The perception of hazy gloss

Peter Vangorp

Department of Computer Science, Edge Hill University, Ormskirk, UK

Pascal Barla

Manao Project, Inria Bordeaux–Sud-Ouest, Talence, France

Roland W. Fleming

Department of Psychology, University of Giessen, Giessen, Germany

Most previous work on gloss perception has examined the strength and sharpness of specular reflections in simple bidirectional reflectance distribution functions (BRDFs) having a *single* specular component. However, BRDFs can be substantially more complex and it is interesting to ask how many additional perceptual dimensions there could be in the visual representation of surface reflectance qualities. To address this, we tested materials with two specular components that elicit an impression of hazy gloss. Stimuli were renderings of irregularly shaped objects under environment illumination, with either a single Ward specular BRDF component (Ward, 1992), or two such components, with the same total specular reflectance but different sharpness parameters, yielding both sharp and blurry highlights simultaneously. Differently shaped objects were presented side by side in matching, discrimination, and rating tasks. Our results show that observers mainly attend to the sharpest reflections in matching tasks, but they can indeed discriminate between single-component and two-component specular materials in discrimination and rating tasks. The results reveal an additional perceptual dimension of gloss—beyond strength and sharpness— akin to "haze gloss" (Hunter & Harold, 1987). However, neither the physical measurements of Hunter and Harold nor the kurtosis of the specular term predict perception in our tasks. We suggest the visual system may use a decomposition of specular reflections in the perception of hazy gloss, and we compare two possible candidates: a physical representation made of two gloss components, and an alternative representation made of a central gloss component and a surrounding halo component.

Introduction

Many natural materials exhibit more complex specular reflections than can be captured by the reflectance models that are typically found in the gloss perception literature. For instance, materials with a rough base layer coated with a clear varnish, such as metallic car paints (Figure 1A), Christmas ornaments, candy apples, or varnished plastics (Figure 1B); or materials with a specular base layer coated with a layer of dirt or grease (Figure 1C) are poorly approximated by traditional bidirectional reflectance distribution function (BRDF) models. Such materials have a distinctive "hazy" appearance, leading Hunter and Harold (1987), in their seminal work on the measurement of appearance, to include a parameter known as *haze gloss* to describe such appearance characteristics. Sometimes a similar appearance is created by a mixture of microfacet distributions at a single layer (Cook & Torrance, 1982), such as partially polished metals (Figure 1D), or by complex diffraction effects (Krywonos, Harvey, & Choi, 2011) due to roughness at scales comparable to visible wavelengths.

In intuitive terms, hazy reflections are those in which a relatively sharp (distinct) reflection is superimposed with, or surrounded by, a blurry "bloom" or fringe, similar to the effect of viewing a light source through haze or mist. For example, in Figure 1A, the sharp edges of the highlight are surrounded by a fuzzy "glow", which gives the material a specific appearance that is different from either the sharp reflections or the blurry highlights on their own. It is this composite impression of "haze" that we seek to understand. The precise nature of the visual representation of hazy gloss—and whether the visual system decomposes the reflections into distinct subcomponents (e.g., separate sharp and blurry terms)—remains poorly understood.

In computer graphics and materials science it has been noted that many reflectance measurements require more complex models to achieve acceptable fitting quality according to various error metrics and also
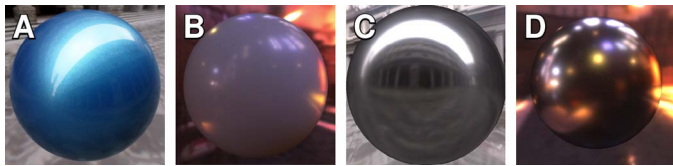
Figure 1. Material classes for which traditional BRDF models were shown to be insufficient: (A) metallic car paint (Rump, Müller, Sarlette, Koch, & Klein, 2008), (B) varnished plastic (Löw et al., 2012), (C) greasy steel (Matusik, Pfister, Brand, & McMillan, 2003), and (D) partially polished nickel (Ngan et al., 2005).

according to visual inspection of rendered images. These shortcomings require, for example, two or more simple gloss components (Cook & Torrance, 1982; Lafortune, Foo, Torrance, & Greenberg, 1997; Ngan, Durand, & Matusik, 2005), a more elaborate single gloss component (Bagher, Soler, & Holzschuch, 2012; Church, Takacs, & Leonard, 1989; Freniere, Gregory, & Chase, 1997; Löw, Kronander, Ynnerman, & Unger, 2012), or a simple gloss component combined with a diffraction component (Holzschuch & Pacanowski, 2017). A key question this raises is to what extent, and under which parameter conditions, does the human visual system distinguish such materials from those with a simple, single gloss component.

To date, the human visual perception of such hazy gloss materials has barely been investigated. Instead, the focus in material perception has so far been on both real and rendered diffuse materials (e.g., te Pas & Pont, 2005), including textured materials (e.g., Kim, Marlow, & Anderson, 2014) and macroscopic surface roughness (e.g., Ho, Landy, & Maloney, 2008; Marlow, Kim, & Anderson, 2012), and on specular and single-component glossy materials (e.g., Fleming, Dror, & Adelson, 2003; Olkkonen & Brainard, 2010). The perceptual representation of general glossy materials was proposed to encode at least six appearance aspects (Hunter & Harold, 1987), although only two independent dimensions were found in perceptual studies on a single class of materials such as real paint samples (O'Donnell & Billmeyer, Jr., 1986) or rendered materials (Pellacini, Ferwerda, & Greenberg, 2000; Wills, Agarwal, Kriegman, & Belongie, 2009).

The following matching, discrimination, and rating experiments are designed to investigate the conditions in which haze gloss is encoded in the perceptual representation of glossy materials. This work explores the space of hazy materials, including both plastics and metals, and focuses on the parameter ranges that exhibit the most pronounced haze by examining different but related sets of stimulus materials in each experiment.

Methods

Stimuli

The stimuli were computer-generated animations of a random blob shape with various materials. The blob shape was constructed from a sphere, displaced outward by a Perlin noise source (Perlin, 2002). Uniform random orientations of the blob shape were used to prevent observers from basing their judgments on specific local features in the image. The objects were continuously rotating back and forth around their vertical axis over a 30° angle and at a maximum angular velocity of 15°/s to exhibit the typical motion of highlights and reflections over the surface (Koenderink & van Doorn, 1980) and to reduce the possibility of the visible side of the shape affecting the perceived material (Vangorp, Laurijssen, & Dutré, 2007). The objects were illuminated by light probes from the (Debevec, 1998) light probe library, specifically the "Grace Cathedral" probe (Experiments 1 and 2) and the "Galileo's Tomb" probe (Experiment 3). The rendering was done in real time for Experiments 1 and 2, hence allowing participants in the matching task to vary parameters smoothly while the object was rotating. The real-time rendering engine used an OpenGL implementation of filtered importance sampling (Křivánek & Colbert, 2008) on an NVIDIA Quadro 4000 GPU (NVIDIA, Santa Clara, CA). The experiments were displayed on a Samsung SyncMaster 27 in. LCD monitor (Samsung, San Jose, CA) at its native 1920 × 1080 resolution and 15–30 frames per second. For Experiment 3, short stimulus video clips were recorded using the same rendering engine and played back during the experiment. Experiment 3 was implemented using the Psychtoolbox-3 extension for MATLAB (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Participants were seated approximately 1 m from the screen. At this viewing distance, each stimulus subtended approximately 5.7° of visual angle. Representative examples of the stimulus presentation for each experiment are shown in Figure 2.

The plastic-like class of materials (Experiment 1) had a constant green diffuse component with RGB reflectance $\rho_d = (0.1, 0.3, 0.1)$. The dark shiny plastic (Experiment 2) and silvery metallic (Experiment 3) materials had no diffuse component (i.e., diffuse reflectance $\rho_d = 0$). Both plastics had a dielectric Fresnel reflectance with a typical refractive index of $\eta = 1.5$, while the metal used a conductor Fresnel reflectance with a complex refractive index of $\eta = 0.145$, $\kappa = 3.19$, which is typical of silver. Each material had either a single or two linearly summed specular components, each of which was modeled with an isotropic Ward BRDF (Ward, 1992) with a variable sharpness pa-

A. Matching task | B. Discrimination task | C. Rating task

Figure 2. Representative screenshots of the stimulus presentation for each task. The framed insets show image zooms on instructions and controls. The objects are rotating back and forth around their vertical axis. In the matching task (A) the target and match stimuli are shown side by side. In the discrimination task (B) the top-left stimulus is made of a two-component material, while the other three use a single-component material. In the rating task (C) the six surface quality sliders are shown at the bottom.

rameter α (see Figure 3). This parameter controls the sharpness of the reflections, *with smaller values corresponding to higher sharpness*. It ranges from 0.2 (blurry) down toward 0 (sharp) in a perceptually linear progression (Pellacini et al., 2000). For the green plastic-like materials, each specular component had a reflectance $\rho_s = 0.15$. For the dark shiny plastic material, the two specular reflectances varied between 0.05 (faint) and 0.25 (bright), while always summing to 0.30 to preserve the total reflected energy. In this paper we will use the term *two-component* for materials with two specular components. We will not count diffuse components in our terminology. Single-component versions of the plastic materials were produced with a single specular component with reflectance $\rho_s = 0.30$. For the metallic material, the two specular reflectances instead summed to 0.95, which is typical of silver. Each specular component reflectance varied between 0.1583 and 0.7917, such that the relative specular reflectances were the same for both dark plastic and silver materials.

We will refer to the sharpness parameters of the narrow (α_n) or wide (α_w) specular components. Alternatively, we will refer to their average sharpness ($\alpha_{avg} = (\alpha_n + \alpha_w)/2$) and their difference from the average sharpness ($\Delta\alpha = |\alpha_n - \alpha_w|/2$). These are the two sets of independent sharpness parameters in the analysis of experiments. Accordingly, we will also refer to the reflectance of the narrow ($\rho_{s,n}$) or wide ($\rho_{s,w}$) specular components.

The complete shading model, depicted in Figure 3, is the sum of the diffuse and two specular components:

$$f_r = \frac{\rho_d}{\pi} + F(\theta_d)$$
$$\times \left[\frac{\rho_{s,n}}{4\pi\alpha_n^2\sqrt{\cos\theta_i\cos\theta_o}}\exp\left(-\frac{\tan^2\theta_h}{\alpha_n^2}\right) \right.$$
$$\left. + \frac{\rho_{s,w}}{4\pi\alpha_w^2\sqrt{\cos\theta_i\cos\theta_o}}\exp\left(-\frac{\tan^2\theta_h}{\alpha_w^2}\right) \right] \quad (1)$$

where θ_i, θ_o, θ_h, and θ_d are the elevation angles of the incoming light direction, the outgoing light direction,
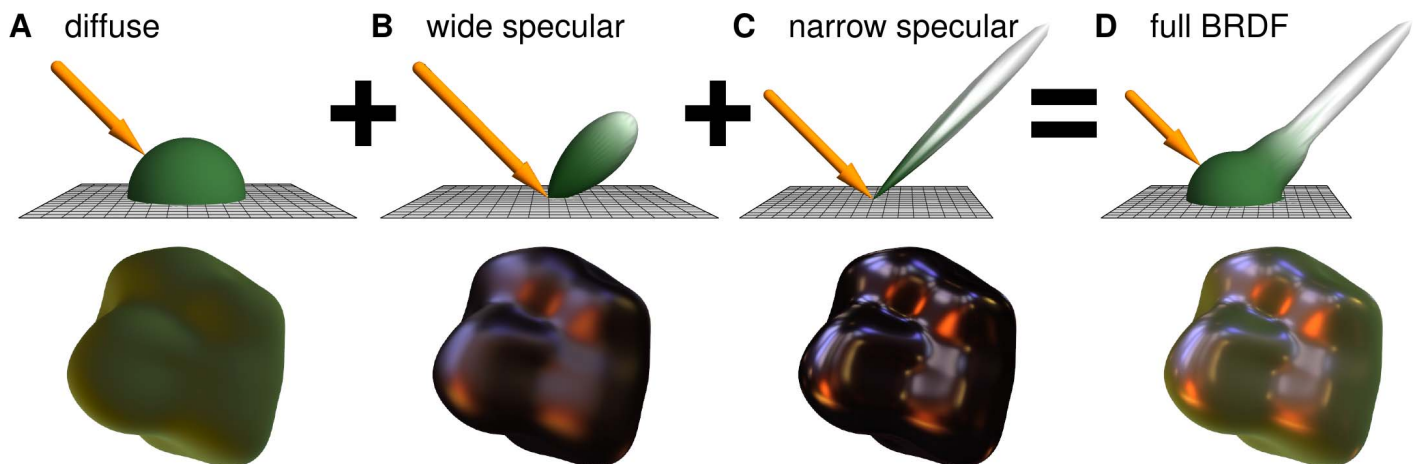

Figure 3. Graphical representation of our BRDF model (Equation 1) for incoming light elevation angle $\theta_i = 45°$, indicated by the orange arrows. The BRDFs of a diffuse component (A) and two different glossy components for blurry (B) and sharp reflections (C) are summed to obtain the full BRDF (D). Note that the cube root of the BRDF is plotted as is common practice to shorten the narrow spike of sharp components. Example objects made from the component materials are shown in the bottom row.
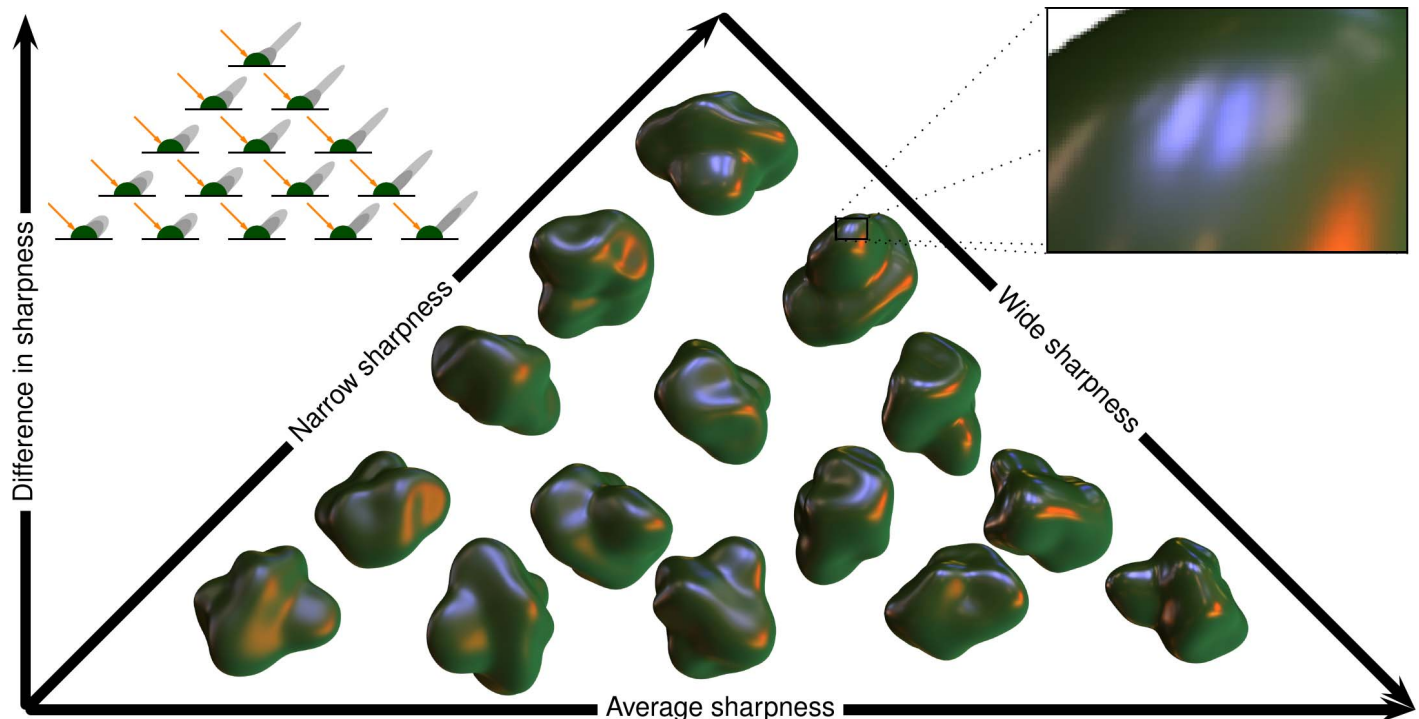
Figure 4. Triangular space of two-component glossy materials. Example target materials from the matching task are displayed at their correct location in the space. Only parameter values inside the triangle are valid according to our model (Equation 1). This limits the range of slider controls. Left inset: graphical representations of the corresponding BRDFs. Right inset: a zoom on a hazy reflection.

the halfway vector, and the difference vector, respectively (Rusinkiewicz, 1998). The Fresnel factor F modulates the angular contribution of the compound specular term and depends on the refractive index of the material. Note that we do not modify the direction of reflections as in off-specular models (Colbert, Pattanaik, & Krivanek, 2006; Lafortune et al., 1997) as this would require additional components to maintain the Helmholtz reciprocity principle of light transport (Helmholtz, 1856) and often produces unrealistic appearance.

This sum of two specular components is the simplest approximation to the true BRDF of various types of hazy materials. This approximation suffices for several types of layered materials such as metallic car paints and varnished or greasy surfaces (Upstill, 1990). It breaks down for highly complex scattering in multi-layered materials such as human skin, for which more advanced mathematical models exist (Donner & Jensen, 2005; Kubelka & Munk, 1931). A similar sum of specular components also occurs in polished materials, due to diffraction effects (Krywonos et al., 2011).

Participants

The participants in the following experiments were all university students aged 18–30. All participants were

volunteers, were naïve to the purpose of the experiments, and reported having normal or corrected-to-normal visual acuity and normal color vision. All experiments were conducted in accordance with the Helsinki protocol, with informed consent confirmed prior to the collection of data, and experimental procedures approved by the local ethics committee of the University of Giessen Psychology Department. Each participant performed only one of the following experiments. Each session lasted up to 1 h although no time limits were enforced.

Experiment 1: Matching task

This task used the green plastic materials. In each trial, participants were presented with two rotating objects side by side, as shown in Figure 2A. Their task was to adjust a single reflectance parameter of the object on the right (*match stimulus*) until it appeared to be the same material as (or as close as possible to) to the object on the left (*target stimulus*). The target material was a two-component material with two different sharpness parameters from the range $\alpha \in \{0.0705, 0.0985, 0.1265, 0.1545, 0.1825\}$. Figure 4 illustrates the resulting triangular stimulus space. The different parameter combinations were presented in pseudorandom order by the presentation software.

Free parameter	Constant parameter
Narrow sharpness α_n	Wide sharpness α_w
Wide sharpness α_w	Narrow sharpness α_n
Average sharpness α_{avg}	Difference in sharpness $\Delta\alpha$
Difference in sharpness $\Delta\alpha$	Average sharpness α_{avg}

Table 1. Overview of the two-component matching conditions.

In one block of trials (*two-component condition*), a single slider gave the observer control over a single parameter of the match material, while the coupled parameter was held constant at the correct value of the target material; that is, when the observer controlled the narrow sharpness, the wide sharpness was automatically kept at the same value as in the test stimulus and vice versa. Also, when the observer controlled the average sharpness, the difference in sharpness was kept at the same value as in the test stimulus and vice versa (see Table 1). Therefore, for the two-component condition, it was always possible to navigate to exactly matching material parameters.

In the other block of trials (*single-component condition*), a single slider gave the observer control over the sharpness parameter of a single-component material presented on the right. The target material was still a two-component material as above, so an exact match could never be reached. However, this allowed us to measure how participants perceptually weighed the different components of the two-component BRDFs: would they match based on the narrow component, the wide component or some average value, when forced to compromise?

In both blocks, the 16 participants were instructed to "match the material on the right to the material on the left in terms of the sharpness or blurriness of the reflections." Each participant performed this task for 15 (target materials) × 5 (slider control conditions) × 3

(repetitions). Participants viewed the stimuli on the monitor under normal office lighting.

Participants were not told what the parameters of the materials were or which parameter was controlled by the slider in the current condition. Instead they could only try out the effect of the slider and explore the range of materials that could be obtained. This avoids potential misunderstandings caused by language or terminology. Participants were able to perceive the effect of the slider on the material, even though the range of the slider was very limited in a few of the conditions. Participants were generally satisfied with the closest match that they had found, even if they did not always perceive it as a *perfect* match.

Based on early pilot trials, matching two parameters simultaneously using two sliders or a two-dimensional control would be too difficult. The two sliders or dimensions would not necessarily be perceptually meaningful or intuitive, and it would be more difficult to explore an entire two-dimensional range than a one-dimensional range of materials. In the authors' experience, participants would often settle for a bad match in a two-dimensional condition. Hence, only one-dimensional conditions are used in the present study.

Results

Figure 5 shows the results of the four slider control conditions in the first block of trials. Most participants perform quite well when the slider controls the average sharpness (Figure 5A), difference in sharpness (Figure 5B), or narrow sharpness (Figure 5C), although in each case a few participants perform close to chance. On the other hand, when the slider controls the wide sharpness (Figure 5D), performance is closer to chance than to ideal performance.
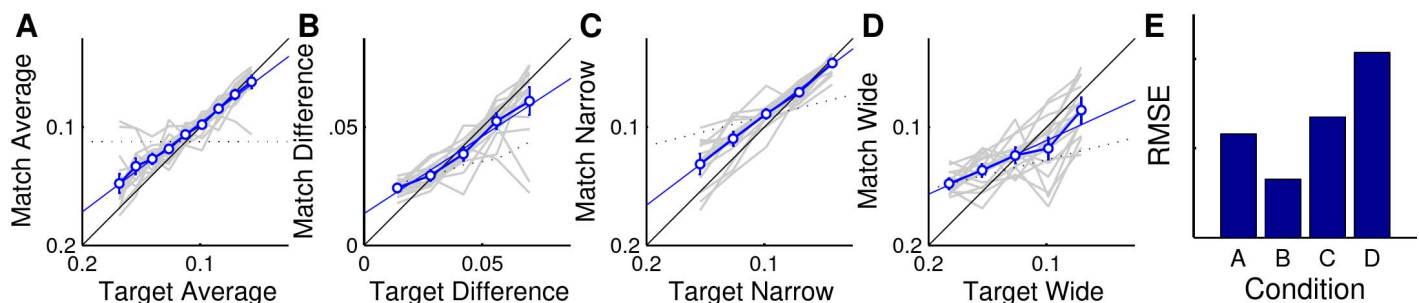

Figure 5. Results of the two-component matching task. Participants controlled a single parameter: (A) average sharpness, (B) difference in sharpness, (C) sharpness of the narrow component, or (D) sharpness of the wide component. The thick blue line indicates the mean matched parameter, with error bars delimiting the 95% confidence interval. The thin blue line is the linear regression result with $R^2 = 0.58, 0.33, 0.56,$ and 0.14, respectively. The separate curves for the $N = 16$ participants are shown in light gray. The diagonal is the veridical match. Chance performance is indicated by the dotted line, which sometimes has an unusual shape because of the binning of the results and the restrictive triangular gloss space. (E) The RMS error quantifies the departure from veridicality in each condition.
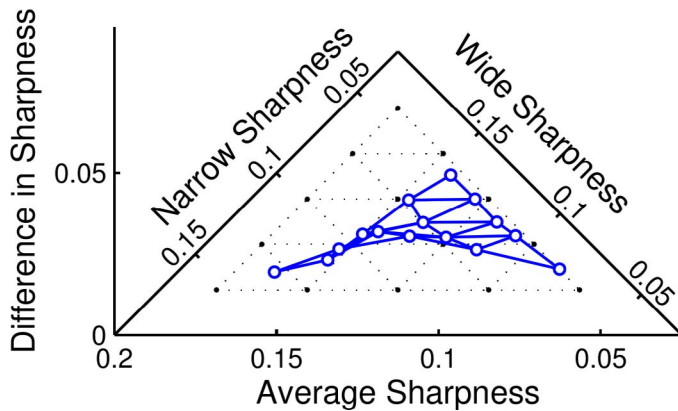
Figure 6. Gloss space representation of the results of the two-component matching task (smaller values of α lead to sharper reflections). The blue circles represent the average matches for each target. The black dots are the true target positions. Connecting lines clarify the deformed structure of the perceived gloss space.

Some of the lines representing chance performance have unusual shapes because of the restrictive gloss space. For target materials in a corner of the space, some of the slider controls have a short valid range. This also reduces the error bars representing the 95% confidence intervals of the mean match. Moreover, the means and confidence intervals are not all based on the same number of trials because the target stimuli are binned according to the relevant parameter.

Figure 6 shows the gloss space representation of the results of the two-component matching task. The space of average matches is severely compressed towards the middle of the valid space. The variance of these average matches is so large that representing it as ellipses would clutter the graph. The least compression and the best matches are in the region of high narrow sharpness and medium to high wide sharpness (i.e., the rightmost two diagonal columns of the plot). Interestingly this is the region where the authors find the two components most clearly distinguishable and get the most salient impression of haze (see Figure 4). One interpretation could be that a perceptual decomposition of sharp and wide specular components is difficult unless there is a clear and distinct difference between these two components: otherwise they are treated as a single component and there is no impression of haze.

However, taken in isolation, these results do not prove that participants perceptually separated the material into two components. In each condition only a single parameter is kept correct, while the other *three* vary: one directly, the other two indirectly. In all cases where performance is high, the narrow sharpness varies and can be used to produce the match, while in the unsuccessful case, the narrow sharpness is kept constant. This means that observers might attend to the narrow shading features to judge material similarity in

this task. Observers might also attend to other shading features related to hazy gloss, which are made more prominent when the narrow component is sharpest.

Figure 7 shows the results of the single-component matching task. Each graph shows the same data, but binned according to different axes. The same binning caveat as for Figure 5 applies, but in this condition the range of the slider control is the same for each target material.

Since there is no correct match in this condition, success must be defined as the consistency between participants and consistency with plausible models (see prediction curves in Figure 7). None of the participants reported that it was impossible to provide a perfect match. The shape and position of the mean curve in each of the graphs is very similar to the prediction that observers match the narrow component (green).

Figure 8 shows the gloss space representation of the results of the single-component matching task. Single-component materials can be thought of as two-component materials with difference in sharpness $\Delta\alpha = 0$ (i.e., located on the average sharpness axis). The arrows show that observers tended to choose a single sharpness close to that of the narrow component, which lies in the direction 45° down and to the right. The best match usually deviates a little from that narrow sharpness towards the average sharpness, which lies straight below the target. The polar plots show skewed and even bimodal probability distributions of the matches for some of the targets with a small difference in sharpness. This bimodality exists both between and within participants, and provides an initial hint that observers may be able to attend to each of the two components in the BRDF independently; that is, that some kind of decomposition into distinct causes or layers occurs.

Together the findings suggest that at least for some parameter ranges, particularly on the upper right edge of the stimulus space, participants are able to match the properties of two-component specular materials. However, to put this more rigorously to the test, we performed a discrimination task to measure the parameter ranges for which participants can reliably distinguish between single- and two-component materials by focusing on that upper right edge of the space and also examine the relative strength of the two components as an additional parameter.

Experiment 2: Discrimination task

We performed a four-alternative forced-choice (4AFC) discrimination task in which one of the four presented stimuli had a two-component material, while the other three all had identical single-component
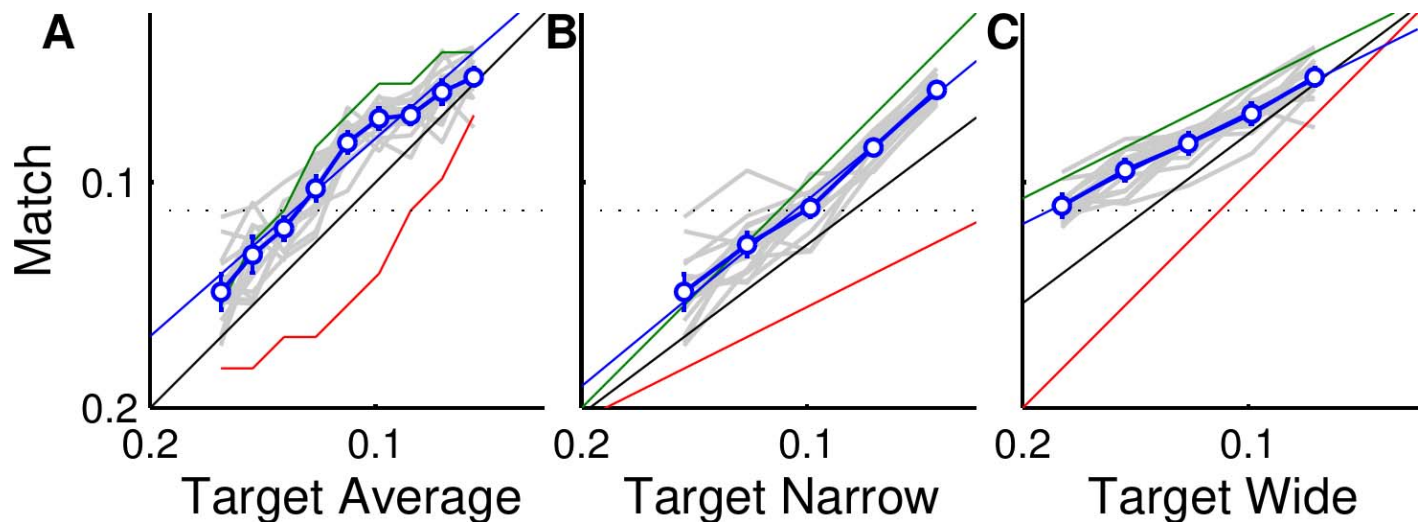
Figure 7. Results of the single-component matching task. Participants controlled a single sharpness parameter. The same data are binned according to (A) average sharpness, (B) sharpness of the narrow component, or (C) sharpness of the wide component of the two-component target. The thick blue line indicates the mean matched parameter, with error bars delimiting the 95% confidence interval. The thin blue line is the linear regression result with $R^2 = 0.53$, 0.62, and 0.22 respectively. The separate curves for the $N = 16$ participants are shown in light gray. The prediction that observers match the narrow component is shown in green. The prediction that observers match the wide component is shown in red. The prediction that observers match the average sharpness is shown in black. Chance performance is indicated by the dotted line.

BRDFs, using the dark shiny plastic material, as shown in Figure 2B. On each trial, participants simply had to identify which material was the odd one out; their ability to do so provides a test of the extent to which
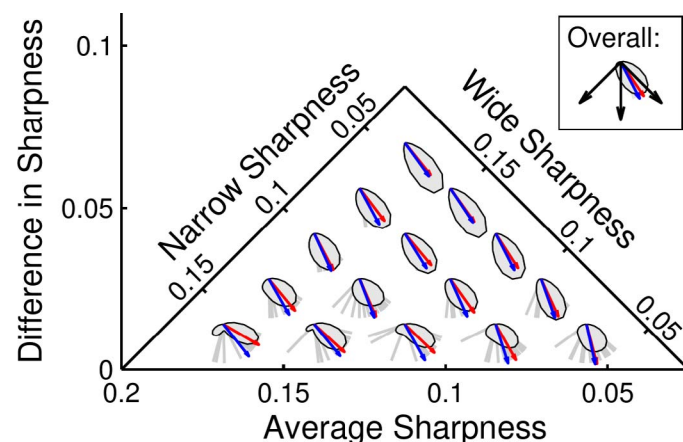

Figure 8. Gloss space representation of the results of the single-component matching task (smaller values of α lead to sharper reflections). Single-component materials can be thought of as two-component materials with difference in sharpness $\Delta\alpha = 0$ (i.e., located on the average sharpness axis). At the position of each two-component target material, the blue and red arrows point in the direction of the mean and median single-component match respectively. The separate matches for the $N = 16$ participants are shown as light gray lines. The probability distribution of the matches is shown as a small polar plot derived from the separate matches using kernel density estimation. The inset shows the average over all targets.

they could perceive differences between single- and two-component materials.

The two-component target material had a narrow component with sharpness $\alpha_n = 0.0396$ and a wide component with sharpness chosen from the range $\alpha_w \in \{0.0833, 0.1125, 0.1417, 0.1708, 0.2\}$. In other words, the difference in sharpness was chosen from the range $\Delta\alpha \in \{0.0218, 0.0364, 0.0510, 0.0656, 0.0802\}$. The contrast between the two specular components was varied by changing their reflectances. These were chosen from the range $\rho_s \in \{0.05, 0.10, 0.15, 0.20, 0.25\}$ in such a way that they always summed to the same total reflected energy as the single-component material with reflectance $\rho_{s,total} = 0.30$ (i.e., there was a tradeoff between the reflectance of the two components, such that when one was high, the other was low). The resulting rectangular stimulus space is shown in Figure 9. Note that this space is not directly comparable to the triangular space used in the matching experiment, as the dimensions vary different parameters of the reflectance model. In three separate blocks of trials, the single-component material (i.e., the distractors) used the sharpness parameter of either the narrow or wide component of the target material, or used the average sharpness. Three such distractors are shown in the left column of Figure 10.

The diffuse component in Experiment 1 could have obscured the intended differences in the specular components; hence, we used $\rho_d = 0$. Unlike the matching experiment, we performed the 4AFC dis-crimination task in a darkened room to improve the
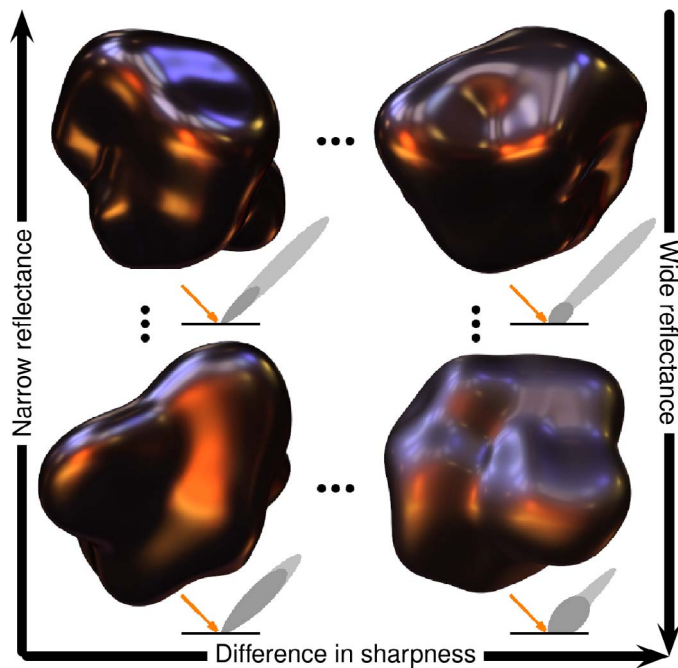
Figure 9. Space of two-component glossy materials used in the discrimination experiment. Example materials of the four corners of the space and graphical representations of their BRDFs are displayed at their correct location in the space.
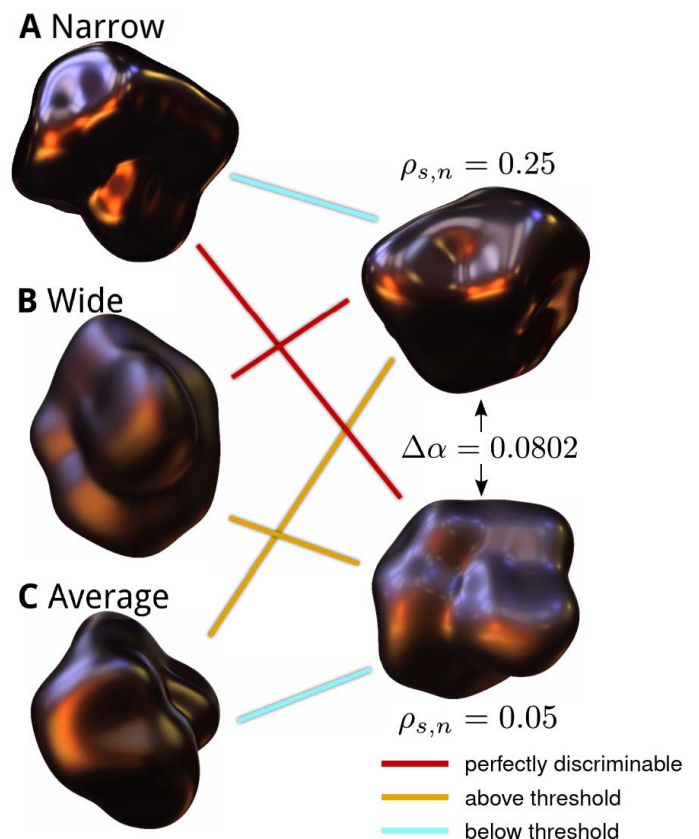

Figure 10. Single-component distractors (left column) are chosen to match the narrow (A), wide (B), or average (C) sharpness of two-component stimuli (right column), here with $\Delta\alpha = 0.0802$. The connections use the same color code as in Figure 11: cyan lines are below discrimination thresholds, orange lines are above, and red lines correspond to 100% correct discrimination.

contrast of the stimuli on the computer screen in order to further optimize the performance of the observers.

The 15 participants were instructed to indicate "which material looks different from the others in terms of the sharpness or blurriness of the reflections." Each participant performed this task for 25 (target materials) $\times$ 3 (single-component conditions) $\times$ 10 (repetitions).

Results

Figure 11 shows the results of the discrimination task. As before, performance increases as the difference in sharpness between the two components grows. For all three classes of distractor, there are ranges of corresponding two-component materials that are easily distinguished.

The main differences between conditions may be explained by the position of the single-component stimulus as a limiting case of the two-component material. When the reflectance of the narrow component tends to 0.30, the reflectance of the wide component becomes so low that the two-component material consists mostly of its narrow component. Conversely, when the reflectance of the narrow component goes toward 0, the two-component material becomes similar to its wide component. This would explain the opposite slopes of the performance curves

relative to the $\rho_{s,n}$ axis in the first and second plots. However, when the single-component material uses the average sharpness, the task also becomes more difficult as the reflectance of the narrow component decreases (Figure 11C). However, Figure 10 suggests that a simple interpretation based solely on the narrow component is unlikely. It visualizes the discrimination task using images of the actual stimuli for the rightmost column of Figure 9. The connecting lines follow the color code of Figure 11, with red lines corresponding to most discriminable pairs. In particular, it is unlikely that the red connection between the single-component stimulus in Block A and the two-component stimulus at the bottom right is due to a vanishing narrow component alone, since the narrow component remains clearly visible when $\rho_{s,n} = 0.05$.

Nevertheless, the discrimination experiment suggests that participants relied heavily on the narrow component since best performance was obtained in the second condition (in which the single-component distractors had the same sharpness as the wider of the two
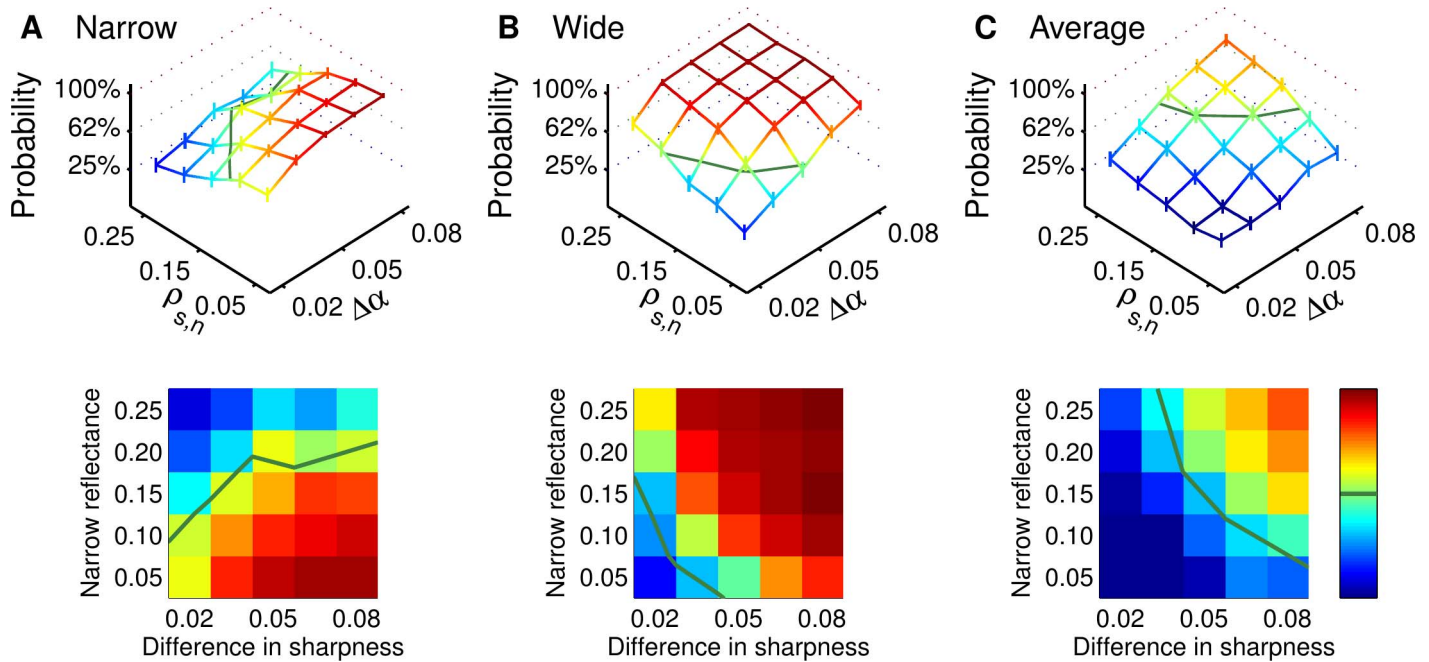
Figure 11. Results of the discrimination task for $N = 15$ participants for the conditions where the single-component material used the (A) narrow, (B) wide, or (C) average sharpness of the two-component target material. In the top row, the vertices of the color-coded surface indicate mean performance over all participants as a function of the difference in sharpness ($\Delta\alpha$) and the reflectance of the narrow component ($\rho_{s,n}$), with error bars delimiting the 95% confidence interval. The dotted lines indicate chance (25%), threshold (62.5%), and ideal performance (100%). The bottom row is a top-down view of the same results. The dark green line in both rows is the contour of threshold performance.

components). As in the matching experiment (see Figure 6), performance was best when the two lobes were easily distinguishable from one another. In all conditions there were parameter ranges for which participants could distinguish the presence of a two-component target from corresponding single-component distractors. However, this in itself does not tell us about the subjective interpretation of these differences. Does the ability to distinguish between different BRDFs reflect a distinct perceptual parameter, or are the image differences detectable but not interpretable? To test this, we asked participants to perform a series of ratings for two-component materials with different parameter values.

Experiment 3: Rating task

In each trial, a single stimulus was presented and participants were asked to provide subjective ratings of six different surface qualities by adjusting the slider positions as shown in Figure 2C. Each slider had two text labels at opposite ends to indicate the range of potential values.

The materials had a silver-metallic appearance, with parameters similar to the two-component target material in the discrimination task. The sharpness of

the narrowest component was $\alpha_n = 0.0396$ and the sharpness of the widest component was chosen from the range $\alpha_w \in \{0.0833, 0.1125, 0.1417, 0.1708, 0.2\}$. In other words, the difference in sharpness was chosen from the range $\Delta\alpha \in \{0.0218, 0.0364, 0.0510, 0.0656, 0.0802\}$. The diffuse reflectance $\rho_d = 0$ and the reflectances of the two components summed to the total $\rho_{s,total} = 0.95$, which is typical for silver. The ratio of the reflectances of the two components was chosen from the range $\rho_{s,\{n,w\}}/\rho_{s,total} \in \{1/6, 2/6, 3/6, 4/6, 5/6\}$. This experiment used a different lighting environment representing another church interior, Galileo's Tomb (Debevec, 1998). The resulting rectangular stimulus space is shown in Figure 12.

The 14 participants were instructed to "rate the presented material on the following six different continuous scales related to gloss appearance":

- glossy versus matte,
- sharp versus blurry,
- not hazy versus hazy,
- polished versus unpolished,
- low versus high friction, and
- not coated/varnished versus coated/varnished.

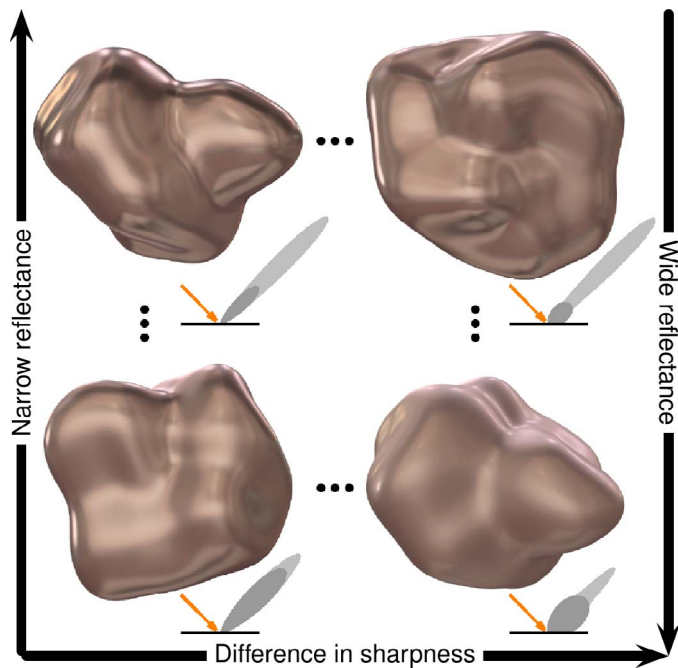Each participant performed this task for 25 (materials) × 2 (repetitions).

Figure 12. Space of two-component glossy materials used in the rating experiment. Example materials of the four corners of the space and graphical representations of their BRDFs are displayed at their correct location in the space.

Results

Figure 13 shows the mean results of the rating task across participants. There are several notable aspects of the results. First, for the properties that directly refer to optical appearance of the material (matte, blurry, and hazy), there are clear and systematic patterns of responses. This is reflected in the range of mean values that participants used, which, for these three properties, were all greater than 0.4 out of a possible 1.0 (see Figure 14). In contrast, for the properties that refer to the way the object was created or its haptic qualities (unpolished, high friction, and coated/varnished), the responses were restricted to ranges less than 0.3, suggesting that the observers did not consistently see large differences between the stimuli in these qualities.

We find that for this range of stimuli, matteness, blurriness, and appearing unpolished are associated with materials with a prominent wide component, more or less irrespective of the difference in sharpness between the two components, illustrated by the fact that the highest rating values appear all across the bottom edge of the stimulus space. In contrast, haze and appearing coated or varnished are also associated with a prominent wide component, but only when there is a large difference in sharpness between the two components, illustrated by the fact that the highest rating values appear only on the right side of the bottom edge. In other words, the subjective impression
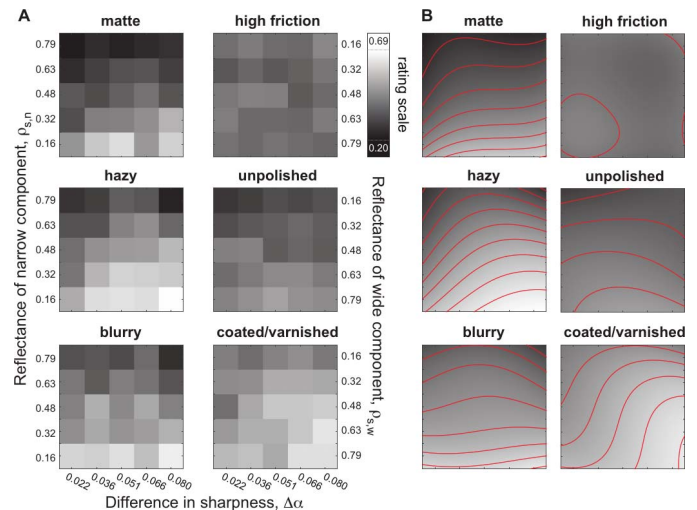

Figure 13. (A) Mean ratings across participants for each of the six properties. Gray values indicate mean rating, consistently normalized such that black represents the lowest and white the highest average ratings across all stimuli and properties. (B) Fits of a third-order polynomial surface to the data presented in each panel of A, with superimposed contours to facilitate visualization of the main trends in each plot.

of hazy or layered materials is crucially associated with a "bloom" or "halo" around sharp reflections. There is little in the way of systematic effects on the perception of friction: essentially all the materials were perceived as having low friction (mean rating of 0.38, and a range of just 0.1).

The difference between blurry/matte/unpolished appearances and hazy-coated/varnished appearances is made clearer through a principal component analysis (PCA) on the mean responses. Figure 15 plots the factor loadings of each property in the space spanned by the first two principal components, which together
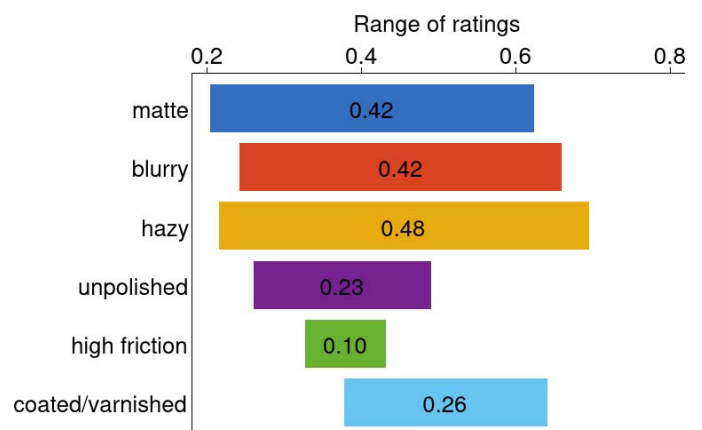

Figure 14. Range of mean responses given to different stimuli for each of the six subjective properties, which serves as a measure of how consistently different from one another the stimuli appeared to the observers.
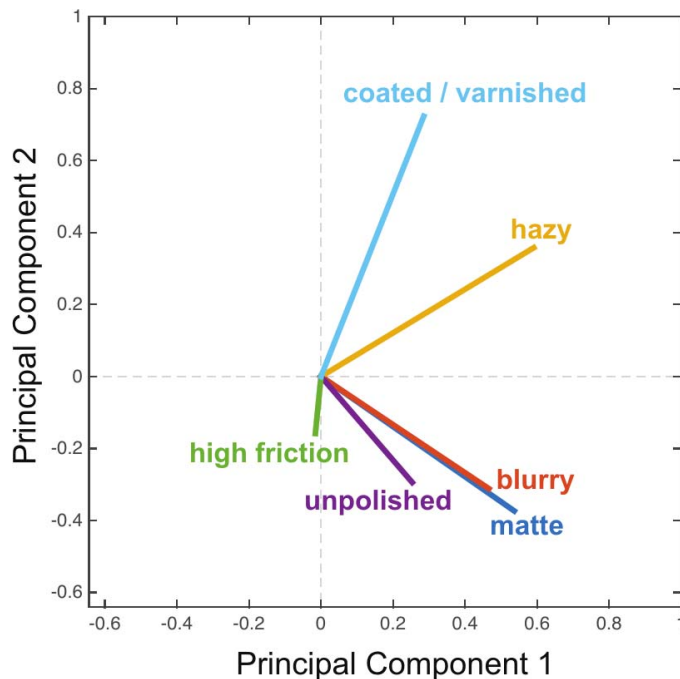
Figure 15. Directions of the six rating scales represented by unit vectors in the principal component space, projected onto the plane formed by the first two principal components which together account for 93.6% of the variance. Nearly orthogonal rating scale directions, such as blurry and hazy, are almost completely uncorrelated. This indicates that their trends in Figure 13 are indeed different and the scales represent distinct perceptual dimensions.

account for 93.6% of the variance in the mean data. The plot shows how blur, matteness, and appearing unpolished are correlated with one another, while haziness and "coatedness" are approximately orthogonal to this cluster. This suggests that haze truly is a distinct perceptual dimension of reflectance, independent of blur, and associated with the presence of both sharp and blurry reflections simultaneously.

Discussion

Summary of results

The two-component matching task showed that observers were able to match three of the four parameters (average and difference in sharpness as well as the sharpness of the narrow component, but not of the wide component; see Figure 5). This seems to suggest that they have access to a perceptual representation of the distinct components of the BRDF, as if it were perceptually segmented in some way. The best matches were obtained when the narrow component is sharpest, as illustrated by the regions of least com-

pression in Figure 6. A bias toward the narrow component was also observed in the single-component matching task, as shown both by the predictions of Figure 7 and the probability distributions of Figure 8. However, the match was not exact, and slightly less sharp than the narrow component.

The discrimination task showed that there are high performance regions in the rectangular stimulus space (including the reflectance dimension) where observers can clearly distinguish between two-component targets and any of the single-component distractors. This provides further support to the idea that participants are perceptually sensitive to the additional complexity of two-component BRDFs. However, as seen in Figure 11, performance regions differed markedly between Block A (narrow sharpness assigned to single-component materials) and Blocks B or C (wide or average sharpness assigned to single-component materials). Performance thus depended on whether the narrow sharpness of the two-component target material was identical (Block A) or different (Blocks B and C) from the single-component distractors. The best discrimination results were obtained in Block B, where the narrow and single component sharpnesses are the most different.

Finally, the rating task suggests that the presence of two components with different sharpness levels is associated with haziness and the impression that a material has been coated or varnished. The fact that this appearance is particularly associated with a prominent wide component that is much broader than the narrow component on which it is superimposed further suggests that the visual system somehow decomposes two-component BRDFs into distinct terms of some kind. Indeed, when the two components are not distinguishable from one another, reflections are not perceived as hazy, suggesting that there is an intimate connection between perceiving haze, and parsing reflections into multiple contributions.

Candidate interpretations

Hunter and Harold (1987) proposed a number of haze measurement techniques. Even though they are not intended as models of human visual processing, they are widely used in industrial applications. The one retained by the ASTM (1997) is a ratio of reflectances H/S, with S measured in the specular direction and H measured in an off-specular direction, for an incident light elevation of $\theta_i = 30°$. The choice of angle θ_{off} between specular and off-specular directions may vary depending on the materials considered; standard values are 2° and 5°. Note, however, that in all cases, such measurements do not involve any explicit decomposition of the reflections into multiple terms. When plotted

A. Haze ratings | B. Haze ($\theta_{off} = 5°$) | C. Haze ($\theta_{off} = 20°$) | D. Excess kurtosis | E. Halo energy | F. Central sharpness
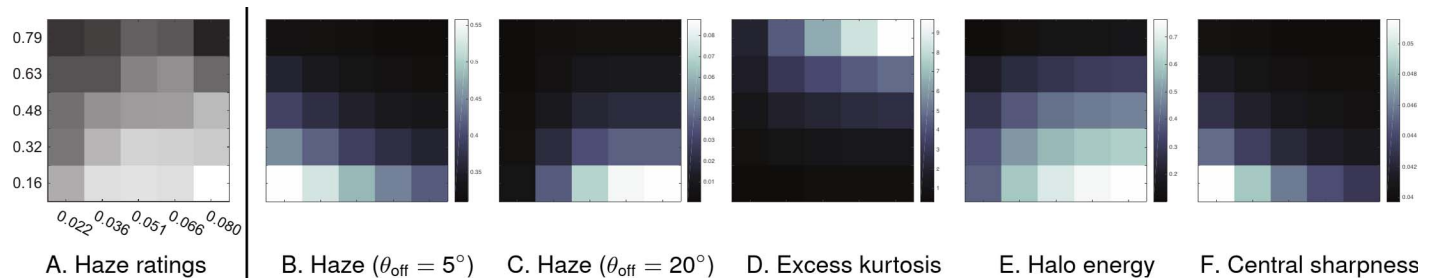
Figure 16. Candidate interpretations for (A) the mean haziness ratings (repeated from Figure 13); each plot has been normalized independently as shown by their respective color bars. Hunter's haze measurements at $\theta_{off} = 5°$ (B) do not appear related to perceived haze. Only when pushed to $\theta_{off} = 20°$ (C) do they begin to show a similar distribution. Excess kurtosis (D) is necessary to elicit hazy gloss percepts, but is not related to perceived haze either. Using the decomposition explained in Figure 18, we extract a halo component (E) whose energy is distributed similarly to haze ratings. The sharpness α_c of the other (central) component (F) is distributed along a different diagonal direction in the stimulus space and is similar to (B).

in the rectangular stimulus space of the discrimination and rating experiments, these two haze measurements give essentially the same results. As seen in Figure 16B, Hunter's haze for $\theta_{off} = 5°$ does not appear to be aligned with participants' haze ratings. Only if we push the value of θ_{off} to 20° do we obtain a similar distribution of values (Figure 16C). The choice of θ_{off} must thus be tailored to the class of materials at hand, which is not general enough for a perceptual theory of hazy gloss.

It is highly unlikely that the human visual system refers to only two specific angles to infer haze gloss, but rather to some image measurements derived from the whole material response. One simple possibility is to consider higher-order moments of the compound specular term. Summing multiple BRDFs with different widths affects the kurtosis of the compound distribution. Indeed, Figure 17A shows that the BRDF becomes leptokurtic when a second specular component (of different sharpness) is used. Note that this is different from looking at the kurtosis of the image histogram (e.g., Motoyoshi, Nishida, Sharan, & Adel-

son, 2007): here we only consider the reflectance distribution function. This function is similar to taking the pixel values along a line through a highlight in an image. A single Ward BRDF component is a Gaussian and therefore has excess kurtosis $k = 0$. The excess kurtosis of a two-component BRDF—a mixture of Gaussians—is positive and can be computed analytically. Note, however, that kurtosis, like Hunter's haze measurement, does not explicitly decompose the specular reflections into the component Gaussian. As shown in Figure 16D, we find that kurtosis is not aligned with participants' haziness ratings. Hence, even though a material must exhibit kurtosis to elicit hazy gloss, kurtosis does not seem to characterize our subjective experience of haziness, at least for this range of materials.

We therefore propose that the visual system may represent haze by decomposing the material response into two distinct components, or causal layers. One obvious choice for such a decomposition would be the physical components themselves (i.e., the broad and narrow specular terms). This decomposition separates the composite reflection into *two superimposed layers*, much like the decomposition of image patches into two superimposed surfaces in transparency perception (Anderson, 1997; Barrow & Tenenbaum, 1978; Beck, Prazdny, & Ivry, 1984; Katz, 1935; Metelli, 1974). This would be broadly consistent with our findings that observers base many of their judgments on the narrow component. However, this is not the only possible decomposition. An alternative that we find also predicts many aspects of the results—especially the subjective ratings of haze—is a "hybrid" decomposition, consisting of a central Ward BRDF component that matches the reflectance peak and spread around the specular direction, and a surrounding halo component that corresponds to the (positive) residual reflectance (if there is any). In other words, it is a decomposition of the reflection into *spatially adjacent, juxtaposed components*, rather than two superimposed
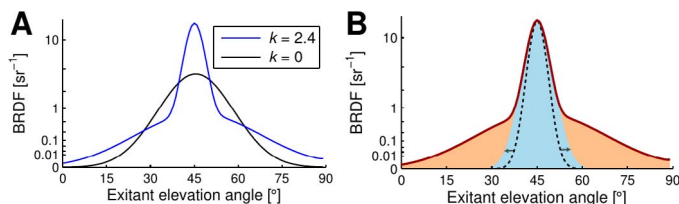

Figure 17. (A) Comparison of the kurtosis of two- versus single-component specular BRDFs as functions of the exitant angle for an incident angle of $\theta_i = 45°$. The two-component BRDF (blue line) has positive excess kurtosis k, whereas the single Ward BRDF component (black line) has excess kurtosis, $k = 0$. (B) The hybrid decomposition is obtained in two steps: the narrow component is first scaled and enlarged to yield a central peak (in blue), which is then subtracted from the compound specular term to yield the surrounding halo (in orange).
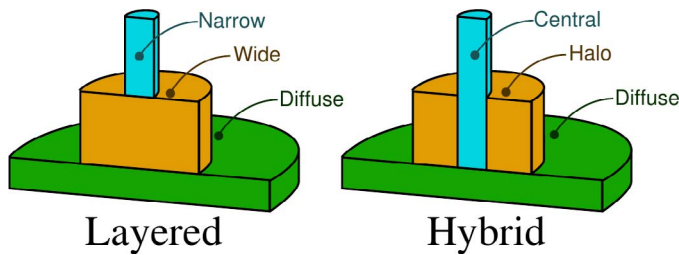
Figure 18. Schematic illustration of the layered decomposition based on the physical reflectances, and the hybrid decomposition that better predicts perceived haziness.

layers. Figure 17B illustrates the procedure we use to perform this decomposition: we scale a Ward BRDF component to match the peak reflectance (dashed curve) and make it as wide as possible (in blue) to fit inside the compound specular term; the halo component (in orange) is then simply obtained by subtraction. We write ρ_c and α_c for the reflectance and sharpness of the central component, and use $\rho_h = 1 - \frac{\rho_c}{\rho_{s,n} + \rho_{s,w}}$ to characterize the halo energy. The latter is equal to 0 when $\alpha_c = \alpha_n = \alpha_w$ as the central component encompasses the whole compound specular term. We plot the halo energy in the rectangular stimulus space in Figure 16E. It shows a clear alignment with mean haziness ratings. We also plot the sharpness α_c of the central component in Figure 16F. Note its strong similarity to Hunter's haze measurement at $\theta_{off} = 5°$ (Figure 16B). This is because the central component is the most prominent component in a direction 5° off the specular direction in our stimulus set, making this angular configuration ill-adapted to measure haze in our case.

The hybrid decomposition thus suggests that the human visual system may decompose compound BRDFs into two perceptual components that are different from the physical specular terms. Figure 18 illustrates the difference between the physical, or layered, decomposition and the hybrid decomposition (we add the diffuse layer in green for completeness), inspired by a similar illustration about lightness perception (Gilchrist, 1994, figure 1.2, p. 30). If our perceptual representation indeed follows this hybrid decomposition, then we might expect the central and halo components to account for the results of our experiments. To this end, we have performed a number of linear regressions on our experimental data using either the parameters of the physical or hybrid decompositions. A representative subset of this analysis is presented in Figure 19. It should be noted that the evidence derived from the current experiment is mixed, so additional, more targeted experiments would be required to strongly distinguish between layered and hybrid decompositions.
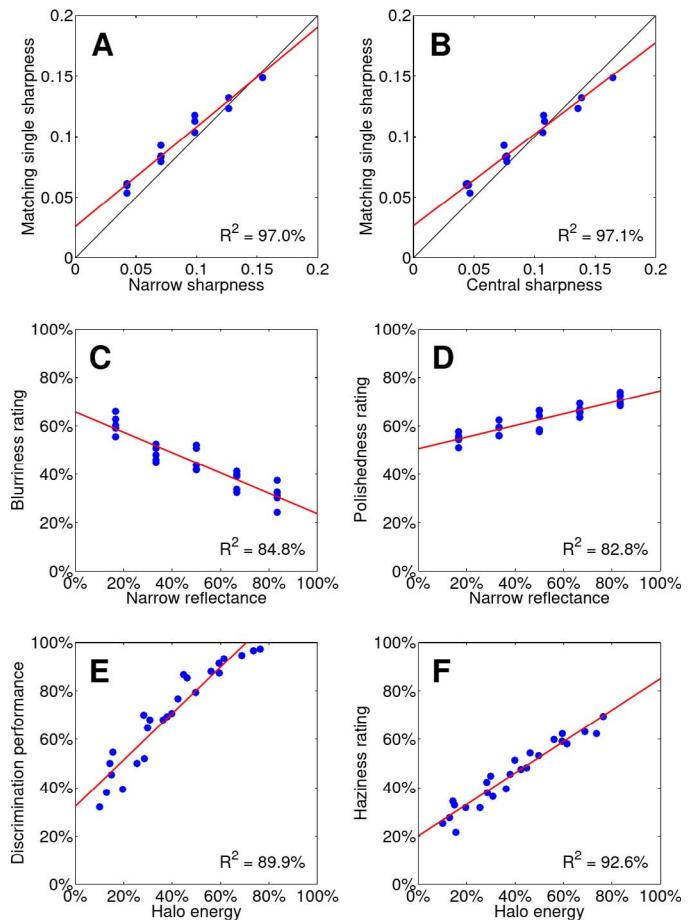

Figure 19. The sharpness data from the single-component matching task shows a strong linear correlation (red line) both with the narrow sharpness α_n of the physical decomposition (A) and with the central sharpness α_c of the hybrid decomposition (B). Using these sharpness parameters as direct predictors of the best matching single sharpness (black diagonal) yields slightly more accurate predictions in the wide end or middle part of the sharpness range, respectively. The mean blurriness (C) and polishedness ratings (D) both show a linear correlation with the reflectance of the narrow component $\rho_{s,n}$ of the physical decomposition. The performance data from the discrimination task (block A) between a two-component material and a single-component distractor matching the narrow sharpness (E) show a reasonably linear correlation to the halo energy ρ_h in the hybrid decomposition. The mean haziness ratings (F) show a strong linear correlation with the halo energy.

Our first observation is that the sharpness parameters α_n and α_c of the physical and hybrid decompositions are strongly correlated ($R^2 = 99.9\%$) in the matching tasks (α_n is fixed in the other tasks). In particular, they both show a strong linear correlation with matching data in the single-component task, as shown in Figures 19A and B. However, neither of these parameters provide an accurate enough account as data is more compressed toward the middle of the range.

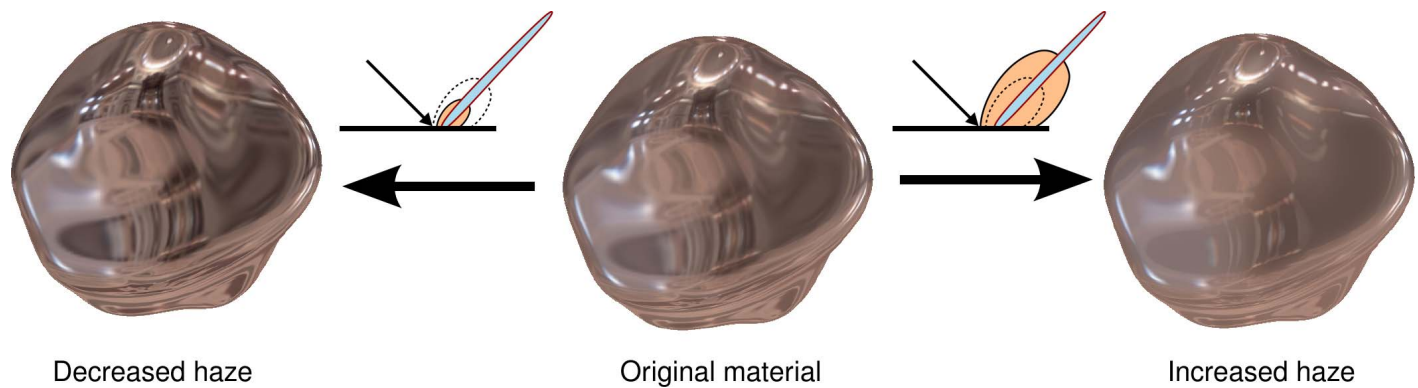| Decreased haze | Original material | Increased haze |

Figure 20. The halo intensity of the original material (dashed curve) may be increased (right) or decreased (left), yielding a corresponding increase or decrease in perceived haziness.

This might be due to an inherent difficulty in the matching of single- and two-component materials, which is also suggested by the bimodality of results shown in Figure 8.

In contrast, in the discrimination and rating tasks, the physical and hybrid decompositions each are better predictors for specific results. For example, Figures 19C and D show that both the blurriness and polishedness ratings are well correlated with the reflectance $\rho_{s,n}$ of the narrow component of the physical decomposition. This is not the case for any of the parameters of the hybrid decomposition.

The discrimination task results show large regions of high performance that differ between experiment blocks. We suggest that these differences can be partly explained by the hybrid decomposition. Indeed, as shown in Figure 19E, performances, when comparing to distractors of narrow sharpness, are linearly correlated to the halo energy ρ_h, suggesting that discrimination was directly based on haze (for a visual example, see Figure 10A). However, when comparing to distractors of wide or average sharpness (Blocks B and C), performances are inversely related to the sharpness of the central component α_c, and show nonlinearities (especially in Block C). We leave the study of these nonlinearities to future work, as they do not concern the perception of hazy gloss.

More importantly, the halo energy ρ_h exhibits a strong linear correlation with mean haziness ratings, as shown in Figure 19F. This confirms the observed similarity between Figures 16A and E. None of the parameters from the physical decomposition provides such a strong correlation, suggesting the hybrid decomposition better characterizes subjective ratings of gloss haze.

Image cues to hazy gloss

Figure 20 shows the effect of manipulating the intensity of the halo term while leaving the central peak of the hybrid decomposition unchanged. The material appears less hazy or more hazy when the halo intensity is decreased or increased respectively, suggesting that the hybrid decomposition could be useful for the editing of material appearance in computer graphics applications. Moreover, since the central peak and halo of the hybrid decomposition are additive, they may be computed separately during rendering, and the resulting peak and halo images added together in post-process. As a result, the editing of Figure 20 may be equivalently obtained by manipulating the intensity of the halo image before adding it to the peak image. This raises the question of which image cues are involved in the perception of hazy gloss, and how they could be manipulated to alter the perception of haziness.

In terms of the image cues they produce, objects exhibiting hazy gloss may be related to glare effects produced by strong light sources or highlights. However, glare is mainly due to properties of the retina; it has recently been suggested that it might be directly processed by on–off cells (Sato, Motoyoshi, & Sato, 2016). In contrast, haze is a property of the object surface, which makes it dependent on object shape (e.g., curvature and slant). Moreover, the perception of hazy gloss will likely be influenced by the environment lighting, similar to other types of gloss (Doerschner, Boyaci, & Maloney, 2010). An exciting avenue of future work is thus to understand how the human visual system encodes hazy gloss and distinguishes it from glare in images.

Conclusions

We have shown through a series of three experiments that haze likely constitutes a quality of perceived gloss that is distinct from blur or contrast. The matching experiment results showed a range of configurations

where hazy gloss makes a visual difference to surface appearance, namely when the narrow component of the material is sharpest. The discrimination experiment showed that haze could indeed be used as a cue to compare material appearances when the narrow component of target and distractors are similar (Block A). The rating task verified that haziness is not only a readily perceivable material quality, but that it is distinct from the blur quality, as evidenced by the distinct directions for haze and blur in the 2D PCA plot.

We have proposed that the visual system parses compound BRDFs into multiple components, which can form the basis of matching and discrimination. In particular, we find that certain aspects of our data can be explained by a nonphysical decomposition into a central reflection peak flanked by a halo component. We suggest that it is the presence of the halo component that is responsible for the perception of hazy gloss. Although the data from the matching experiment does not distinguish between the hybrid and a layered decomposition into the physical components, the hybrid model does provide a better account of the discrimination (Block A) and rating (haziness) experiments. Our findings suggest that the standard industrial measurements of haze gloss may need adjustment to account for human perception across a wider range of materials. Further experiments will also be required to assess the perceptual relevance of the hybrid decomposition.

Acknowledgments

Commercial relationships: none.
Corresponding author: Peter Vangorp.
Email: peter.vangorp@edgehill.ac.uk.
Address: Department of Computer Science, Edge Hill University, Ormskirk, UK.

References

Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception*, *26*(4), 419–453. [PubMed] [Article]

ASTM. (1997). *Standard test methods for measurement of gloss of high-gloss surfaces by goniophotometry* (Standard No. E430-97). West Conshohocken, PA: American Society for Testing and Materials. [Article]

Bagher, M. M., Soler, C., & Holzschuch, N. (2012). Accurate fitting of measured reflectances using a shifted gamma micro-facet distribution. *Computer Graphics Forum*, *31*(4), 1509–1518. [Article]

Barrow, H. G., & Tenenbaum, J. M. (1978). Recovering intrinsic scene characteristics from images. In A. Hanson & E. Riseman (Eds.), *Computer vision systems* (pp. 3–26). New York: Academic Press.

Beck, J., Prazdny, K., & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception & Psychophysics*, *35*(5), 407–422. [Article]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. [PubMed] [Article]

Church, E. L., Takacs, P. Z., & Leonard, T. A. (1989). The prediction of BRDFs from surface profile measurements. In J. C. Stover (Ed.), *Proceedings of SPIE: Vol. 1165. Scatter from optical components* (pp. 136–150). Bellingham, WA: SPIE. [Article]

Colbert, M., Pattanaik, S., & Krivanek, K. (2006). BRDF-shop: Creating physically correct bidirectional reflectance distribution functions. *IEEE Computer Graphics & Applications*, *26*(1), 30–36. [PubMed] [Article]

Cook, R. L., & Torrance, K. E. (1982). A reflectance model for computer graphics. *ACM Transactions on Graphics*, *1*(1), 7–24. [Article]

Debevec, P. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In M. F. Cohen (Ed.), *Proceedings of ACM SIGGRAPH 98* (pp. 189–198). New York, NY: ACM. [Article]

Doerschner, K., Boyaci, H., & Maloney, L. T. (2010). Estimating the glossiness transfer function induced by illumination change and testing its transitivity. *Journal of Vision*, *10*(4):8, 1–9, doi:10.1167/10.4.8. [PubMed] [Article]

Donner, C., & Jensen, H. W. (2005). Light diffusion in multi-layered translucent materials. *ACM Transactions on Graphics*, *24*(3), 1032–1039. [Article]

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, *3*(5):3, 347–368, doi:10.1167/3.5.3. [PubMed] [Article]

Freniere, E. R., Gregory, G. G., & Chase, R. C. (1997). Interactive software for optomechanical modeling. In R. E. Fischer, R. B. Johnson, R. C. Juergens, W. J. Smith, & P. R. Yoder, Jr. (Eds.), *Proceedings of SPIE, Vol. 3130. Lens design, illumination, and optomechanical modeling* (pp. 128–133). Bellingham, WA: SPIE. [Article]

Gilchrist, A. (1994). Absolute versus relative theories of lightness perception. In A. Gilchrist (Ed.), *Lightness, brightness, and transparency* (pp. 1–34). Hillsdale, NJ: Erlbaum.

Helmholtz, H. (1856). *Handbuch der physiologischen Optik* [*Handbook of physiological optics*]. Leipzig, Germany: Voss.

Ho, Y.-X., Landy, M. S., & Maloney, L. T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science*, *19*(2), 196–204. [Article]

Holzschuch, N., & Pacanowski, R. (2017). A two-scale microfacet reflectance model combining reflection and diffraction. *ACM Transactions on Graphics*, *36*(4): 66, 1–12.

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

Katz, D. (1935). *The world of colour* (Trans. from German by R. B. MacLeod & C. W. Fox). London, UK: Kegan Paul, Trench, Trübner.

Kim, J., Marlow, P. J., & Anderson, B. L. (2014). Texture-shading flow interactions and perceived reflectance. *Journal of Vision*, *14*(7):1, 1–19, doi:10.1167/14.7.1. [PubMed] [Article]

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(Suppl. 1), 1–16.

Koenderink, J. J., & van Doorn, A. J. (1980). Photometric invariants related to solid shape. *Optica Acta*, *27*(7), 981–996. [Article]

Krywonos, A., Harvey, J. E., & Choi, N. (2011). Linear systems formulation of scattering theory for rough surfaces with arbitrary incident and scattering angles. *Journal of the Optical Society of America A*, *28*(6), 1121–1138. [Article]

Kubelka, P., & Munk, F. (1931). Ein Beitrag zur Optik der Farbenstriche [An article on optics of paint layers] (S. H. Westin, Trans.). *Zeitschrift für technische Physik*, *12*(112), 593–601. [Article]

Křivánek, J., & Colbert, M. (2008). Real-time shading with filtered importance sampling. *Computer Graphics Forum*, *27*(4), 1147–1154. [Article]

Lafortune, E. P. F., Foo, S.-C., Torrance, K. E., & Greenberg, D. P. (1997). Non-linear approximation of reflectance functions. In G. S. Owen, B. Mones-Hattal, & T. Whitted (Eds.), *Proceedings of ACM SIGGRAPH 97* (pp. 117–126). New York: ACM. [Article]

Löw, J., Kronander, J., Ynnerman, A., & Unger, J. (2012). BRDF models for accurate and efficient rendering of glossy surfaces. *ACM Transactions on Graphics*, *31*(1):9, 1–14. [Article]

Marlow, P., Kim, J., & Anderson, B. (2012). The perception and misperception of specular surface reflectance. *Current Biology*, *22*(20), 1909–1913. [Article]

Matusik, W., Pfister, H., Brand, M., & McMillan, L. (2003). A data-driven reflectance model. *ACM Transactions on Graphics*, *22*(3), 759–769. [Article]

Metelli, F. (1974). Achromatic color conditions in the perception of transparency. In R. B. McLeod & H. L. Pick (Eds.), *Perception: Essays in honor of J. J. Gibson* (pp. 93–116). New York: Cornell University Press.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, *447*(7141), 206–209. [PubMed] [Article]

Ngan, A., Durand, F., & Matusik, W. (2005). Experimental analysis of BRDF models. In K. Bala & P. Dutré (Eds.), *Rendering Techniques '05: Proceedings of the 16th Eurographics Symposium on Rendering* (pp. 117–126). Natick, MA: AK Peters. [Article]

O'Donnell, F. X. D., & Billmeyer, F. W., Jr., (1986). Psychometric scaling of gloss. In J. J. Rennilson & W. N. Hale (Eds.), *Review and evaluation of appearance: Methods and techniques. ASTM STP 914* (pp. 14–32). West Conshohocken, PA: American Society for Testing and Materials.

Olkkonen, M., & Brainard, D. H. (2010). Perceived glossiness and lightness under real-world illumination. *Journal of Vision*, *10*(9):5, 1–19, doi:10.1167/10.9.5. [PubMed] [Article]

Pellacini, F., Ferwerda, J. A., & Greenberg, D. P. (2000). Toward a psychophysically-based light reflection model for image synthesis. In K. Akeley (Ed.), *Proceedings of ACM SIGGRAPH 2000* (pp. 55–64). New York: ACM. [Article]

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. [PubMed] [Article]

Perlin, K. (2002). Improving noise. *ACM Transactions on Graphics*, *21*(3), 681–682. [Article]

Rump, M., Müller, G., Sarlette, R., Koch, D., & Klein, R. (2008). Photo-realistic rendering of metallic car paint from image-based measurements. *Computer Graphics Forum*, *27*(2), 527–536. [Article]

Rusinkiewicz, S. M. (1998). A new change of variables for efficient BRDF representation. In G. Drettakis & N. Max (Eds.), *Rendering Techniques '98, Proceedings of the 9th Eurographics Workshop on Rendering* (pp. 11–22). Vienna, Austria: Springer. [Article]

Sato, H., Motoyoshi, I., & Sato, T. (2016). On–off asymmetry in the perception of blur. *Vision Research*, *120*, 5–10. [Article]

te Pas, S. F., & Pont, S. C. (2005). A comparison of material and illumination discrimination performance for real rough, real smooth, and computer generated smooth spheres. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization* (pp. 75–81). New York: ACM. [Article]

Upstill, S. (1990). *The RenderMan companion: A programmer's guide to realistic computer graphics*. Boston, MA: Addison-Wesley Professional.

Vangorp, P., Laurijssen, J., & Dutré, P. (2007). The influence of shape on the perception of material reflectance. *ACM Transactions on Graphics*, *26*(3): 77, 1–9. [Article]

Ward, G. J. (1992). Measuring and modeling anisotropic reflection. *Computer Graphics*, *26*(2), 265–272. [Article]

Wills, J., Agarwal, S., Kriegman, D., & Belongie, S. (2009). Toward a perceptual space for gloss. *ACM Transactions on Graphics*, *28*(4):103, 1–15. [Article]