Three perceptual dimensions for specular and diffuse reflection

MATTEO TOSCANI, Justus-Liebig-Universität Gießen, Germany
DAR'YA GUARNERA, NTNU - Norwegian University of Science and Technology, Norway
GIUSEPPE CLAUDIO GUARNERA, University of York, United Kingdom and NTNU - Norwegian University of Science and Technology, Norway
JON YNGVE HARDEBERG, NTNU - Norwegian University of Science and Technology, Norway
KARL R. GEGENFURTNER, Justus-Liebig-Universität Gießen, Germany

Previous research investigated the perceptual dimensionality of achromatic reflection of opaque surfaces, by using either simple analytic models of reflection, or measured reflection properties of a limited sample of materials. Here we aim to extend this work to a broader range of simulated materials. In a first experiment, we used sparse multidimensional scaling techniques to represent a set of rendered stimuli in a perceptual space that is consistent with participants' similarity judgments. Participants were presented with one reference object and four comparisons, rendered with different material properties. They were asked to rank the comparisons according to their similarity to the reference, resulting in an efficient collection of a large number of similarity judgments. In order to interpret the space individuated by multidimensional scaling, we ran a second experiment in which observers were asked to rate our experimental stimuli according to a list of 30 adjectives referring to their surface reflectance properties. Our results suggest that perception of achromatic reflection is based on at least three dimensions, which we labelled "Lightness", "Gloss" and "Metallicity", in accordance with the rating results. These dimensions are characterized by a relatively simple relationship with the parameters of the physically based rendering model used to generate our stimuli, indicating that they correspond to different physical properties of the rendered materials. Specifically, "Lightness" relates to diffuse reflections, "Gloss" to the presence of high contrast sharp specular highlights and "Metallicity" to spread out specular reflections.

Additional Key Words and Phrases: Perception, BRDF, dimensionality

ACM Reference Format:
Matteo Toscani, Dar'ya Guarnera, Giuseppe Claudio Guarnera, Jon Yngve Hardeberg, and Karl R. Gegenfurtner. 2020. Three perceptual dimensions for specular and diffuse reflection. *ACM Transactions on Applied Perception* 1, 1, Article 1 (January 2020), 27 pages. https://doi.org/10.1145/3380741

1 INTRODUCTION

In everyday life, we are usually extremely good at visually distinguishing between materials and inferring their properties (for a recent review, see [29]). For example, color vision helps detecting edible fruits against a foliage background [68] or to determine whether a fruit is ripe [90]. Colors are useful to segment a scene in its components, according to their material properties, and to retrieve information from memory [36, 109]. Also, humans can identify a wide range of materials from briefly presented photographs [88, 105, 106] and group them

Authors' addresses: Matteo Toscani, Justus-Liebig-Universität Gießen, Gießen, Germany; Dar'ya Guarnera, NTNU - Norwegian University of Science and Technology , Gjøvik, Norway; Giuseppe Claudio Guarnera, University of York, York, United Kingdom, NTNU - Norwegian University of Science and Technology, Gjøvik, Norway; Jon Yngve Hardeberg, NTNU - Norwegian University of Science and Technology, Gjøvik, Norway; Karl R. Gegenfurtner, Justus-Liebig-Universität Gießen, Gießen, Germany.

XXXX-XXXX/2020/1-ART1 $15.00
https://doi.org/10.1145/3380741

in meaningful categories according to their properties [7, 32, 104]. Visual recognition of materials allows us to determine their mechanical properties, e.g. whether a surface is slippery, cold, or fragile. For instance, softness ratings for static objects systematically depend on the optical properties [80].

Recognising material properties is a difficult task for the visual system. Inferences on the material of a surface are possible because different materials structure light in a characteristic way. However, linking reflected light to the material properties of the reflecting surface is an under-determined problem. There are virtually infinitely many combinations of illumination and surface properties, which produce the same pattern of cone photoreceptor excitations on the retina. In other words, the visual system has only access to the proximal stimulus (the retinal image), and this can be caused by different environmental properties (distal stimulus) [1, 3]. For example, a local edge could correspond to an object boundary, a fold in a surface, the boundary of a shadow, or a change in surface pigmentation [53]. Thus, the correspondence between perception and its physical cause is far from being trivial and subject to investigations for over a century [42, 43]; for a recent review see [29]. Two main approaches have been proposed to address this problem. The so called *inverse optics* approach proposes that in order to recover surface reflection properties (e.g. albedo or gloss), the visual system explicitly estimates and discounts the contributions of illumination and geometry to the observed luminances e.g. [63, 83–85]. For instance, the complex interaction between the appearance of a specular highlight and the perceived surface curvature of a three-dimensional shape was interpreted as evidence that the visual system employs a physical model of the interaction of light with curved surfaces [12] Alternatively, the visual system could bypass the problem of estimating the individual contributions to the reflected light by means of simple image statistics which correlate with perceptually relevant properties of the distal stimulus (e.g. surface albedo or gloss), allowing to directly estimate them (see for review [29, 93]).

The more complex the reflective properties of a surface, the more appealing the need of a simple theoretical approach. For instance, it was proposed that the visual system exploits the correlation between the skewness of the luminance histogram of the light reflected from a surface and the surface's specularity, so that a surface characterized by a positively skewed histogram appears glossy [69, 87]. However, human gloss judgments of a broader set of naturalistic stimuli seemed to correlated more with image contrast than with the skewness of the luminance histogram, and histogram-equalized versions of the same images appeared even glossier than the originals, despite a lack of skewness [104].

Since the visual system is likely to exploit regularities observed in the real world, the complex patterns of natural reflection could provide the visual system the diagnostic image features for directly perceiving material properties [30]. Therefore, naturalistic, complex and physically accurate stimuli are to prefer when investigating perception of material properties, and complex physical models of reflection can be useful to understand perception.

A radiometric description of the way light is reflected by a given material, is given by its Bidirectional Reflectance Distribution Function (BRDF) [73]. BRDF describes the ratio of outgoing radiance reflected from a surface to the incoming irradiance as a function of the incident and exitant directions [37].

Although this formulation allows a systematic account of the behaviour of the light reflected from an opaque surface of a homogenous material, it is highly complex, since every combination of incident and exitant directions over the hemisphere require an individual measurement. Thus, several attempts have been made to describe BRDFs by relatively simple parametric models, often used in computer graphics (for reviews, see [29, 37]). Apart from color, most of these models separate reflectance into diffuse (following Lambert's law [56]) and specular dimensions and have parameters controlling the their angular distributions [14, 29, 37]. Alternative approaches use measured BRDF data, for instance from existing datasets such as MERL [64], UTIA [27] and RGL [23], to convert BRDFs into products of compact non-parametric lower-dimensional factors [6], or to derive low-dimensional embeddings, in which the manifold representation can be used for rendering [89].

Although some of these models are based on appearance, the perceptual relevance of their dimensions have typically not been empirically tested. However, establishing the perceptual dimensionality of the BRDF is the basis for the definition of any perceptually based representation of materials. For instance, concerning color, the infinitely dimensional space of spectral wavelength distributions could be successfully reduced to only three dimensions, the excitation of each of the three classes of cones [42, 65]; see also [41]; for a recent rewiew, see [35]. The transformation of spectra into a three-dimensional representation is the basis for the definition of color spaces.

Research extended beyond color by attempting to individuate the perceptually relevant dimensions of specular reflections. Billmayer and O'Donnell [11], asked participants to judge the difference in gloss between pairs of samples (black, grey or white samples with varying gloss level). They used multidimensional scaling (MDS) techniques to place the stimuli in a metrical space, consistent with the observers' judgements. MDS results suggested that observers did not base their judgements on more than one dimension, indicating that perceived BRDF is essentially low dimensional.

The following investigations proposed two dimensions to explain perception of diffuse and specular reflections. Pellacini *et al.* [25, 81] used MDS to investigate the dimensionality of gloss perception of simulated surfaces and related perceived gloss to the parameters of the model they used for the simulations. Specifically, they used an isotropic version of Ward's model, in which one parameter regulates the diffuse component and two parameters define the energy of the specular component and the spread of the specular lobe [102]. Although the model describes the diffuse and the specular components by three parameters, results showed that observers rated the similarity between simulated surfaces mostly based on two dimensions. A neural substrate for these parameters was recently presented by Nishio *et al.* [75].

Wills a*et al.* [108], applied sparse MDS to similarity judgments between couples of simulated surfaces whose reflections were based on a database of measured BRFS (MERL [64]), rather than parametric BRDF. Two dimensions explained over 95% of the variance of human judgments. They reported two pronounced trends in the two-dimensional space they individuated: one opposing darker to brightest BRDFs, the other one, opposing diffuse BRDFs and those having a strong specular component. As the authors acknowledge, their results depend on the particular surfaces and viewing conditions employed in the experiment. However, despite the fact that different datasets were used, results suggest that the perceptual space of reflectance can be approximated with a one or two-dimensional space. This would predict that presenting participants with an increased variety of materials would again yield a similar one or two-dimensional solution. Alternatively, other dimensions may emerge.

Here, we aim to investigate the perceptual dimensions of achromatic reflection with a broader range of materials than in previous research. We decided to define our experimental stimuli according to a parametric BRDF model rather than using measured BRDFs, in order to expand our range of simulated materials over the one defined by the MERL database used by Wills *et al.* [108]. Although the Ward model is probably the most used model in material perception studies with simulated surfaces (*e.g.* [13, 22, 25, 30, 44, 49, 50, 60, 61, 75, 78, 81, 82, 92, 98, 99, 101, 104, 110]), this model was recently shown to be unable to capture some of the perceptually relevant aspects of the physics of reflection [38, 39]. Thus, we rendered different materials based on a slightly more complex model, characterized by five parameters [19, 37, 57]. This model, referred hereafter as the ABC-model, is thought to be physically accurate despite the relatively small number of parameters [57].

In a first experiment, we used sparse MDS techniques, similar to Wills *et al.* [108], to assign a metrical embedding to our set of stimuli (*i.e.* to place our stimuli in a perceptual space), consistent with a large number of similarity judgments on our samples. Results suggested that perceived similarity was based on more than two dimensions, the perceived specular reflections being represented by two different dimensions. MDS produces metrical spaces (*i.e.* Euclidean embeddings) whose distances between elements are unaffected by rotations or inversions, thus the axes can be arbitrary rotated obtaining equivalent solutions, and the interpretation of a chosen set of axes might not be straightforward. Thus, in order to interpret the MDS results, we ran a second experiment in which

observers were asked to rate our experimental stimuli according to a list of 30 adjectives referring to their surface reflectance properties. We applied PCA (Principal Components Analysis) on the ratings, and examinated adjectives dominating the first three dimensions, which we could therefore interpret as mostly related to “gloss”, “lightness” and “metallicity”. Crucially, these stimuli along these three MDS dimensions were characterized by different physical properties, as indicated by the parameters of the ABC-model. Furthermore, stimuli which scored high on the MDS embedding dimension aligned with the “gloss” PCA dimension, were characterized by high image contrast, whereas the ones scoring high on the dimension aligned with the “metallicity” component, by high skewness of the luminance histogram.

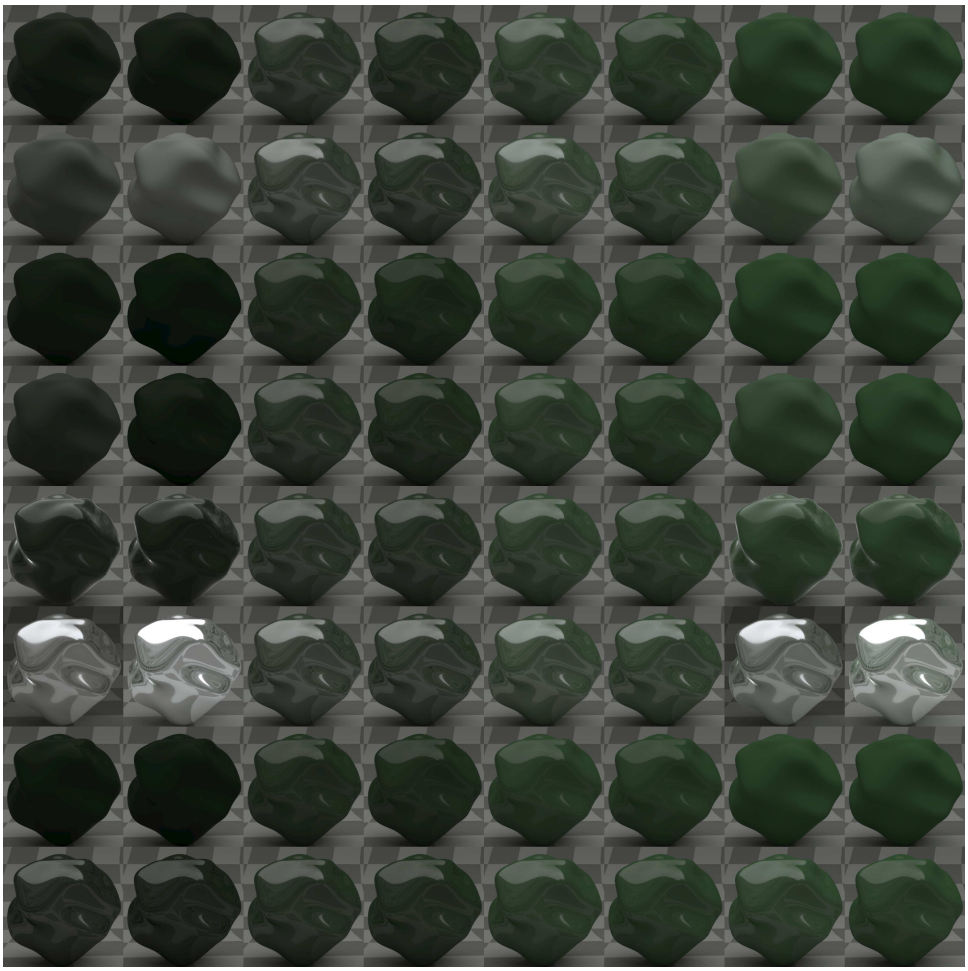

Fig. 1. Experimental Stimuli. The corresponding parameters of the ABC-model are reported in Table 1. Contrast is enhanced to improve visibility.

2 GENERAL METHODS

2.1 Stimuli

We rendered five different three-dimensional "blobby" shapes using 64 different materials, as defined by different combination of parameters of the ABC-model, for a total of 320 renderings. Each object was embedded in a different scene. In each scene, the object was laying on the floor, with a wall behind it. The floor and the wall were covered with a checkerboard pattern (see Figure 1). The scenes were illuminated by a complex naturalistic light probe converted into grey scale (Courtyard of the Doge's palace, Venice, Italy light probe [33]). We used naturalistic illumination because it was shown that humans are better in judging material properties under these circumstances [30]. The images were rendered with Mitsuba renderer [103], at a resolution of 512×512 pixels.

2.2 Shapes

The shapes were generated using sinusoidal perturbations of spheres. This family of shapes has been extensively used in vision research before [16–18, 31, 70, 71, 76, 99]. We generated 5 instances of classes of shapes, as defined by the amplitude and frequency of their sinusoidal components (Fig. 1). The generation algorithm starts with a sphere, and recursively applies 5 sinusoidal perturbations to its vertices. The perturbation Amplitude is defined as percentage of the sphere radius, the Frequency of each perturbation is defined by the number of cycles of the sinusoid within the sphere. For the five shapes, the frequency (F) and amplitude (A) parameters are (F=2, A=3%), (F=3, A=4%), (F=4, A=3%), (F=6, A=3%), (F=7, A=2%).

In order to prompt observers to base their similarity judgments on perceived material properties, in every trial the five different surfaces were presented with different shapes, so that an image pixel-wise comparison was not possible.

2.3 BRDF model

The ABC-model is a variation of the physically-based Cook-Torrance model [20]. The model accounts for both specular and diffuse reflections (modelled according to Lambert's law [56]). Modelling of the specular reflections is based on the microfacets theory [96]. This theory assumes that surfaces that are not perfectly smooth consist of a collection of small mirror-like *microfacets*. The microfacets are not perfectly aligned, with their orientation being described by a statistical distribution, which explains why shiny objects can show blurred specular highlights, instead of a perfectly specular reflection (mirror—like).

Moreover, the model includes a term accounting for the Fresnel effect [34], which has a perceptually relevant impact on material appearance [38, 39]. In Figure 2 we report a comparison illustrating the difference between a rendering of a sphere accounting for the Fresnel effect and without accounting for it.

In the ABC model, the distribution of micro facets is modelled by three parameters, which therefore characterize the rough specular term: the **amplitude** (**A**), the **width** of the specular peak (**B**) —large values correspond to sharp specular highlights—, and the **falloff rate** (**C**) of the specular lobes. The **Fresnel term** is controlled by means of the **Index of Refraction** (**IOR**) parameter; please note that the *IOR* defined in the ABC model does not have a direct physical interpretation, since its range does not correspond to real materials. Instead, its role is simply to modulate the Fresnel effect. Finally, another parameter weights the **diffuse reflection** term (**KD**), representing Lambertian reflections. Overall, the model is described by the following equation:

$$f(V_i, V_r) = \frac{KD}{\pi} + \frac{A}{(1 + B(1 - \langle H, N \rangle))^C} \cdot \frac{Fresnel(IOR, \Theta_h) \cdot G(V_i, V_r, N)}{\langle V_i, N \rangle \cdot \langle V_r, N \rangle} \tag{1}$$

where V_i and V_r are the incoming and exitant directions, N is the surface normal at a given point on the surface, H is *half-way vector - i.e.* the normalized vectorial sum of V_i and V_r -, Θ_h is the angle between N and H, G models masking and self-shadowing among microfacets and $\langle \cdot, \cdot \rangle$ represents the dot product of two vectors.

We defined 64 different materials spanning the gamut of the five ABC model parameters. The region of the space within the gamut is defined as the combination of parameters for which energy was conserved. Energy conservation assumes that the energy reflected cannot exceed incident energy:

$$\forall v_i, \int_{\Omega_+} f(V_i, V_r)\langle V_r, N\rangle d\omega_r \leq 1 \tag{2}$$

where Ω_+ is the unit hemisphere over the surface. In our implementation of the ABC model, we experimentally verified that energy conservation holds under the following constraints:

$$\begin{aligned}
&0 \leq A \leq 20000;\\
&(4.85 \cdot A + 0.4021) < B < (-0.012 \cdot A^2 + 250.6 \cdot A + 180);\\
&(0.4272 \cdot A^{0.1216}) \leq C \leq (2.309 \cdot A^{-0.04639});\\
&1.01 \leq IOR \leq min(9.31, 2.355 \cdot C^{8.097} + 2.1).
\end{aligned}$$

Figure 1 shows the full set of images, arranged by increasing **diffuse reflection** (**KD**, from left to right) and **amplitude** of the specular lobe (**A**, from up to down). Although perceived albedo increases from left to right and gloss from top to bottom, it is visible that this arrangement is far from fully accounting for the perceptual differences between our images. For example, some images in the top are clearly more specular than others at the bottom, presumably because of the effect of other parameters than the **amplitude** of the specular lobe.

Figure 3 depicts the 3D slice of the material gamut in the ABC cube, which extends beyond the portion occupied by the selected 64 measured BRDFs from the MERL database, as fitted with the ABC-model parameters. These measured BRDFs have been selected out of the 100 available in the database, by removing materials for which the fitting provides unstable results (such as very anisotropic samples, like chrome-steel or very specular, such as chrome) and fabric materials, for which the BRDF is not sufficient to faithfully describe the appearance, rather influenced by the geometry of the threads and the translucency [37, 40].

In Figure 4 we report the 3D section of the ABC model, relative to the **A**, **C**, **IOR** parameters, and the coordinates of the 64 MERL points as fitted in the model. As a comparison, the gamut of the Ward model would be a subset of the plane with equation **IOR** = 1. The lack of the Fresnel term causes the Ward BRDF model to diverge in appearance from real materials [39, 72], thus highlighting the importance of using a physically-based model, such as the ABC, in perceptual experiments. Overall, the space of reflectances covered in our experiments is clearly wider than the ones in the MERL materials and the gamut of the Ward BRDF gamut (see Figure 3 and Figure 4).

Please note that, once fitted to the ABC model, the appearance of some measured materials in the MERL dataset might not be faithfully reproduced [46]; similar considerations apply to any current analytical BRDF model [37]. Several works have pointed out a number of measurement artifacts in the MERL database, such as optical aberrations, discontinuities, and extrapolation artifacts for data above 75 degrees [24, 45, 66], which cause the measured materials in the MERL dataset to differ from the physical counterpart. Such artifacts explain in part the difficulty to faithfully reproduce the appearance of some materials in the MERL dataset, and highlight the importance of using a gamut beyond the one defined by the dataset itself.

2.4 Apparatus

Images were displayed on an Eizo CG223W 10 bit LCD monitor, which was connected to a Dell Precision 380 computer (Dell Inc., Round Rock, Texas, USA). The images were rendered in linear RGB and gamma corrected according to the gamma functions of each of the red, green and blue channels of the monitor. The experiment was controlled using the Psychophysics Toolbox [52] in MATLAB (The Mathworks Inc., Natick, MA, USA). The CIE-xyY chromatic coordinates of the monitor RGB channels were $R = [0.653; 0.333; 35.0]$, $G = [0.204; 0.677; 64.1]$,

r	c	KD	A	B	C	IOR	r	c	KD	A	B	C	IOR
1	1	0.015	0.44	2.5361	0.38661	1.1974	1	1	0.105	2500.33	147030.9761	1.2312	3.2239
2	1	0.015	0.44	2.5361	0.38661	2.1011	2	1	0.105	2500.33	147030.9761	1.2312	7.277
3	1	0.015	0.44	290.2617	0.38661	1.1974	3	1	0.105	2500.33	416838.9232	1.2312	3.2239
4	1	0.015	0.44	290.2617	0.38661	2.1011	4	1	0.105	2500.33	416838.9232	1.2312	7.277
5	1	0.015	10000	48500.4021	1.3093	1.1974	5	1	0.105	7500.11	328453.8932	1.3298	3.2239
6	1	0.015	10000	48500.4021	1.3093	9.3036	6	1	0.105	7500.11	328453.8932	1.3298	7.277
7	1	0.015	10000	1306180	1.3093	1.1974	7	1	0.105	7500.11	912609.8083	1.3298	3.2239
8	1	0.015	10000	1306180	1.3093	9.3036	8	1	0.105	7500.11	912609.8083	1.3298	7.277
1	2	0.015	0.44	2.5361	2.3986	1.1974	1	2	0.105	2500.33	147030.9761	1.4812	3.2239
2	2	0.015	0.44	2.5361	2.3986	9.3036	2	2	0.105	2500.33	147030.9761	1.4812	7.277
3	2	0.015	0.44	290.2617	2.3986	1.1974	3	2	0.105	2500.33	416838.9232	1.4812	3.2239
4	2	0.015	0.44	290.2617	2.3986	9.3036	4	2	0.105	2500.33	416838.9232	1.4812	7.277
5	2	0.015	10000	48500.4021	1.5061	1.1974	5	2	0.105	7500.11	328453.8932	1.4608	3.2239
6	2	0.015	10000	48500.4021	1.5061	9.3036	6	2	0.105	7500.11	328453.8932	1.4608	7.277
7	2	0.015	10000	1306180	1.5061	1.1974	7	2	0.105	7500.11	912609.8083	1.4608	3.2239
8	2	0.015	10000	1306180	1.5061	9.3036	8	2	0.105	7500.11	912609.8083	1.4608	7.277
1	3	0.055	2500.33	147030.9761	1.2312	3.2239	1	3	0.135	0.44	2.5361	0.38661	1.1974
2	3	0.055	2500.33	147030.9761	1.2312	7.277	2	3	0.135	0.44	2.5361	0.38661	2.1011
3	3	0.055	2500.33	416838.9232	1.2312	3.2239	3	3	0.135	0.44	290.2617	0.38661	1.1974
4	3	0.055	2500.33	416838.9232	1.2312	7.277	4	3	0.135	0.44	290.2617	0.38661	2.1011
5	3	0.055	7500.11	328453.8932	1.3298	3.2239	5	3	0.135	10000	48500.4021	1.3093	1.1974
6	3	0.055	7500.11	328453.8932	1.3298	7.277	6	3	0.135	10000	48500.4021	1.3093	9.3036
7	3	0.055	7500.11	912609.8083	1.3298	3.2239	7	3	0.135	10000	1306180	1.3093	1.1974
8	3	0.055	7500.11	912609.8083	1.3298	7.277	8	3	0.135	10000	1306180	1.3093	9.3036
1	4	0.055	2500.33	147030.9761	1.4812	3.2239	1	4	0.135	0.44	2.5361	2.3986	1.1974
2	4	0.055	2500.33	147030.9761	1.4812	7.277	2	4	0.135	0.44	2.5361	2.3986	9.3036
3	4	0.055	2500.33	416838.9232	1.4812	3.2239	3	4	0.135	0.44	290.2617	2.3986	1.1974
4	4	0.055	2500.33	416838.9232	1.4812	7.277	4	4	0.135	0.44	290.2617	2.3986	9.3036
5	4	0.055	7500.11	328453.8932	1.4608	3.2239	5	4	0.135	10000	48500.4021	1.5061	1.1974
6	4	0.055	7500.11	328453.8932	1.4608	7.277	6	4	0.135	10000	48500.4021	1.5061	9.3036
7	4	0.055	7500.11	912609.8083	1.4608	3.2239	7	4	0.135	10000	1306180	1.5061	1.1974
8	4	0.055	7500.11	912609.8083	1.4608	7.277	8	4	0.135	10000	1306180	1.5061	9.3036

Table 1. ABC model parameters of the 64 materials used in the experiments. “r” and “c” refer to the rows and columns of the examples in Figure 1.

and $B = [0.151; 0.065; 7.8]$. Gamma curves and chromatic coordinates were measured with a Konica Minolta CS-2000A Spectroradiometer (Konica Minolta Sensing Inc., Singapore). Participants were seated in a dark room, with their heads stabilized by a chinrest with 38 cm distance between the forehead and the centre of the screen. For both the experiments, the scenes were superimposed on a grey background (CIE-xyY = $[0.331; 0.334; 53.45]$).

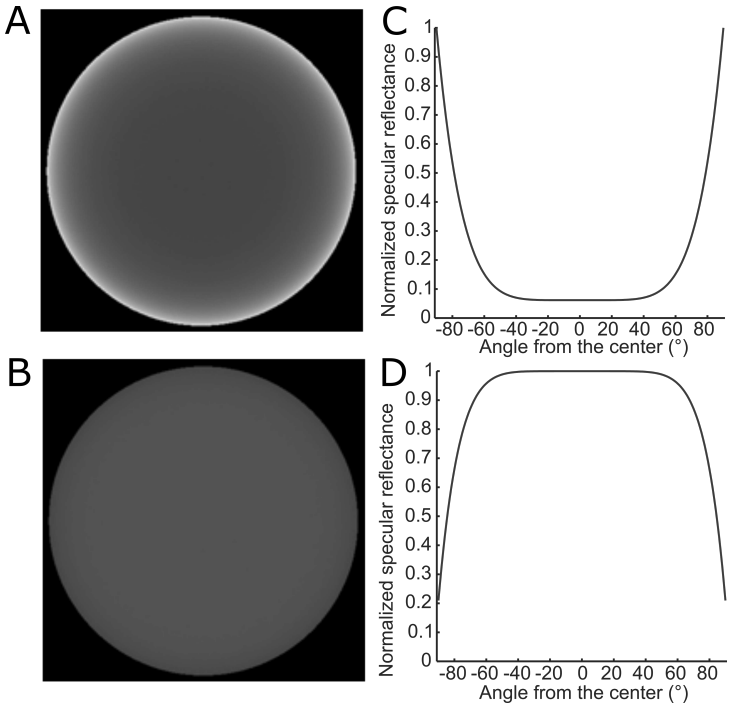

Fig. 2. (A,B) Two spheres rendered under uniform white illumination, using a microfacet BRDF model accounting for the Fresnel effect (A), and without accounting for it (Ward model, B). C,D) Specular reflectance, moving along a diameter of the sphere. When accounting for the Fresnel effect, the reflectance increases moving away from the centre, such as in real world dielectric materials (C), whereas using the Ward model the reflectance decreases (D).

3 EXPERIMENT 1: MULTIDIMENSIONAL SCALING

3.1 Participants

Eight students from the Justus—Liebig University of Giessen volunteered to take part in the experiment. All volunteers were naïve to the purpose of the experiment, and they had normal or corrected-to-normal visual acuity. They all provided written informed consent in agreement with the Declaration of Helsinki, and all procedures were approved by the local ethics commission of the department (approval number 2017—0030). Volunteers were reimbursed for their participation.

3.2 Procedure

In each trial, observers were presented with five rendered scenes, each depicting a blobby shape on a checkerboard—patterned surface. Four of these scenes were presented in a 2×2 arrangement on the left of the screen, representing the comparison shapes. The reference shape was presented on the right side of the screen see Figure 5. Observers ran 600 experimental trials; in each of which five different materials were chosen from all 7624512 possible five—elements combinations of the 64 materials shown in see Figure 1. Since we decided to replicate the methods of Wills and colleagues [108], this choice was randomized, with no replacement. However, adaptive sampling methods could have optimized information gain, e.g. [55]. The correspondence between the five materials and the five shapes was randomized, as well as the allocation of each comparison to the four positions on the left side of the screen.

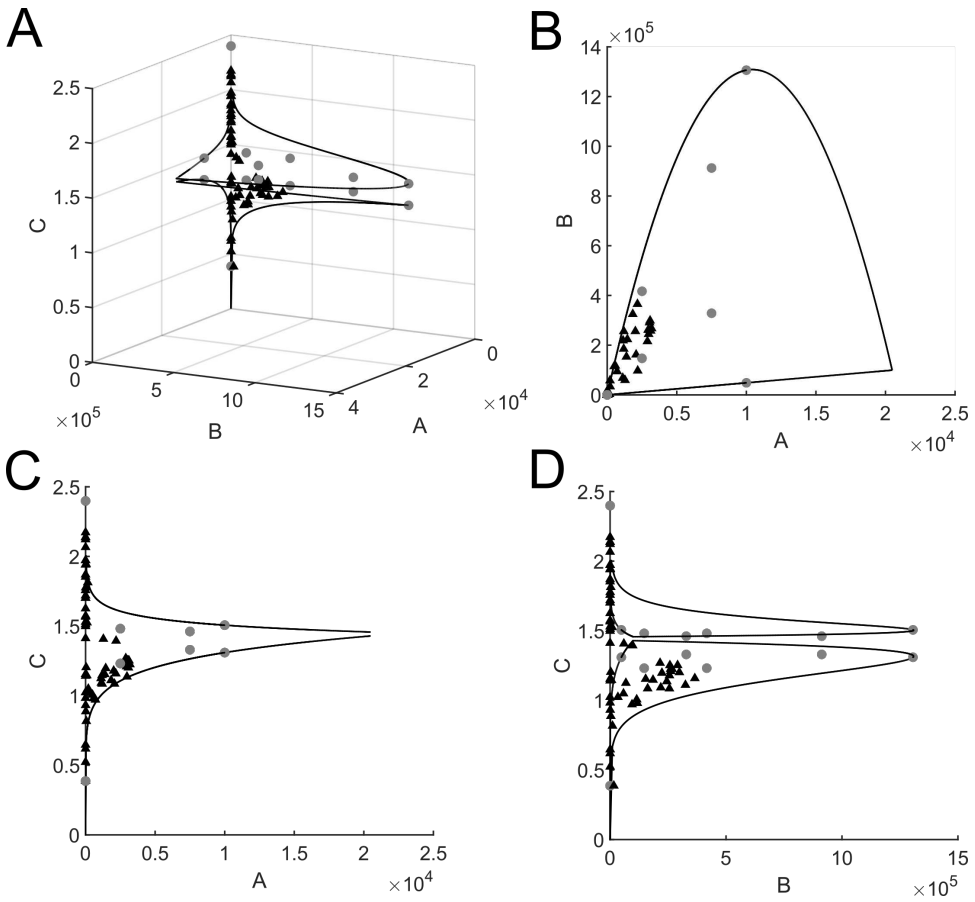

Fig. 3. ABC model gamut. A) View of the 3D slice of the ABC model relative to the A, B and C parameters. The triangles represent the coordinates of the 64 MERL materials used to tune the range of parameters used in this study. The solid black lines enclose the range of parameters for which our implementation of the model conserves energy; the gray dots represent the combination of parameters used in this study. B-D) Projections on the AB, AC and BC planes respectively.

Participants were instructed to select with the mouse the comparison shape that was most similar in terms of surface material properties to the ones of the reference shape. After selecting one shape, that shape disappeared and the participant had again to select the comparison which was most similar to the reference, among the remaining three. Then, this procedure was iterated one more time, so that all the comparisons were ranked in terms of similarity with the reference. Thus, for each trial we obtained the four couples of the reference with each of the comparisons, ranked based on the similarity between reference and comparison.

3.3 Multidimensional scaling

We adopted a sparse generalized non-metric multidimensional scaling algorithm [2, 108]. This procedure uses paired comparisons between three elements (which of two elements is closer to a third one) to assign each element to a $n - dimensional$ Euclidean embedding that is consistent with the comparisons. An advantage of MDS based on similarity judgments, is that there is no need for the experimenter to provide verbally labelled categories or

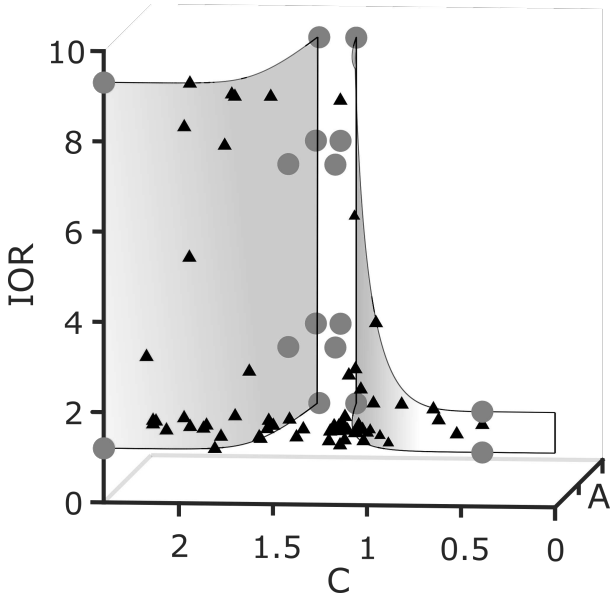

Fig. 4. ABC model gamut. View of the 3D slice of the ABC model relative to the *A*, *C* and *IOR* parameters. The black triangles represent the coordinates of the 64 MERL materials used to tune the range of parameters used in this study. The two shaded surfaces enclose the ABC model gamut. The parameters used in this study are indicated by the grey circles. As a comparison, the Ward gamut is limited to a subset of the plane with equation $IOR = 1$.

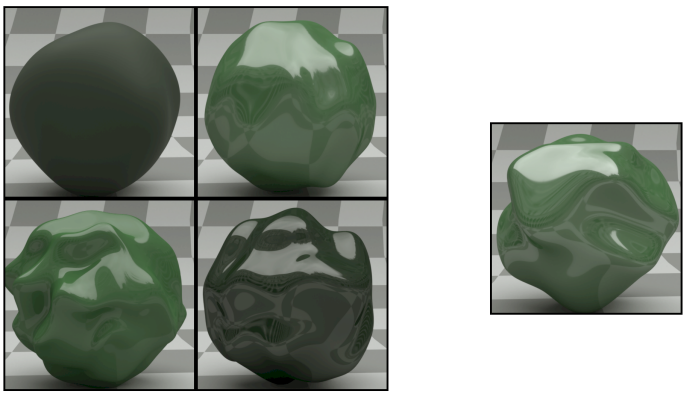

Fig. 5. Multidimensional Scaling experiment. Four comparison shapes were presented on the left, one reference on the right. The five shapes always had a different geometry and different material properties. Contrast is enhanced to improve visibility.

dimensions to participants, which might bias the results. In fact, participants presumably use whatever cues they have available to judge the similarity between pairs of stimuli.

In our paradigm, we can derive six paired comparisons from the four ranked comparison-reference couples in each trial. Thus, the MDS algorithm could make use of 3600 paired comparisons (6 rankings × 600 trials) for each participant.

We used the Generalized Non-metric Multidimensional Scaling Toolbox 1.0 [2]. Briefly, given n elements to assign to a Euclidean metrics according to a number of paired comparisons, the goal of the MDS algorithm is to find an embedding matrix X which explains the paired comparisons represented by the following inequality:

$$||x_i - x_r|| < ||x_j - x_r||, \tag{3}$$

with $(x_i, x_j, x_r) \in X$ and x_r being the reference stimulus, x_i the one comparison judged as closer to the reference, and x_j the other comparison in the paired comparison triplet. Each of the elements of the triplet (x_i, x_j, x_r) belongs to a different row of X, whose columns represent the dimensions of the embedding. The algorithm finds a Gram matrix $K = X^T X$, which is then decomposed to recover X. Since not all the inequalities are necessarily consistent with a possible embedding, for each comparison the slack variable ϵ_{ijr} is added to allow for inequality violation. Thus, the problem of finding X could be reduced to an optimization problem directed to minimize the total amount of slack. The maximum dimensionality of X is the number of elements n.

However, the embedding of the smallest possible dimensionality is to prefer, because of reducing complexity and in order to avoid overfitting. Thus, in the minimization process the rank of K (which is the same as the rank of X) is penalized, so that linear dependencies between the columns are introduced and dimensionality reduction is feasible:

$$\arg\min_{K,\epsilon} \sum_{ijr} \epsilon_{ijr} + \lambda rank(K), \tag{4}$$

with the inequalities (including their slack terms) being the constraints for the minimization. Crucially, a positive scalar lambda (λ) controls the penalization of the rank of K: the higher λ, the lower rank is preferred, decreasing the complexity (and thus the dimensionality) of X. In order to choose λ, we run a k-fold cross-validation. Thus, for 50 values of $\lambda = [02.5]$, we randomly partitioned our set of paired comparisons in k equally sized subsamples, and iteratively fitted an embedding X to each of the subsamples.

For each fitted subsample s, we computed a *training error*, as the proportion of inequalities from s violated by the fitted X. *Testing error* was computed in the same way, based on the inequalities from the remaining subsamples. *Training error* tend to increase with increasing λ, because of the lower complexity of X. Ideally, *testing error* initially decreases with increasing λ, because the reduction of complexity of X improves the generalization of the fit. After a certain value of λ, *testing error* increases again, because the fitted X is too simple to properly fit the data. Thus, in our analyses we chose the value of λ which produced the minimum *testing error*.

We first fitted an embedding to all participants after leaving one our and used it to predict the responses of the left out participant. The prediction errors were similar to the training errors, suggesting generality across participants. Given that participants tended to be consistent, we fitted a single embedding to the paired comparisons triplets pooled across participants.Since for every participant we had a total of 3600 paired comparisons, we fitted the final embedding on 28800 paired comparisons, nearly the double of the number used by Wills and colleagues [108]. After that, we computed the explained variance by each of the dimensions of the embedding and restricted our further analyses and interpretation to its main three dimensions (*i.e.* the ones which explained more variance). Since the estimation of the embedding is based on distances, which are unaffected by rotations or inversions, finding meaningful axes is a subsequent problem.

4 RESULTS

Figure 6 shows the results for one example participant. According to 5—folds cross—validation, *training error* (red line, Figure 6A) increases with λ, as a consequence of reducing the complexity of the embedding, which with increasing λ becomes too simple to properly fit the data. As expected, *testing error* (blue line, Figure 6A) initially

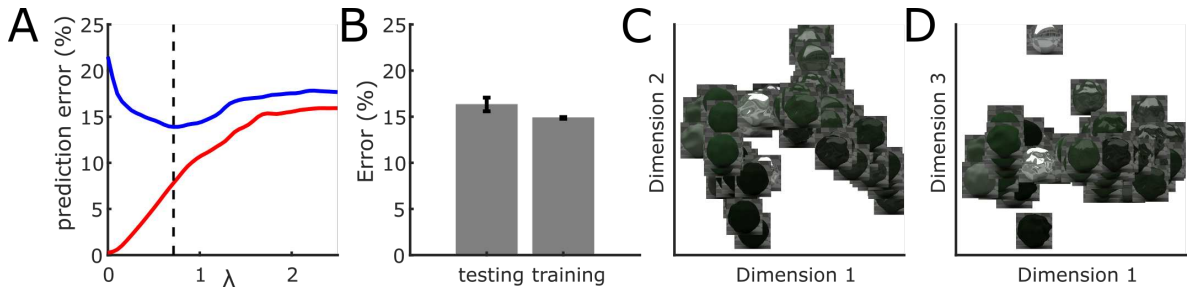

Fig. 6. MDS results for one participant and consistency between participants. A) Cross-validation for fitting the embedding to the responses of participant 1. λ parameter on the x-axis, prediction error on the y-axis. Red line represents training error, blue line testing error. Black dashed vertical line indicates the value of λ at which the testing error reaches its global minimum. B) Consistency between participants. Bars indicate the testing error and the testing error averaged across iterations (y-axis). Error bars depict the standard error of the mean.C-D) Arrangement of the stimuli according to the three main dimensions of the embedding fitted to the data from participant 1. Units are arbitrary but comparable between dimensions.

decreases with increasing λ, indicating that the low *training error* for the smallest λ values was due to overfitting. After reaching a minimum, testing error increases, presumably because the model is too simple to account for the data. The λ value at this minimum (indicated by the black dashed line, Figure 6A) is chosen to fit the embedding.

The first three dimensions of the fitted embedding are shown in Figure 6C-D. Dimension 1 seems to arrange the stimuli from matte to shiny (Figure 6C-D, x-axis). Dimension 2 seems to represent albedo, with low values for dark surfaces Figure 6C, y-axis). Dimension 3 seems related to specularity, with high values for broad specular highlights. However, the rotation of these axis is arbitrary and needs to be appropriately chosen.

The different observers showed similar patterns. We checked between—observers consistency by fitting the embedding to the pooled data of all the observers after excluding one and computing the testing error on her/his data. We iterated this procedure for each observer and averaged the testing error across observers. Training error was computed for each fit, and averaged across iterations. Figure 6B shows averaged *training* and *testing error*.

We note that the *testing error* is rather low (~16%) and close to the *training error* (~15%), indicating that the structure of the embedding generalizes well across observers. In fact, *testing* and *training* error did not statistically differ on average ($t(7) = 1.819$, $p = 0.112$). Given the high consistency, we fitted the embedding on the pooled data across all observers. Results are shown in Figure 7.

Again, *training error* (red line), Figure 7A increases with λ and *testing error* (blue line), Figure 7B presents a non-monotonic trend with a clear global minimum, indicating the best trade-off between over-fitting and simplicity of the model. Figure 7B shows the explained variance of the first 15 dimensions of the embedding defined by the chosen λ (dashed black line in Figure 7A). The first three dimensions explain 85% of the variance. This is less than in previous results [108], where over 95% of the variance was contained within the first two dimensions. In order to include 95% of the variance of our fitted embedding we would need to include at least six dimensions.

This result suggests that a consistent portion of the variability in human similarity judgments on material properties of opaque surfaces is not included within two dimensions. This could be due to previous studies [108] not using such an extensive data set. Although more than three dimensions might be perceptually meaningful, we focus our subsequent analyses and discussion on the first three dimensions of the embedding, so that each of them individually explain a consistent portion of variance (63%, 14%, 9%). With respect of these first three dimensions, it is worth noting that the arrangements of the stimuli according to the embedding fitted to the pooled data (Figure 7C-D) is very similar to the one obtained by fitting the embedding to each single participant (Figure 6C-D), suggesting generality, as indicated by previous analysis. However, the axes of the fitted embedding

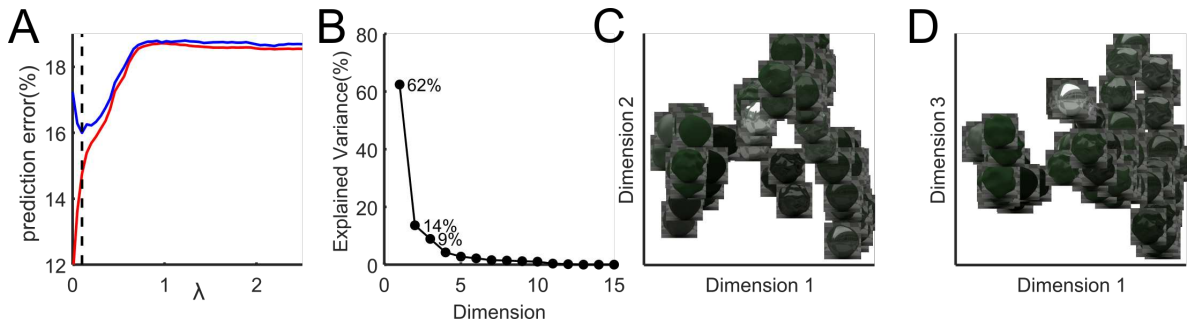

Fig. 7. MDS results for the data pooled across participants. A) Cross-validation. λ parameter on the x-axis, prediction error on the y-axis. Red line represents training error, blue line testing error. Black dashed vertical line indicates the value of Îż at which the testing error reaches its global minimum. B) Explained variance of the embedding dimensions on the y-axis, dimensions ordered by explained variance on the x-axis. Numerical labels next to the data-points indicate the explained variance of the first (62%), the second (14%) and the third main dimensions (9%). C-D) Arrangement of the stimuli according to the three main dimensions of the embedding. Units are arbitrary but comparable between dimensions.

need to be rotated in a meaningful way. In order to do so, we ran a rating experiment with the aim of collecting verbal labels for our stimuli.

5 EXPERIMENT 2: RATINGS

Observers were asked to rate our experimental stimuli according to a list of adjectives referring to their surface reflectance properties. Since the meanings of some adjectives partially overlap, ratings for different adjectives were expected to correlate. We applied a PCA on the ratings in order to express the ratings in a set of orthogonal dimensions. In particular, we were interested in the first three components, in order to use them to interpret the MDS results.

5.1 Adjectives

We selected a list of thirty German adjectives (see Table 2) by informally interviewing five German native speakers. They were presented with a version of Figure 1 in which the images were not systematically ordered, and they had to produce a number of adjectives that could be used to describe the surface properties of the shapes in the figure. We did not ask for a specific number of adjectives, so that the participants' only concern was to propose potential descriptors. Even though we asked for surface properties, some of the adjectives were about the objects' shapes, and thus they were removed (e.g. bumpy, uneven, dented, round). Also, idiosyncratic words like "mercurial" were removed, yielding in total a list of thirty adjectives. After identifying a list of several potential material appearance descriptors, to reduce their number, Serrano and colleagues [86] asked a number of participants which ones did or did not apply to a set of materials. Based on the answers, they could select a relatively small list of representative and intuitive material descriptors to develop a representation for material appearance. In our study we aimed to associate a semantic meaning to the MDS dimensions we individuated. Therefore, these dimensions do not have to be associated to single adjectives, rather, each of them would be related to all the adjective with different weight, helping us in the interpretation of the MDS results.

None of the five people who participated to the interview took part in Experiment 1 nor in Experiment 2. The list of adjectives partially overlapped with the one from [86]: mid-level features like soft, hard, matte, glossy, bright or rough are present in both lists. However, although we selected a higher number of descriptors, our method did not produce high-level features like plastic-like, fabric-like, ceramic-like, tint of reflections, strength of reflections or sharpness of reflections. This might be because in [86], the initial list of features was chosen

by the authors, whereas we let naïve participants propose descriptors, probably missing technical features like strength or sharpness of reflections.

5.2 Participants

Eight students from the Justus—Liebig University of Giessen volunteered to take part in the experiment. All volunteers were naïve to the purpose of the experiment, and they had normal or corrected—to—normal visual acuity. They all provided written informed consent in agreement with the Declaration of Helsinki.

5.3 Procedure

Observers were presented with each of the combinations between the 64 rendered materials and the 30 adjectives, in random order, for a total of 1920 trials. The task was to indicate (by moving a slider controlled with the mouse) how much each adjective was appropriate to describe the surface of the shape presented on the screen, from 0% to 100%. To complete the experiment, participants took two sessions of approximately 1.5 hours each. The shape was randomized in each trial, among the five different geometries we used to render our 64 materials.

5.4 Analysis

As a first step, we converted the ratings for the different adjectives and different participants into a common scale, by z—transforming the ratings of each adjective across materials, separately for each participant. Then, we assessed between-participants consistency by computing Pearson's correlation coefficients between the ratings of each of the 28 couples of participants. We then used PCA to transform the ratings, averaged across participants, into a set of orthogonal dimensions each given by a linear combination of the ratings for the different adjectives (with different weights for the different adjectives). We selected the three dimensions that explain the most variance of the full dataset, and used them as a reference to orient the axis of the selected three dimensions of the MDS embedding. Specifically, we used *Procrustes superimposition* [48] to align the dimensions of the embedding to the ones obtained with PCA on the ratings. Alignment was achieved by translating, rotating and uniformly scaling the MDS embedding dimensions. These transformations did not change the performance of the embedding in accounting for the similarity judgments, *i.e.* they produced an equivalent embedding.

After aligning, we computed the correlation between each dimension of the MDS embedding and the corresponding PCA dimension (in terms of explained variance). This correlation reflects how much the two experiments individuated similar dimensions. In order to assess the statistical significance of these correlations we did a permutation test by computing the correlation coefficient for each pair of dimensions 1000 times, for each of which the correspondence between ratings and materials was randomized. This procedure allowed to estimate the distributions of the correlation coefficients under the null—hypotheses that materials occupy different positions (chosen at random) along the dimensions of the embedding and the PCA dimensions. The alternative hypothesis is that they are placed on a similar space, e.g. a material scoring high on the first PCA dimension would also score high on the first dimension of the embedding. Finally, we labelled the PCA dimensions according to the adjectives which weighted more on them, and used these labelled dimensions to interpret the corresponding dimensions of the MDS embedding.

6 RESULTS

Ratings were overall consistent across participants, as indicated by the correlation between the ratings of each couple of participants (mean Pearson's r= 0.523, range=[0.385 0.646]). PCA shows that it is possible to explain 93% of the variance of the ratings with the first three principal components. The weights of each adjective on these components are reported in Table 2.

Adjectives	1st PC (gloss)	2nd PC (lightness)	3rd PC (metallicity)
Blurry (Unscharf)	-0.12131	-0.10341	0.034326
Bright (Leuchtend)	0.177281	0.170038	**0.226158**
Brilliant (Brilliant)	0.2288	0.033156	0.051862
Cold (Kalt)	0.152984	-0.19037	0.169742
Colorfull (Bunt)	0.036302	**0.372449**	-0.27486
Colorless (Farblos)	-0.02973	**-0.38873**	0.293367
Dark (Dunkel)	-0.05378	**-0.48295**	**-0.31523**
Distorted (Verzerrt)	0.107422	-0.01631	0.052818
Dull (Stumpf)	-0.1889	-0.08562	0.150405
Fluent (Fliessend)	0.144624	0.023092	-0.01093
Glassy (Gläsern)	**0.249675**	-0.00495	-0.09853
Gleaming (Schimmernd)	**0.256094**	0.042191	0.000346
Glittering (Glitzernd)	0.234546	0.02902	0.10361
Glossy (Glänzend)	**0.256592**	0.018187	-0.02146
Gummy (Gummiert)	-0.16155	0.086637	0.03933
Hard (Hart)	0.202152	-0.12567	0.026314
Lacquered (Lackiert)	**0.251072**	-0.03231	-0.1655
Light (Hell)	0.060371	**0.41282**	**0.533121**
Matte (Matt)	**-0.26153**	0.048376	0.136152
Metallic (Metallisch)	0.200586	-0.04029	**0.243132**
Nacreous (Perlmuttartig)	0.174533	0.049995	-0.06608
Natural (Natürlich)	-0.09454	**0.3035**	**-0.34663**
Patterned (Strukturiert)	0.044206	0.075896	-0.04375
Rough (Rau)	-0.11454	-0.0042	0.062888
Silky (Seiden)	0.240748	0.044459	-0.10473
Soft (Weich)	-0.20495	0.118343	0.025597
Specular (Spiegelnd)	**0.257036**	0.022359	0.039565
Velvety (Samten)	-0.20381	0.072783	0.113802
Warm (Warm)	-0.11906	**0.260496**	-0.17326
Wet (Nass)	0.193906	0.010851	-0.17524

Table 2. PCA results. Weights of each adjective on the first three principal components. The adjectives, presented in German for the experiment, are translated here into English. The highest 20% (in absolute value) of the weights is marked bold.

The first component is characterized by high positive weights of "Specular", "Glossy", and "Gleaming", "Lacquered", "Glassy", and the highest negative weight (in absolute value) of "Matte"; therefore we labelled it as **Gloss**. The second component opposes the high positive weight of "Light" to the high negative weight of "Dark", therefore we labelled it as **Lightness**. In addition, the **Lightness** component opposes high positive weight for "colourful" and high negative weight for "colorless". This is consistent with the idea that the lightness dimension relates to the diffuse reflection component of the stimuli. In fact, all the chromatic information was represented by the albedo parameter; thus when the albedo was low, the stimuli tended to look achromatic. Also, the **Lightness** component has positive weights for the adjectives "warm" and "natural". The third principal component has high positive weights of "Light", "Metallic" and "Bright", and high negative weights for "Natural" and "Dark". Crucially, the weights for "Glossy", "Specular" and "Gleaming" are relatively close to zero, as opposed to the ones

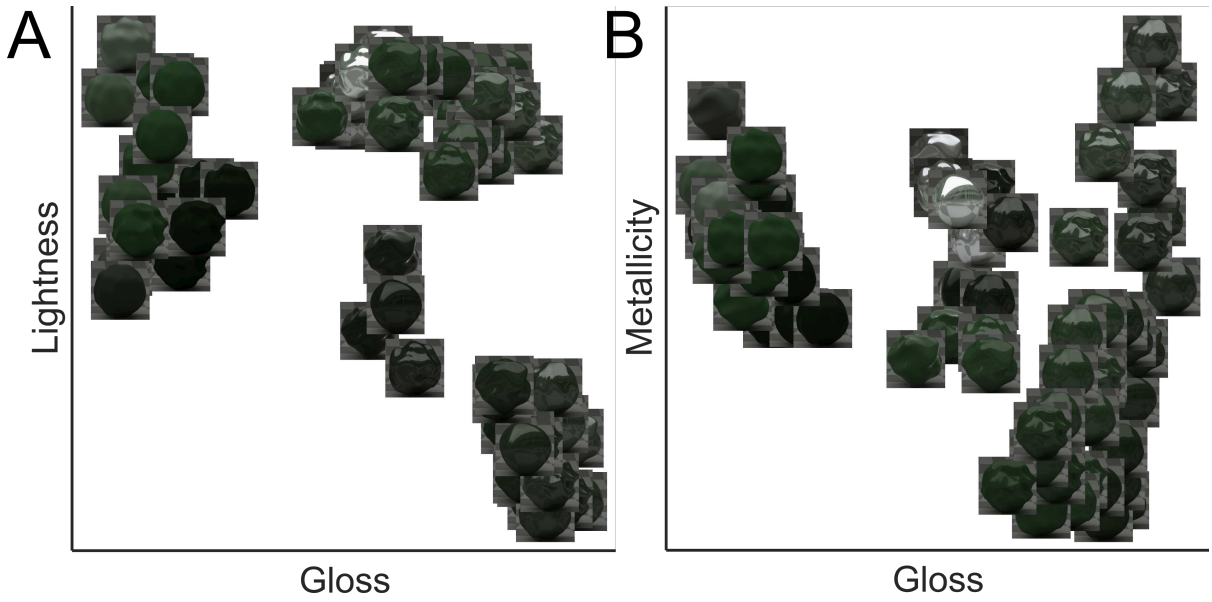

Fig. 8. Arrangement of the stimuli according to the three main dimensions of the embedding, after procrustes superimposition. Stimuli are arranged in the Lightness—Gloss plane (A) and in the Metallicity—Gloss plane (B). Units are arbitrary but comparable between dimensions.

for the first component. In fact, this third component seems to present high values for metallic, (non—natural and light) materials, which are not rated as glossy or specular. Thus, we labelled it as **Metallicity**.

We now have two three—dimensional representations for the 64 images. One is an MDS embedding based on visual similarity judgments, and the other one is based on a PCA of property judgments. We can evaluate the similarity of the two representations by computing the correlations of the respective factor loadings for the images. This was done after a *Procrustes superimposition* to align the dimensions of the MDS embedding to the ones obtained with PCA. Pearson's correlation coefficients were rather high for all the three dimensions (r= 0.871, 0.724 and 0.5659; for the first, the second and the third dimensions of the embedding with **Gloss**, **Lightness** and **Metallicity**, respectively). Correlation coefficients were significantly higher than chance, as they were not contained within the bootstrapped confidence intervals obtained under the null hypothesis that materials occupy random positions along the dimensions of the embedding and the PCA dimensions. The 95% confidence intervals were [0.05 0.33], [0.037 0.34], [0.027 0.33], respectively. This result indicates that, after aligning the embedding's dimensions with the first three principal components, materials are arranged in the embedding and in the principal components in a similar way, as far as the first three dimensions are concerned. Therefore, we could use the first three principal components to interpret the three chosen dimensions of the aligned embedding. Figure 8 shows how the different materials are arranged according to the three dimensions of the embedding after the *Procrustes superimposition.*

The first three dimensions of the transformed embedding are similar to the ones of the original embedding. In fact, Dimension 1, now labelled as **Gloss** because of its correlation with the first principal component, seems to arrange the stimuli from matte to shiny. Dimension 2 labelled as **Lightness**, seems to represent perceived albedo, with low values for dark surfaces. Dimension 3 labelled as **Metallicity** seems related to specularity, with high values for broad specular highlights, which give a "metallic" look to surfaces.

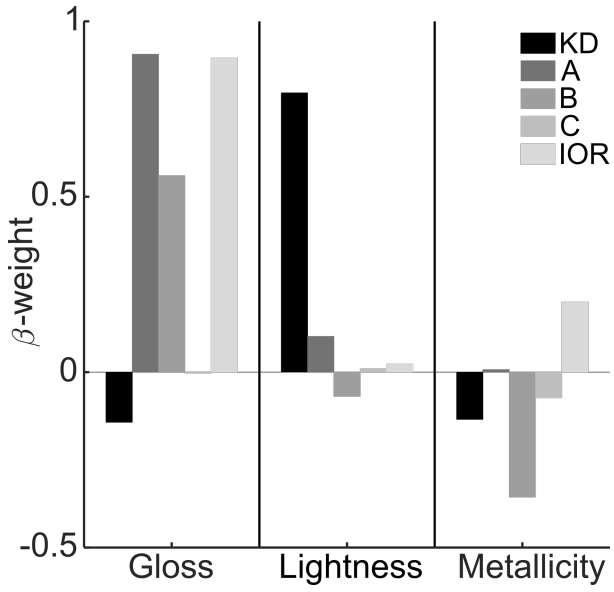

Fig. 9. Regression results of the scores of the three embedding dimensions as a function of the ABC-model parameters. Beta-weights for each of the model parameters (y—axis), for the three dimensions of the MDS embedding (x—axis). Different colors (from dark to light) indicate the five different parameters, as illustrated in the legend.

In the next sections, we attempt to better understand these three dimensions by relating them to the parameters of the ABC—model used to render the materials, and to the statistics of the luminance histograms of the simulated light reflected from the surfaces.

6.1 Relationship with the ABC—model's parameters

In order to understand the physical meaning of the three dimensions of the embedding (**Gloss**,**Lightness** and **Metallicity**), we related them to the parameters of the ABC-model, which control the simulated physics of the rendering system. To do so, we linearly regressed the scores of our simulated materials as a function of each of the 5 parameters of the ABC-model, for each of the three dimensions separately. The values of the parameters were z-transformed, so that the regression coefficients (Îš-weights) are comparable with each other, indicating the contribution of the model parameters on each of the three dimensions. β—weights are presented in Figure 9.

Results indicate that the **Gloss** dimension increases with the amplitude of the specular lobe (**A**), the spread of the specular lobe (**B**) and the index of refraction (**IOR**). According to the formulation of the ABC—model as we used it to create our stimuli, large values of **B** imply sharp specular highlights (see equation 1). Thus, the **Gloss** dimension seems to be selective for bright (high amplitude) sharp specular highlights, and high reflections also at non-grazing angles, as indicated by the relationship with the index of refraction. The **Lightness** component is nearly exclusively related to the diffuse term **(KD)**, suggesting that observers could isolate the diffuse reflections and based their judgments on them independently of the specular components. Interestingly, the **Metallicity** component is negatively related to **B** and **KD**, and positively related to **IOR**. This indicates that object with high scores on **Metallicity** have low diffuse reflections and broad specular highlights, as opposed to objects with high **Gloss**, which are characterized by sharp specular reflections. High index of refraction and low diffuse reflections seem a general feature of the two specular components **Gloss** and **Metallicity**.

This analysis shows that the three dimensions individuated by the MDS and the rating experiment, relate to different parameters of the ABC—model in a relatively simple way.

6.2 Relationship with image statistics

We aimed to relate the dimensions we individuated to previous research about the relationship between the luminance histogram statistics and perception of surface properties. To do so, we extracted the luminance histograms of our rendered stimuli, by rendering a black version of each surface in its position to generate a mask for segmenting it from its background. RGB values were converted to the corresponding luminance values (cd/m^2) as the images were presented on the experimental screen. After that, we computed the first four moments of the luminance histograms: mean (**Mean**), standard deviation (**Std**), skewness (**Skew**), and Kurtosis (**Kurt**) following previous the work on image statistics and perception of surface properties . We also included a measure of image contrast (**Cont**) following Sharan and colleagues [87], defined as the standard deviation divided by the mean of the luminance histogram. Given that the top values of the luminance histogram were found to be particularly diagnostic for surfaces' albedo [87, 98–100], we included in our analyses the average of the luminance values comprised within the 100th percentile (**Top**). These image statistics were related to the **Gloss**, **Albedo** and **Metallicity** dimensions by means of linear classification. Specifically, for each dimension we divided our stimuli in low and high scores, by median split, using these two classes as labels for the linear classification. The naïve Bayesian classifier was trained to distinguish between low and high scores based on the image statistics. In order to limit over-fitting, we iteratively left out each single stimulus and trained the classifier on the others. Performance was evaluated on the classification of the left-out stimulus at each iteration and then averaged across iterations. Thus, at each iteration classification could be right or wrong, depending on the position of the left out stimulus respect to the classification border determined in the training phase. The proportion of correct classifications across iterations gives the performance of the classifier given a set of dimensions (i.e. a combination of image statistics).

Classification was done based on all the 63 combinations (of 1 to 6 elements) of the six statistics. In order to evaluate the individual contributions of these image statistics despite their inter-correlations, for each statistic we compared the classification performance for all the combinations including and all the ones excluding that statistic, following our previous work [67, 104]. Namely, for each statistic, we averaged the classification performance across all the combinations including that statistic, and subtracted from this the performance averaged across all the other combinations. Classification results are shown in Figure 10.

The **Gloss** dimensions is dominated by image contrast, with a lower contribution of skewness. Conversely, the role of contrast in the **Metallicity** dimension is minimal, and classification along this dimension is dominated by skewness. Interestingly, image contrast and skewness have been extensively reported as perceptual correlates of specular reflection. From our results, it seems that they relate to different perceptual dimensions of specular reflection. The **Lightness** dimension is dominated by standard deviation and the mean, presumably because the central portions of the luminance histogram are the most informative about the albedo of glossy surfaces [99].

7 DISCUSSION

We used sparse multidimensional scaling to individuate a metrical embedding to represent distances between the stimuli we generated within the gamut of the physically based BRDF ABC-model, consistent with the similarity judgments produced by our observers. Results suggest that perceived similarity was based on more than two dimensions. In fact, the first three dimensions of the embedding explain 85% of its total variance, as opposed to previous MDS results [108], which could explain over 95% of the variance of the fitted embedding with two dimensions, roughly corresponding to lightness and gloss. Presumably, this due to the larger gamut of materials we used, as provided by the higher complexity and physical accuracy of the BRDF ABC-model. In order to

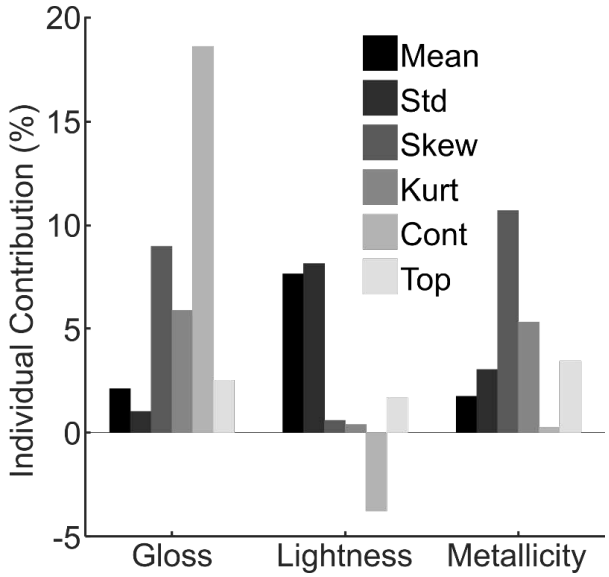

Fig. 10. Classification results based on image statistics. Individual contribution of each image statistic (y-axis), for the three dimensions of the MDS embedding (x-axis). Different colors (from dark to light) indicate the six different image statistics, as illustrated in the legend.

interpret the MDS results, we ran a rating experiment and applied PCA on the ratings for 30 adjectives. The first three PCA dimensions explained about 93% of the variance in the ratings.

Our approach was to start with as many adjectives as participants could think of for describing our stimuli, then determine the number of underlying uncorrelated dimensions to explain the ratings. Therefore, we could confidently assume that the first three PCA dimensions are capturing the qualities humans use to verbally describe visual material properties. When additional properties are considered, like tactile or subjective (e.g. quality or attractiveness) are rated, presumably humans base their ratings on different underlying dimensions and the first three PCA components can only explain 56% of the variability [26].

According to the PCA loadings, we interpreted the three first PCA dimensions as related to **Gloss**, **Lightness** and **Metallicity**. These adjectives used for the labels of the first three PCA dimensions do not have to be isolated by each dimension. In fact, the first dimension, which we labelled **Gloss**, is also related to the *metallic* and *bright* attributes. This is consistent with the results of Serrano and colleagues [86], showing significant correlations between *Glossy*, *Metallic-like* and *Bright* ratings. Similarly, the third dimension (**Metallicity**) is also associated to the *light* and *bright* attributes. However, it is not associated to *Glossy*, and -as expected from the PCA analyses, it is characterized by a pattern of weights that makes it different from the other dimensions. Our results do not imply that the **Metallicity** label reflects its meaning, since the full combination of attributes needs to be considered (i.e. material with high **Metallicity** score do not just have to be very much metal-like, but also e.g., *bright*, *light* and *non-natural* looking). The mapping of reflection properties into an intuitive semantic space is beyond the scope of our study and already provided by [86]. Such a space has been used to determine which aspect of BRDFs need to be preserved and which can be approximated to fit rendered images onto display gamuts (i.e. gamut mapping) [91]. Noteworthy, *metallic-like* and *bright* attributes were the attributes most important for preserving material appearance.

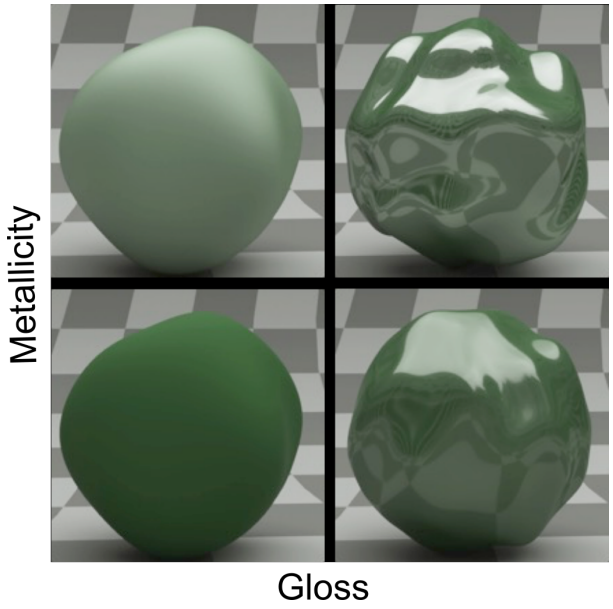

Fig. 11. Example of Gloss and Metallicity differences. The materials are chosen from the arrangement presented in Figure 8, so that the two examples on each row exhibit similar scores on the Metallicity dimension, and the ones on each column on the Gloss dimension. The stimuli on the left column appear matte, hover they clearly differ along the metallic, suggesting that the variability expressed by these examples cannot be arranged on a single dimension.

Since the rotation of the axes of the MDS embedding dimensions is arbitrary, we could align it with the first three PCA dimensions, to interpret the embedding dimensions accordingly. Thus, we arranged our stimuli in a three-dimensional space defined by a diffuse reflection dimension (**Lightness**) and two specular reflection dimensions (**Gloss** and **Metallicity**). Figure 11 shows four examples of variations along these two specular dimensions.

These dimensions seem related to the physical properties of the rendered images (as controlled by the ABC-model's parameters) in a rather simple way. The **Gloss** dimension seems selective for high contrast (*i.e.* high scaling factor of the specular lobe) narrow/sharp specular highlights, and high reflections at grazing angles. Conversely, the **Metallicity** dimension scores high for surfaces defined by broad specular highlights. Several investigators have proposed that the size and sharpness of the specular highlights influence gloss perception [9, 10, 15, 28, 61, 86]. Here we show that this influence applies to different perceptual aspects of specular reflection in an opposite way, with high contrast narrow specular highlight being a cue for one perceptual aspect of specularity and broad specular highlight for another. However, it was reported that human ratings of metallic-like appearance relate to the shape of the specular highlights only for middle amplitude range. For high values of the amplitude, the shape becomes increasingly irrelevant [86]. This could not have been shown in our results, since the regression analysis we performed did not model interaction between parameters.

Pellacini and colleagues [81] proposed and formally described two dimensions: contrast gloss and distinctiveness of image (*DOI gloss*). This allows a direct comparison with our findings, but differences in the two paradigms have to be kept in mind. In the Pellacini's et al. study [81], observers were asked to judge similarity in terms of gloss, whereas we asked our observers to rank the comparison objects in terms of how similar their material is to the reference. In order to place our stimuli in the BRDF space in which *contrast gloss* and *DOI gloss* are defined, we used a BRDF parameter remapping technique [38, 39] to determine the optimal set of Ward-model's

parameters for each of the materials used in our study. This technique makes use of a genetic algorithm to explore a given BRDF parameter space (i.e. ABC-model space) and find the set of parameters in another BRDF space (Ward-model space) yielding the most similar result, in terms of perceived image difference. Thus, we could use the equations proposed in [81] to compute the position of our stimuli along the *contrast gloss* a *distinctiveness of image* dimensions. We observed a close relationship between *DOI gloss* and our **Gloss** dimension (Pearson's r=0.79). **Gloss** is indeed characterized by relatively high amplitude (**A** in the ABC-model) and low width (i.e. high **B** values in the ABC-model) of the specular lobe, which together make the reflected image of the environment sharp and visible. Consistently, *DOI gloss* is negatively related to **Metallicity** (r= -0.29), as this dimension is characterized by broad specular lobes (i.e. low **B** values). *Contrast gloss* is positively related to both **Gloss** and **Metallicity** (r= 0.4, r=0.32, respectively). This is not surprising, given that *contrast gloss* is expressing the contrast between the diffuse and the specular components (see equation 4 in [81]), with high values for low diffuse and high specular contributions. This also explains why the correlation between *contrast gloss* and **Lightness** is negative, although relatively small (r= -0.12). Finally, also *DOI gloss* exhibits a negative correlation with **Lightness**. Although this relationship is relatively small (r= -0.14), as well as the correlation between *contrast gloss* and **Lightness**, it might reflect a perceptual interaction between perceived specularity and lightness. It was indeed reported that black glossy surfaces appear shinier than white glossy surfaces having identical specular reflections [11].

The notion that perception of specular reflection does not rely on a single physical measure is not novel [15], since the early work of Hunter, where perceptual gloss was parsed into six different aspects [47]. Each of the dimensions we proposed can capture one or more of these aspects. For instance, *Distinctness-of-reflected-image gloss* seems specific to the **Gloss** dimension and negatively associated to the **Metallicity** dimension, as the first one is characterized by low and the second one by high width of the specular lobe. *Specular gloss* (i.e. the perceived shininess or brilliance of highlights) is probably expressed by both **Gloss** and **Metallicity**, but absent in **Lightness**, as this dimension isolates the diffuse component. Similarly, *Sheen* (perceived gloss at grazing angles) is probably captured by both **Gloss** and **Metallicity**, since both are associated with relatively high values of the **IOR** coefficient, which controls the expression of specular reflections at grazing angles. The dimensions we propose might fail to capture *Contrast gloss* (i.e. contrasts between specularities and the rest of a surface), as the diffuse component is isolated and seems not to influence the specular dimensions. The other aspects (i.e. *Haze* - presence of a hazy appearance adjacent to the highlights, and *Absence-of-surface-texture gloss* - perceived smoothness of a surface) are difficult to relate to our stimuli and results.

The idea that different perceptual proximal attributes contribute to the percept of specular reflection has been recently proven by Marlow *et al.* [61], who showed that a weighted combination of a series of individual cues, such as perceived contrast, depth, coverage and sharpness of the specular highlights, could account for 94% of the variance in the glossiness ratings obtained from different participants. However, their approach suggests that different proximal aspects of specular reflection, as they are perceived, are integrated into a single dimension of perceived specularity. This is consistent with the low dimensionality of perception of BRDF as reported in early studies [25, 81, 108]. Specifically, Wills and colleagues [108] proposed a two dimensional space, defined by one dimension representing the continuum between dark and light materials, and a second dimension opposing diffuse and specular reflection. Here we propose that perception of complex BRDF is of higher dimension, with more than one dimension coding for specular reflection. In fact, when observers are asked to judge the similarity of different materials, they base their judgments also on differences between different aspects of specular reflection, as related to the **Gloss** and the **Metallicity** dimensions we individuated. In the study by Marlow *et al.* [61], participants were asked to rate a series of rendered surfaces according to their level of specular reflection (*i.e.* gloss). This task was not meant to reveal different aspects of perceived specular reflection, but to provide a synthetic index of perceived gloss (or specularity) in order to relate it to perceived proximal image properties. Therefore, our results are not at odds with the ones from Marlow and colleagues, but complementary.

With multidimensional scaling on paired comparisons, observers did not need to be asked for a synthetic judgment of verbally labelled category (e.g. gloss), and presumably participants used the stimulus properties which are perceptually available, to judge the similarity between pairs of stimuli, in a relatively task-free manner. Since this same method has also been used by Wills and colleagues [108], we believe that the difference in our results is due to the BRDF gamut we used, which is broader than the one defined by the portion of the MERL database used by Wills and colleagues.

Interestingly, the Lightness dimension is nearly exclusively related to diffuse reflection. Conversely, the one or two dimensions proposed in previous investigations [11, 81], did not incorporate the diffuse component in isolation. This might be due to the smaller range of materials that was used, to the different task (gloss vs material similarity judgments) or reflect a true perceptual interaction between perceived specularity and lightness. The latter case would constitute an example of failure of gloss constancy, as gloss could not be judged independent from lightness [11, 77]. Our results instead suggest that observers could tell apart specular and diffuse reflections, consistently with previous research [8, 99].

The visual system might be able to identify specular reflections, because of its tight relationship with surfaces' geometry. In fact, specular highlights tend to appear next to the luminance maxima in diffuse shading [31, 54] and present the same orientation as the diffuse shading that surrounds them [4, 9, 95]. Crucially, when these constraints are violated, specular reflections tend to appear as matte [9, 50, 61], indicating that these constraints might be used by the visual system to identify specular reflections. Despite the correlation between perceived gloss and skewness [69] or contrast [104], two surfaces characterized by identical skewness could appear as glossy or matte, depending on the position of the specular highlights. When highlights are placed where specular reflection would physically occur according to the surface's geometry, they appear as specular reflections. If the same highlights appear elsewhere, the highlights appear as painted light stripes on that same surface, indicating that the visual system uses physical cues for gloss perception other than or additional to the luminance histogram skewness [4, 49]. In general, the perceived gloss of a surface depends on its 3D shape and illumination field. The "bumpiness" of the 3D texture of a surface influences its perceived gloss [44]. Discrimination performance between different materials depend on 3D shape [101], and the ability of human participants to match the reflection parameters of a rendered shape to another one depends on the similarity between the two shapes [74]. Similarly, two surfaces with different 3D shapes need different reflectance properties to appear to be equally glossy. The same is true when the two surfaces are illuminated by different light-fields [21, 78, 79]. The influence of 3D shape and illumination on the perceived material properties of a surface is such that a completely matte surface could appear glossy, by appropriately choosing the relief height and the illumination angle [107]. The influence of 3D shape and illumination on gloss perception is mediated by the way that shape geometry and structure of the illumination field modulate the perceived size, contrast, sharpness, and depth of specular reflections [61]. In our study the association between material and shape was randomized thus the effect of shape on material perception could not have biased our result. However, our methods did not allow to investigate such effect, as the shape was not systematically varied.

It was reported that the brightest regions of surface are used by the visual system to perceive its lightness (*i.e.* perceived albedo), since they are particularly related to the surface's albedo for matte objects [97, 98, 100]. However, for our stimuli the Lightness dimension was highly related to the mean of the luminance histogram rather than to its top values. This is not surprising, given the finding that the brightest regions of non-matte surfaces are likely to include specular reflections. Thus the central portions of the luminance histogram are the most informative about the albedo of glossy surfaces, and human observers take this in account when producing lightness judgements [99]. It was proposed that certain statistical regularities of the visual input correlate with certain surface properties, being potential useful heuristic for material appearances. Specifically, Motoyoshi and colleagues [69], based on gloss ratings of a limited number of textured surfaces, suggested that the skewness of the luminance histogram is the preferential proximal image cue for gloss perception. Using a broader dataset of

images of natural surfaces, classified as glossy or matte by independent observers, Wiebel and colleagues [104] showed that the standard deviation was predicting human gloss judgments better than skewness, challenging previous results. For the rendered images used here, skewness seems to be a good predictor for gloss or metallicity, which agrees with the results of Wiebel for this class of stimuli.

The fact that certain properties of the luminance histogram are related to different perceptual qualities, does not mean that these properties are necessary and sufficient to elicit these qualities. Identical luminance gratings can look either metallic or matte depending on their perceived three-dimensional shape [62]. Rotating the specular reflections of uniform albedo surfacesdo not change image statistics, but it does affect perceived gloss. Such surfaces do not appear glossy, unless the specular highlights [49] or low-lights [51] are positioned and oriented consistently with the surface's diffuse shading profile. In general, perceived surface reflectance properties depend on several different aspects of a surface, such as its shape, texture and motion [5, 49, 59, 62] that cannot be captured by the luminance histogram.

Here we propose that perception of specular reflection is likely to be more than one-dimensional, but our coarse sampling of the ABC-model materials gamut does not allow us to provide a precise characterization of the perceptual space underlining it. The labels we used for our dimensions are a direct interpretation of the observers' ratings in Experiment 2. Therefore their meaning does not have to correspond to physical reflection properties. Todd and Norman recently investigated metallic appearance of different materials under different illuminations [94]. Observers rated metallic appearance of curved chrome objects simulated with varying roughness, i.e. the degree of microscopic scattering on a surface, from one direction to uniformly scattering light in all directions. They found a complex non-monotonic relationship between roughness and the appearance ratings, which could not have been captured by our linear classification analysis. Furthermore, this relationship changed with the spatial structure of the illumination. Working out this complexity needs specific investigation and is beyond the scope of our study. The ABC-model gamut includes a large variety of physically possible materials, irrespective of their occurrence in the world.

The characterization of the distribution of BRDFs in reality, could help driving further investigation on material perception. Nevertheless, the relationship between the dimensions we individuated and the parameters of the ABC-model gives the basis for a fine investigation of this perceptual space. In analogy with MacAdam's work on color [58], future research could attempt to measure the metric properties of a three-dimensional BRDF perceptual space by measuring discrimination threshold around a set of points systematically spanning that space. Also, the relationship we individuated with different image statistics, might allow to extend the previous work assessing their causal role on appearance [104] beyond a single specular dimension.

8 ACKNOWLEDGMENTS

We thank Roland Fleming for helpful discussions and comments. We also thank Anna Metzger for her support in translating German into English for the rating experiment. The study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 222641018 - SFB/TRR 135, TP A8 and C2. This work was also supported by the projects MUVApp N-250293 and Spectraskin N-288670, both funded by the Research Council of Norway.

REFERENCES

[1] Edward H Adelson and Alex P Pentland. 1996. The perception of shading and reflectance. *Perception as Bayesian inference* (1996), 409–423.

[2] Sameer Agarwal, Josh Wills, Lawrence Cayton, Gert Lanckriet, David Kriegman, and Serge Belongie. 2007. Generalized non-metric multidimensional scaling. In *Artificial Intelligence and Statistics*. 11–18.

[3] Barton L Anderson. 2011. Visual perception of materials and surfaces. *Current Biology* 21, 24 (2011), R978–R983.

[4] Barton L Anderson and Juno Kim. 2009. Image statistics do not explain the perception of gloss and lightness. *Journal of vision* 9, 11 (2009), 1–17.

[5] Kim Juno Anderson, Barton L. 2009. Image statistics do not explain the perception of gloss and lightness. *Journal of Vision* 9, 10 (2009), 1–17.

[6] Mahdi M. Bagher, John Snyder, and Derek Nowrouzezahrai. 2016. A Non-Parametric Factor Microfacet Model for Isotropic BRDFs. *ACM Trans. Graph.* 35, 5, Article 159 (July 2016), 16 pages.

[7] Elisabeth Baumgartner, Christiane B Wiebel, and Karl R Gegenfurtner. 2013. Visual and haptic representations of material properties. *Multisensory research* 26, 5 (2013), 429–455.

[8] Jacob Beck. 1964. The effect of gloss on perceived lightness. *The American journal of psychology* 77, 1 (1964), 54–63.

[9] Jacob Beck and Slava Prazdny. 1981. Highlights and the perception of glossiness. *Attention, Perception, & Psychophysics* 30, 4 (1981), 407–410.

[10] Julia Berzhanskaya, Gurumurthy Swaminathan, Jacob Beck, and Ennio Mingolla. 2005. Remote effects of highlights on gloss perception. *Perception* 34, 5 (2005), 565–575.

[11] Fred W Billmeyer Jr and Francis XD O'Donnell. 1987. Visual gloss scaling and multidimensional scaling analysis of painted specimens. *Color Research & Application* 12, 6 (1987), 315–326.

[12] Andrew Blake and Heinrich. BÃijlthoff. 1990. Does the brain know the physics of specular reflection. *Nature:343(6254).165* (1990).

[13] Adrien Bousseau, James P O'shea, Frédo Durand, Ravi Ramamoorthi, and Maneesh Agrawala. 2013. Gloss perception in painterly and cartoon rendering. *ACM Transactions on Graphics (TOG)* 32, 2 (2013), 18.

[14] Brent Burley. 2012. Physically-based shading at Disney. In *ACM SIGGRAPH 2012 Courses (SIGGRAPH '12)*. ACM, New York, NY, USA, Article 10, 7 pages.

[15] Alice C Chadwick and Robert W Kentridge. 2015. The perception of gloss: a review. *Vision research* 109 (2015), 221–235.

[16] Steven Cholewiak, Romain Vergne, Benjamin Kunsberg, Steven Zucker, and Roland W Fleming. 2015. Distinguishing between texture and shading flows for 3D shape estimation. *Journal of vision* 15, 12 (2015), 965–965.

[17] Steven A Cholewiak and Roland W Fleming. 2013. Towards a unified explanation of shape from shading and texture. *Journal of Vision* 13, 9 (2013), 258–258.

[18] Steven A Cholewiak, Benjamin Kunsberg, Steven Zucker, and Roland W Fleming. 2014. Predicting 3D shape perception from shading and texture flows. *Journal of Vision* 14, 10 (2014), 1113–1113.

[19] Eugene L Church, Peter Z Takacs, and Thomas A Leonard. 1990. The prediction of BRDFs from surface profile measurements. In *Scatter from Optical Components*, Vol. 1165. International Society for Optics and Photonics, 136–151.

[20] Robert L Cook and Kenneth E Torrance. 1982. A Reflectance Model for Computer Graphics. *ACM Trans. Graph.* 1, 1 (Jan. 1982), 7–24. https://doi.org/10.1145/357290.357293

[21] Katja Doerschner, Huseyin Boyaci, and Laurence T Maloney. 2010. Estimating the glossiness transfer function induced by illumination change and testing its transitivity. *Journal of Vision* 10, 4 (2010), 1–9.

[22] Katja Doerschner, Roland W Fleming, Ozgur Yilmaz, Paul R Schrater, Bruce Hartung, and Daniel Kersten. 2011. Visual motion and the perception of surface material. *Current Biology* 21, 23 (2011), 2010–2016.

[23] Jonathan Dupuy and Wenzel Jakob. 2018. An Adaptive Parameterization for Efficient Material Acquisition and Rendering. *ACM Trans. Graph.* 37, 6, Article 274 (Dec. 2018), 14 pages.

[24] Jonathan Dupuy and Wenzel Jakob. 2018. An Adaptive Parameterization for Efficient Material Acquisition and Rendering. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 37, 6 (Nov. 2018), 274:1–274:18. https://doi.org/10.1145/3272127.3275059

[25] James A Ferwerda, Fabio Pellacini, and Donald P Greenberg. 2001. Psychophysically based model of surface gloss perception. In *Human Vision and Electronic Imaging VI*, Vol. 4299. International Society for Optics and Photonics, 291–302.

[26] Jiri Filip and Martina Kolafova. 2019. Perceptual Attributes Analysis of Real-world Materials. *ACM Transactions on Applied Perception* 16, 1 (2019), 1–21.

[27] JirÃŋ Filip and RadomÃŋr Vávra. 2014. Template-Based Sampling of Anisotropic BRDFs. *Computer Graphics Forum* 33, 7 (October 2014), 91–99. https://doi.org/10.1111/cgf.12477

[28] Roland W Fleming. 2012. Human perception: Visual heuristics in the perception of glossiness. *Current Biology* 22, 20 (2012), R865–R866.

[29] Roland W Fleming. 2014. Visual perception of materials and their properties. *Vision research* 94 (2014), 62–75.

[30] Roland W Fleming, Ron O Dror, and Edward H Adelson. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of vision* 3, 5 (2003), 347–368.

[31] Roland W Fleming, Antonio Torralba, and Edward H Adelson. 2004. Specular reflections and the perception of shape. *Journal of vision* 4, 9 (2004), 798–820.

[32] Roland W Fleming, Christiane B Wiebel, and Karl R Gegenfurtner. 2013. Perceptual qualities and material classes. *Journal of vision* 13, 8 (2013), 1–20.

[33] USC Institute for Creative Technologies. 2017. High-Resolution Light Probe Image Gallery. http://gl.ict.usc.edu/Data/HighResProbes/

[34] Augustin Jean Fresnel. 1834. *Mémoire sur la loi des modifications que la réflexion imprime à la lumière polarisée.* De l'Imprimerie de Firmin Didot Fréres.

[35] Karl R Gegenfurtner and Daniel C Kiper. 2003. Color vision. *Annual review of neuroscience* 26, 1 (2003), 181–206.

[36] Karl R Gegenfurtner and Jochem Rieger. 2000. Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology* 10, 13 (2000), 805–808.
[37] Dar'ya Guarnera, Giuseppe Claudio Guarnera, Abhijeet Ghosh, Cornelia Denk, and Mashhuda Glencross. 2016. Brdf representation and acquisition. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 625–650.
[38] Dar'ya Guarnera, Giuseppe Claudio Guarnera, Matteo Toscani, Mashhuda Glencross, Baihua Li, Jon Yngve Hardeberg, and Karl R Gegenfurtner. 2018. Perceptually validated analytical BRDFs parameters remapping. In *ACM SIGGRAPH 2018 Talks*. ACM, 17.
[39] Dar'ya Guarnera, Giuseppe Claudio Guarnera, Matteo Toscani, Mashhuda Glencross, Baihua Li, Jon Yngve Hardeberg, and Karl R. Gegenfurtner. 2018. Perceptually Validated Cross-Renderer Analytical BRDF Parameter Remapping. *IEEE Transactions on Visualization and Computer Graphics* (2018), 1–15. https://doi.org/10.1109/TVCG.2018.2886877
[40] Giuseppe Claudio Guarnera, Peter Hall, Alain Chesnais, and Mashhuda Glencross. 2017. Woven fabric model creation from a single image. *ACM Transactions on Graphics (TOG)* 36, 5 (2017), 165.
[41] Selig Hecht. 1930. The development of Thomas Young's theory of color vision. *JOSA* 20, 5 (1930), 231–270.
[42] Hermann von Helmholtz. 1852. Über die Theorie der zusammengesetzten Farben. *Annalen der Physik* 163, 9 (1852), 45–66.
[43] Hermann von Helmholtz. 1962. *Helmholtz's treatise on physiological optics*. Vol. 1. Dover Publications.
[44] Yun-Xian Ho, Michael S Landy, and Laurence T Maloney. 2008. Conjoint measurement of gloss and surface texture. *Psychological Science* 19, 2 (2008), 196–204.
[45] Nicolas. Holzschuch and Romain. Pacanowski. 2017. The Effects of Digital Cameras Optics and Electronics for Material Acquisition. In *Proceedings of the Workshop on Material Appearance Modeling (MAM '17)*. Eurographics Association, Goslar Germany, Germany, 25–28. https://doi.org/10.2312/mam.20171328
[46] Nicolas Holzschuch and Romain Pacanowski. 2017. A Two-scale Microfacet Reflectance Model Combining Reflection and Diffraction. *ACM Trans. Graph.* 36, 4, Article 66 (July 2017), 12 pages.
[47] Richard S Hunter. 1937. Methods of determining gloss. *NBS Research paper RP* 958 (1937).
[48] David G Kendall. 1989. A survey of the statistical theory of shape. *Statist. Sci.* (1989), 87–99.
[49] Juno Kim and Barton L Anderson. 2010. Image statistics and the perception of surface gloss and lightness. *Journal of Vision* 10, 9 (2010), 1–17.
[50] Juno Kim, Phillip J Marlow, and Barton L Anderson. 2011. The perception of gloss depends on highlight congruence with surface shading. *Journal of Vision* 11, 9 (2011), 1–19.
[51] Juno Kim, Phillip J Marlow, and L Anderson, Barton. 2012. The dark side of gloss. *Nature Neuroscience* 15 (2012), 1590–1595.
[52] Mario Kleiner, David Brainard, Denis Pelli, Allen Ingling, Richard Murray, and Christopher Broussard. 2007. What's new in Psychtoolbox-3? Perception. 36 ECVP Abstract Supplement 14.
[53] David C Knill and Daniel Kersten. 1991. Apparent surface curvature affects lightness perception. *Nature* 351, 6323 (1991), 228.
[54] Jan J Koenderink and Andrea J van Doorn. 1980. Photometric invariants related to solid shape. *Optica Acta: International Journal of Optics* 27, 7 (1980), 981–996.
[55] Manuel Lagunas, Sandra Malpica, Ana Serrano, Elena Garces, Diego Gutierrez, and Belen. Masia. 2019. A Similarity Measure for Material Appearance. *arXiv preprint:1905.01562* (2019).
[56] Johann Heinrich Lambert. 1760. *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Klett.
[57] Joakim Löw, Joel Kronander, Anders Ynnerman, and Jonas Unger. 2012. BRDF models for accurate and efficient rendering of glossy surfaces. *ACM Transactions on Graphics (TOG)* 31, 1 (2012), 9.
[58] David L MacAdam. 1942. Visual sensitivities to color differences in daylight. *Josa* 32, 5 (1942), 247–274.
[59] Anderson Barton L Marlow, Phillip J. 2016. Motion and texture shape cues modulate perceived material properties. *Journal of Vision* 16, 1 (2016), 1–14.
[60] Phillip J Marlow and Barton L Anderson. 2013. Generative constraints on image cues for perceived gloss. *Journal of vision* 13, 14 (2013), 1–23.
[61] Phillip J Marlow, Juno Kim, and Barton L Anderson. 2012. The perception and misperception of specular surface reflectance. *Current Biology* 22, 20 (2012), 1909–1913.
[62] Phillip J Marlow, Dejan TodoroviÄĞ, and Barton L Anderson. 2015. Coupled computations of three-dimensional shape and material. *Current Biology* 25, 6 (2015), R221–R222.
[63] David Marr. 1982. *Vision: A computational investigation into*. WH Freeman.
[64] Wojciech Matusik. 2003. *A data-driven reflectance model*. Ph.D. Dissertation. Massachusetts Institute of Technology.
[65] James Clerk Maxwell. 1857. XVIII.-Experiments on Colour, as perceived by the Eye, with Remarks on Colour-Blindness. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* 21, 2 (1857), 275–298.
[66] Stephen McAuley, Stephen Hill, Naty Hoffman, Yoshiharu Gotanda, Brian Smits, Brent Burley, and Adam Martinez. 2012. Practical Physically-based Shading in Film and Game Production. In *ACM SIGGRAPH 2012 Courses (SIGGRAPH '12)*. ACM, New York, NY, USA, Article 10, 7 pages. https://doi.org/10.1145/2343483.2343493

[67] Zarko Milojevic, Robert J Ennis, Matteo Toscani, and Karl R Gegenfurtner. 2018. Categorizing natural color distributions. *Vision research* (2018).

[68] John D Mollon. 1989. "Tho'she kneel'd in that place where they grew..." The uses and origins of primate colour vision. *Journal of Experimental Biology* 146, 1 (1989), 21–38.

[69] Isamu Motoyoshi, Shin'ya Nishida, Lavanya Sharan, and Edward H Adelson. 2007. Image statistics and the perception of surface qualities. *Nature* 447, 7141 (2007), 206.

[70] Alexander A Muryy, Roland W Fleming, and Andrew E Welchman. 2016. 'Proto-rivalry': how the binocular brain identifies gloss. *Proc. R. Soc. B* 283, 1830 (2016), 20160383.

[71] Alexander A Muryy, Andrew E Welchman, Andrew Blake, and Roland W Fleming. 2013. Specular reflections and the estimation of shape from binocular disparity. *Proceedings of the National Academy of Sciences* 110, 6 (2013), 2413–2418.

[72] Addy Ngan, Frédo Durand, and Wojciech Matusik. 2005. Experimental Analysis of BRDF Models. In *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques (EGSR '05)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 117–126. https://doi.org/10.2312/EGWR/EGSR05/117-126

[73] Fred E Nicodemus, Joseph C Richmond, Jack J Hsia, IW Ginsberg, and T Limperis. 1992. Radiometry. Jones and Bartlett Publishers, Inc., USA, Chapter Geometrical Considerations and Nomenclature for Reflectance, 94–145. http://dl.acm.org/citation.cfm?id=136913.136929

[74] ShinâĂŹya Nishida and Shinya Mikio. 1998. Use of image-based information in judgments of surface-reflectance properties. *JOSA A* 15, 12 (1998), 2951–2965.

[75] Akiko Nishio, Takeaki Shimokawa, Naokazu Goda, and Hidehiko Komatsu. 2014. Perceptual gloss parameters are encoded by population responses in the monkey inferior temporal cortex. *Journal of Neuroscience* 34, 33 (2014), 11143–11151.

[76] Farley J Norman, James T Todd, and Guy A Orban. 2004. Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science* 15, 8 (2004), 565–570.

[77] Francis X D O'Donnell and Fred W Billmeyer. 1986. Psychometric scaling of gloss. *Review and Evaluation of Appearance: Methods and Techniques*. 914 (1986), 14–32.

[78] Maria Olkkonen and David H Brainard. 2011. Joint effects of illumination geometry and object shape in the perception of surface reflectance. *i-Perception* 2, 9 (2011), 1014–1034.

[79] Maria Olkkonen and David Brainard H. 2010. Perceived glossiness and lightness under real-world illumination. *Journal of Vision* 10, 9 (2010), 1–19.

[80] Vivian C Paulun, Filipp Schmidt, Jan Jaap R van Assen, and Roland W Fleming. 2017. Shape, motion, and optical cues to stiffness of elastic objects. *Journal of vision* 17, 1 (2017), 1–22.

[81] Fabio Pellacini, James A Ferwerda, and Donald P Greenberg. 2000. Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 55–64.

[82] Jonathan B Phillips, James A Ferwerda, and Ann Nunziata. 2010. Gloss discrimination and eye movements. In *Human Vision and Electronic Imaging XV*, Vol. 7527. International Society for Optics and Photonics, 75270Z.

[83] Zygmunt Pizlo. 2001. Perception viewed as an inverse problem. *Vision research* 41, 24 (2001), 3145–3161.

[84] Tomaso Poggio and Christof Koch. 1985. III-Posed problems early vision: from computational theory to analogue networks. *Proc. R. Soc. Lond. B* 226, 1244 (1985), 303–323.

[85] Tomaso Poggio, Vincent Torre, and Christof Koch. 1987. Computational vision and regularization theory. In *Readings in Computer Vision*. Elsevier, 638–643.

[86] Ana Serrano, Diego Gutierrez, Karol Myszkowski, Hans-Peter Seidel, and Belen Masia. 2016. An intuitive control space for material appearance. *ACM Transactions on Graphics* 35, 6 (2016), 1–12.

[87] Lavanya Sharan, Yuanzhen Li, Isamu Motoyoshi, Shin'ya Nishida, and Edward H Adelson. 2008. Image statistics for surface reflectance perception. *JOSA A* 25, 4 (2008), 846–865.

[88] Lavanya Sharan, Ruth Rosenholtz, and Edward Adelson. 2009. Material perception: What can you see in a brief glance? *Journal of Vision* 9, 8 (2009), 784–784.

[89] Cyril Soler, Kartic Subr, and Derek Nowrouzezahrai. 2018. A Versatile Parameterization for Measured Material Manifolds. *Computer graphics forum* 37, 2 (2018), 135–144.

[90] Petroc Sumner and John D Mollon. 2000. Chromaticity as a signal of ripeness in fruits taken by primates. *Journal of Experimental Biology* 203, 13 (2000), 1987–2000.

[91] Tiancheng Sun, Ana Serrano, Diego Gutierrez, and Belen Masia. 2017. Attribute-preserving gamut mapping of measured BRDFs. *In Computer Graphics Forum* 36, 4 (2017), 47–54.

[92] Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Gabriele Simone. 2017. Image contrast measure as a gloss material descriptor. In *International Workshop on Computational Color Imaging*. Springer, 233–245.

[93] William Thompson, Roland W Fleming, Sarah Creem-Regehr, and Jeanine Kelly Stefanucci. 2016. *Visual perception from a computer graphics perspective*. AK Peters/CRC Press.

[94] James T Todd and Farley J Norman. 2018. The visual perception of metal. *Journal of Vision* 18, 9 (2018), 1–17.
[95] James T Todd, Farley J Norman, and Ennio Mingolla. 2004. Lightness constancy in the presence of specular highlights. *Psychological Science* 15, 1 (2004), 33–39.
[96] Kenneth E Torrance and Ephraim M Sparrow. 1967. Theory for off-specular reflection from roughened surfaces. *Josa* 57, 9 (1967), 1105–1114.
[97] Matteo Toscani and Matteo. Valsecchi. 2019. Lightness Discrimination Depends More on Bright Rather Than Shaded Regions of Three-Dimensional Objects. *i-perception: 10(6), 1âĂŞ10.* (2019).
[98] Matteo Toscani, Matteo Valsecchi, and Karl R Gegenfurtner. 2013. Optimal sampling of visual information for lightness judgments. *Proceedings of the National Academy of Sciences* 110, 27 (2013), 11163–11168.
[99] Matteo Toscani, Matteo Valsecchi, and Karl R Gegenfurtner. 2017. Lightness perception for matte and glossy complex shapes. *Vision research* 131 (2017), 82–95.
[100] Matteo Toscani, Sunčica Zdravković, and Karl R Gegenfurtner. 2016. Lightness perception for surfaces moving through different illumination levels. *Journal of vision* 16, 15 (2016), 1–18.
[101] Peter Vangorp and Philip Dutré. 2008. Shape-dependent gloss correction. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*. ACM, 123–130.
[102] Gregory J Ward. 1992. Measuring and modeling anisotropic reflection. In *ACM SIGGRAPH Computer Graphics*, Vol. 26. ACM, 265–272.
[103] Jakob Wenzel. 2010. Mitsuba renderer. http://www.mitsuba-renderer.org.
[104] Christiane B Wiebel, Matteo Toscani, and Karl R Gegenfurtner. 2015. Statistical correlates of perceived gloss in natural images. *Vision research* 115 (2015), 175–187.
[105] Christiane B Wiebel, Matteo Valsecchi, and Karl R Gegenfurtner. 2013. The speed and accuracy of material recognition in natural images. *Attention, Perception, & Psychophysics* 75, 5 (2013), 954–966.
[106] Christiane B Wiebel, Matteo Valsecchi, and Karl R Gegenfurtner. 2014. Early differential processing of material images: Evidence from ERP classification. *Journal of vision* 14, 7 (2014), 1–13.
[107] Maarten WA Wijntjes and Pont Sylvia C. 2010. Illusory gloss on Lambertian surfaces. *Journal of Vision* 10, 9 (2010), 1–12.
[108] Josh Wills, Sameer Agarwal, David Kriegman, and Serge Belongie. 2009. Toward a perceptual space for gloss. *ACM Transactions on Graphics (TOG)* 28, 4 (2009), 103.
[109] Christoph Witzel and Karl R Gegenfurtner. 2018. Color perception: Objects, constancy, and categories. *Annual review of vision science* 4 (2018), 475–499.
[110] Bei Xiao and David H Brainard. 2008. Surface gloss and color perception of 3D objects. *Visual neuroscience* 25, 3 (2008), 371–385.