitudes of gloss are related to physical magnitudes of specular reflectance by a compressive non-linearity (orange data), which deviates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

where perceptual magnitude $\psi \in \mathbb{R} : \psi \in [0, 1]$ and stimulus magnitude $S \in \mathbb{Z} : S \in [0.017, 0.099]$ However, when this perceptual scale is plotted against the specular contrasts (SC) of each image [Fig. 2(b); green data], there is a linear relationship ($\psi = 6.83SC - 0.68$). This is in agreement with the findings of Pellacini, Ferwerda, and Greenberg [2] who used multidimensional scaling methods in their studies.

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 judgments. In other words, the difference between suprathreshold and near-threshold judgments 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.

A. Observers

A distinct group of 23 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).

B. 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 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),

$$S = \frac{\log\left(\frac{\psi - 1.4495}{-1.8472}\right)}{-14.27}, \tag{2}$$

and 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; available in Dataset 1, Ref. [27]) was rendered with the values of specular reflectance listed in Fig. 3.

C. 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

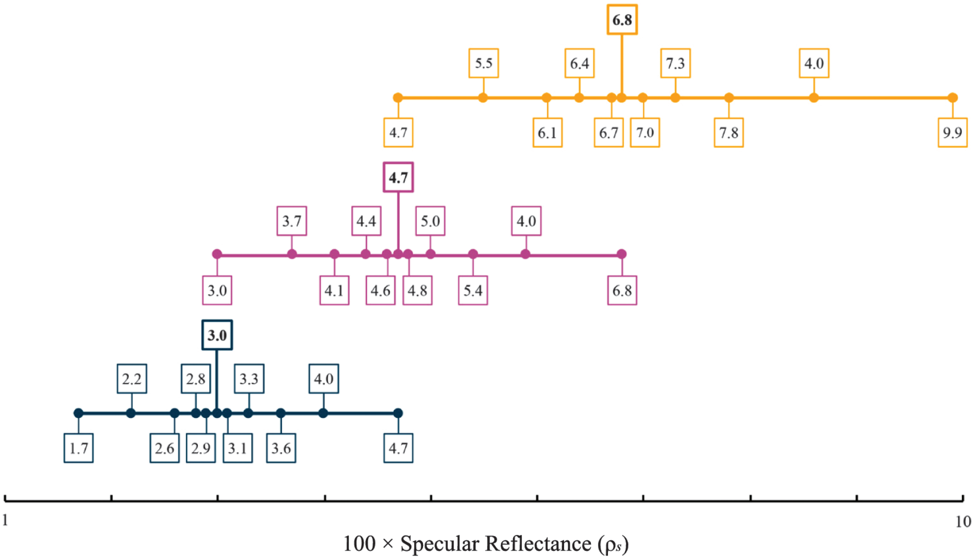

Fig. 3. Rendered values of the Ward specular reflectance parameter (ρ_s) used to estimate discriminability with the method of constant stimuli. Three standard values (shown in bold) and 10 logarithmically scaled comparison values for each standard were calculated at equidistant locations on the perceptually 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.

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 corresponding subset (e.g., if the standard reflectance $\rho_s = 0.030$, then the comparison reflectance $\rho_s \in \{0.017, 0.022, 0.026, 0.028, 0.029, 0.031, 0.033, 0.036, 0.040, 0.047\}$). 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 × 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 s, and the images for the next trial were displayed. In order to limit the total duration of the experiment to approximately 1 h, the images were displayed for a maximum of 5 s before disappearing from the screen, after which the observer could advance to the next trial by entering a response.

D. 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 [29]. The psychometric function slopes for each observer [Fig. 4(a)] 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 ($F(2, 44) = 46.3$, $p < 001$, $\eta_p^2 = .678$). 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. 4(b)]. However, these differences in slope are eliminated when the psychometric functions are plotted on the perceptual scale [Fig. 4(c)]. This result demonstrates that the perceptual scale is responsible for the pattern of 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.* [14]; 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 [29]. According to this between-subjects 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. 4(d), 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

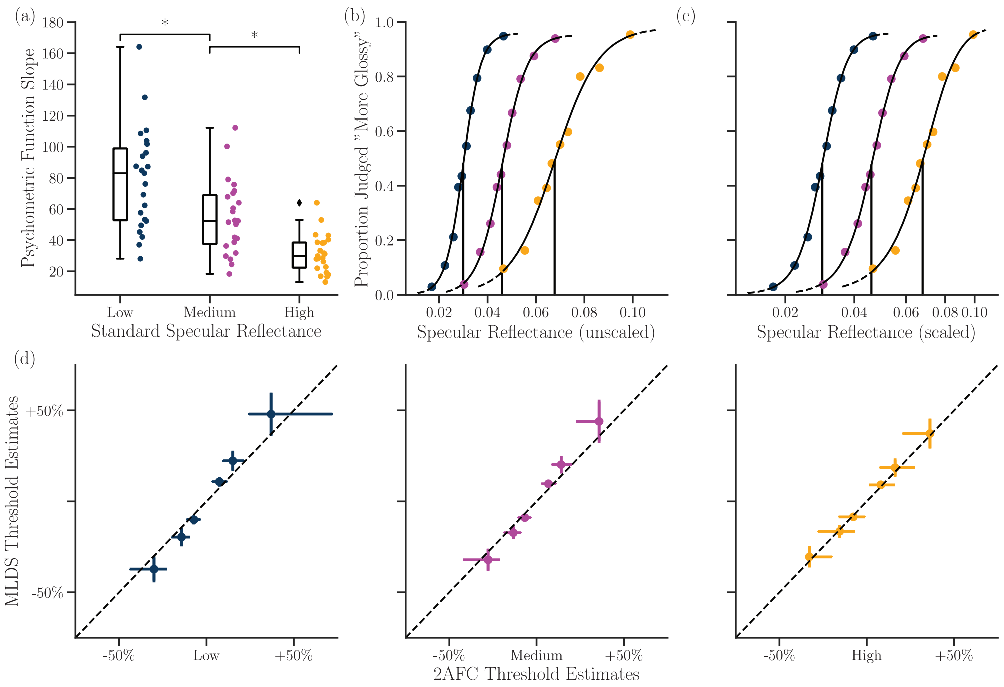

Fig. 4. Discriminability estimates obtained with the method of constant stimuli. The underlying 2AFC data is provided in Data File 2. (a) Psychometric function slopes for individual participants (colored data points) and corresponding box plots for the three standards. Asterisks represent a significance level of $p < 0.01$. (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.

these methods, at least when directly compared on a common metric.

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 formalism of MLDS, sensory representations are modeled as independent, Gaussian random variables with equal variance, while the precision of each trial decision (i.e., which pair is more similar) is estimated by the fitting procedure. Simulated violations of this equal variance assumption about internal sensory noise do not affect the shape of the perceptual scale produced by MLDS [12,13], but may affect discriminability estimates derived from the same perceptual scale [14]. The model assumptions underlying MLDS may also interact with stimulus complexity and dynamic range [11,30], both of which have been shown to affect the perception of gloss [3,31–33]. 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 [5,34]. Such task-dependencies may be particularly relevant when the stimulus property in question ("gloss") consists of multiple appearance dimensions [7,35], 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 [Fig. 2(b)]. Our experiments are therefore analogous to those described by Kingdom [15], who compared scaling and discrimination data from experiments (originally published by Whittle [36,37]) 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. Similarly, if it is assumed that a common transducer function mediates scaling and discriminability of perceived gloss, our results indicate that internal sensory noise remains constant for suprathreshold and near-threshold intervals of specular reflectance. 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 [15,16]. 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 [38]. Weber's law has inspired considerable debate about the transducer functions that relate stimulus and sensation (e.g., [39]), 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 ([38], p. 146). This early observation suggested that a perceptually uniform color space would be a complex mathematical entity, and these complexities were not fully 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 [40,41]. 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 [23,42–46], which therefore means that the validity of any gloss space will be constrained by the viewing conditions chosen for its construction [47]. 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 [48–50], 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 prerequisite 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 [8], and also that such small image differences are not necessarily relevant to the task of scaling material appearance [9]. Our results demonstrate 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 ("DyViTo: Dynamics in Vision and Touch", H2020-MSCA-ITN-2017, Project ID 765121); Deutsche Forschungsgemeinschaft ("Cardinal Mechanisms of Perception", Project ID 222641018, SFB-TRR-135-TP-C1); European Research Council ("SHAPE", ERC-2015-CoG-682859).

Acknowledgment. We would like to thank Saskia L. Honnefeller, 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

1. P. J. Marlow, J. Kim, and B. L. Anderson, "The perception and misperception of specular surface reflectance," Curr. Biol. **22**, 1909–1913 (2012).
2. F. Pellacini, J. A. Ferwerda, and D. P. Greenberg, "Toward a psychophysically-based light reflection model for image synthesis," in *27th Annual Conference on Computer Graphics and Interactive Techniques* (ACM/Addison-Wesley, 2000), pp. 55–64.
3. G. Obein, K. Knoblauch, and F. Viéot, "Difference scaling of gloss: nonlinearity, binocularity, and constancy," J. Vis. **4**(9), 4 (2004).
4. Y. X. Ho, M. S. Landy, and L. T. Maloney, "Conjoint measurement of gloss and surface texture," Psychol. Sci. **19**, 196–204 (2008).
5. J. B. Phillips, J. A. Ferwerda, and A. Nunziata, "Gloss discrimination and eye movements," Proc. SPIE **7527**, 75270Z (2010).
6. F. B. Leloup, M. R. Pointer, P. Dutré, and P. Hanselaer, "Overall gloss evaluation in the presence of multiple cues to surface glossiness," J. Opt. Soc. Am. A **29**, 1105–1114 (2012).
7. R. S. Hunter and R. W. Harold, *The Measurement of Appearance*, 2nd ed. (Wiley, 1987).
8. R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," in *ACM SIGGRAPH 2011 Papers* (ACM, 2011), pp. 40:1–40:14.
9. G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala, "Visual equivalence: towards a new standard for image fidelity," ACM Trans. Graph. **26**, 76 (2007).
10. C. Charrier, K. Knoblauch, L. T. Maloney, A. C. Bovik, and A. K. Moorthy, "Optimizing multiscale SSIM for compression via MLDS," IEEE Trans. Image Process. **21**, 4682–4694 (2012).
11. E. D. Protonotarios, A. Johnston, and L. D. Griffin, "Difference magnitude is not measured by discrimination steps for order of point patterns," J. Vis. **16**(9), 2 (2016).
12. L. T. Maloney and J. N. Yang, "Maximum likelihood difference scaling," J. Vis. **3**(8), 5 (2003).
13. F. A. A. Kingdom and N. Prins, *Psychophysics: A Practical Introduction*, 2nd ed. (Academic, 2016).

14. G. Aguilar, F. A. Wichmann, and M. Maertens, "Comparing sensitivity estimates from MLDS and forced-choice methods in a slant-from-texture experiment," J. Vis. **17**(1), 37 (2017).
15. F. A. A. Kingdom, "Fixed versus variable internal noise in contrast transduction: the significance of Whittle's data," Vis. Res. **128**, 1–5 (2016).
16. F. Devinck and K. Knoblauch, "A common signal detection model accounts for both perception and discrimination of the watercolor effect," J. Vis. **12**(3), 19 (2012).
17. J. A. Ferwerda, F. Pellacini, and D. P. Greenberg, "Psychophysically based model of surface gloss perception," Proc. SPIE **4299**, 4211–4299 (2001).
18. W. Jakob, Mitsuba renderer (2010).
19. J. F. Norman and F. Phillips, Bell Peppers (v1.1) [3D Object Files] (2016).
20. W. J. Adams, J. H. Elder, E. W. Graf, J. Leyland, A. J. Lugtigheid, and A. Muryy, "The Southampton-York natural scenes (SYNS) dataset: statistics of surface attitude," Sci. Rep. **6**, 35805 (2016).
21. H. W. Jensen, "Global illumination using photon maps," in *Eurographics Workshop on Rendering Techniques*, X. Pueyo and P. Schröder, eds. (Springer, 1996), pp. 21–30.
22. D. Geisler-Moroder and A. Dür, "A new ward BRDF model with bounded Albedo," Comput. Graph. Forum **29**, 1391–1398 (2010).
23. R. W. Fleming, R. O. Dror, and E. H. Adelson, "Real-world illumination and the perception of surface reflectance properties," J. Vis. **3**(5), 3 (2003).
24. L. Xia, S. C. Pont, and I. Heynderickx, "The visual light field in real scenes," Iperception **5**, 613–629 (2014).
25. A. Gilchrist, C. Kossyfidis, T. Agostini, X. Li, F. Bonato, J. Cataliotti, B. Spehar, V. Annan, and E. Economou, "An anchoring theory of lightness perception," Psychol. Rev. **106**, 795–834 (1999).
26. E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," ACM Trans. Graph. **21**, 267–276 (2002).
27. J. R. Cheeseman, J. A. Ferwerda, F. J. Maile, and R. W. Fleming, "Supplemental Materials: Scaling and Discriminability of Perceived Gloss," Open Science Framework (2020), https://doi.org/10.17605/OSF.IO/9H75A.
28. J. Peirce and M. Macaskill, *Building Experiments in PsychoPy* (SAGE, 2018).
29. H. H. Schütt, S. Harmeling, J. H. Macke, and F. A. Wichmann, "Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data," Vis. Res. **122**, 105–123 (2016).
30. G. Aguilar and M. Maertens, "Toward reliable measurements of perceptual scales in multiple contexts," J. Vis. **20**(4), 19 (2020).
31. W. J. Adams, G. Kucukoglu, M. S. Landy, and R. K. Mantiuk, "Naturally glossy: gloss perception, illumination statistics and tone mapping," J. Vis. **18**(13), 4 (2018).
32. K. Doerschner, L. T. Maloney, and H. Boyaci, "Perceived glossiness in high dynamic range scenes," J. Vis. **10**(9), 11 (2010).
33. J. B. Phillips, J. A. Ferwerda, and S. Luka, "Effects of image dynamic range on apparent surface gloss," in *Color Imaging Conference* (2009), pp. 193–197.
34. L. T. Maloney and K. Knoblauch, "Measuring and modeling visual appearance," Annu. Rev. Vis. Sci. **6**, 519–537 (2020).
35. M. Toscani, D. Guarnera, G. C. Guarnera, J. Y. Hardeberg, and K. R. Gegenfurtner, "Three perceptual dimensions for specular and diffuse reflection," ACM Trans. Appl. Percept. **1**, 6 (2020).
36. P. Whittle, "Brightness, discriminability and the 'Crispening Effect'," Vis. Res. **32**, 1493–1507 (1992).
37. P. Whittle, "Increments and decrements: luminance discrimination," Vis. Res. **26**, 1677–1691 (1986).
38. G. T. Fechner, *Elements of Psychophysics* (Translated by Adler, H. E.) (Rinehart and Winston, 1966), Vol. **I**.
39. S. S. Stevens, "To honor Fechner and repeal his law," Science **133**, 80–86 (1961).
40. D. L. MacAdam, "Specification of small chromaticity differences," J. Opt. Soc. Am. **33**, 18–26 (1943).
41. T. Smith and J. Guild, "The C.I.E. colorimetric standards and their use," Trans. Opt. Soc. **33**, 73 (1931).
42. Y. X. Ho, L. T. Maloney, and M. S. Landy, "The effect of viewpoint on perceived visual roughness," J. Vis. **7**(1), 1 (2007).
43. J. F. Norman, J. T. Todd, and F. Phillips, "Effects of illumination on the categorization of shiny materials," J. Vis. **20**(5), 2 (2020).
44. P. Vangorp, J. Laurijssen, and P. Dutré, "The influence of shape on the perception of material reflectance," ACM Trans. Graph. **26**, 77 (2007).
45. F. Zhang, H. de Ridder, P. Barla, and S. Pont, "Effects of light map orientation and shape on the visual perception of canonical materials," J. Vis. **20**(4), 13 (2020).
46. J. F. Norman, F. Phillips, J. R. Cheeseman, K. E. Thomason, C. Ronning, K. Behari, K. Kleinman, A. B. Calloway, and D. Lamirande, "Perceiving object shape from specular highlight deformation, boundary contour deformation, and active haptic manipulation," PLoS One **11**, 1–15 (2016).
47. A. Fores, M. D. Fairchild, and I. Tastl, "Improving the perceptual uniformity of a gloss space," in *Color and Imaging Conference* (Society for Imaging Science and Technology, 2014), pp. 7–13.
48. L. L. Kontsevich and C. W. Tyler, "Bayesian adaptive estimation of psychometric slope and threshold," Vis. Res. **39**, 2729–2737 (1999).
49. H. R. Lieberman and A. P. Pentland, "Microcomputer-based estimation of psychophysical thresholds: the best PEST," Behav. Res. Methods Instrum. **14**, 21–25 (1982).
50. A. B. Watson, "QUEST+: a general multidimensional Bayesian adaptive psychometric method," J. Vis. **17**(3), 10 (2017).