

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/220183894>

The influence of shape on the perception of material reflectance

Article in ACM Transactions on Graphics · July 2007

DOI: 10.1145/1275808.1276473 · Source: DBLP

CITATIONS

87

READS

182

3 authors, including:



Peter Vangorp
Edge Hill University

19 PUBLICATIONS 261 CITATIONS

SEE PROFILE



Jurgen Laurijssen
KU Leuven

4 PUBLICATIONS 101 CITATIONS

SEE PROFILE

The Influence of Shape on the Perception of Material Reflectance

Peter Vangorp*

Jurgen Laurijssen†

Philip Dutré‡

Department of Computer Science
Katholieke Universiteit Leuven

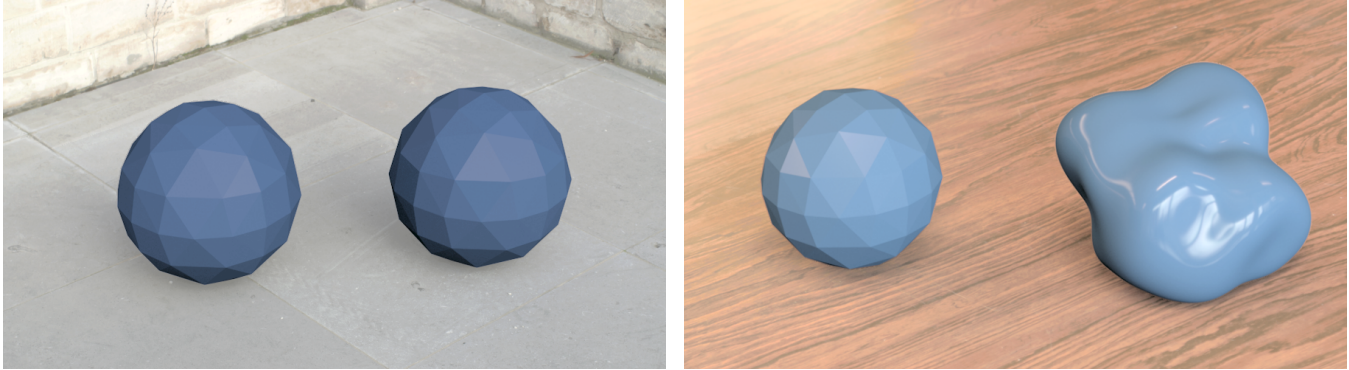


Figure 1: The tessellated spheres in the left image are rendered with two different types of a blue plastic BRDF, yet they are perceived as made from the same material. The objects in the right image are rendered with an identical blue plastic BRDF, yet their appearance is very different.

Abstract

Visual observation is our principal source of information in determining the nature of objects, including shape, material or roughness. The physiological and cognitive processes that resolve visual input into an estimate of the material of an object are influenced by the illumination and the shape of the object. This affects our ability to select materials by observing them on a point-lit sphere, as is common in current 3D modeling applications.

In this paper we present an exploratory psychophysical experiment to study various influences on material discrimination in a realistic setting. The resulting data set is analyzed using a wide range of statistical techniques. Analysis of variance is used to estimate the magnitude of the influence of geometry, and fitted psychometric functions produce significantly diverse material discrimination thresholds across different shapes and materials.

Suggested improvements to traditional material pickers include direct visualization on the target object, environment illumination, and the use of discrimination thresholds as a step size for parameter adjustments.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture; J.4 [Social and Behavioral Sciences]: Psychology

Keywords: visual perception, psychophysics, shading, material editing, geometry

1 Introduction

When people are confronted with a new object for the first time, the initial impression of its nature is primarily estimated from visual input. Visual material perception is one of the most significant clues for determining properties of an object, such as what material it is made of, how smooth it would be to the touch, or how heavy it could be. These observations are based on lifelong experience, built up from childhood by actually confirming visual impressions with other senses. When confidence in the visual input grows, material perception becomes an increasingly cognitive process [Palmer 1975]. In this paper we present an exploratory study of the effects of shape on the perception of materials.

Just as every image editing program needs a color picker, every 3D modeling application depends on a good material picker. Indeed, realistic image synthesis not only depends on accurate light transport simulations, but also on the ability to model virtual scenes accurately. One of the challenging tasks in 3D modeling is assigning an appropriate material description to every object in the scene. Until now, most material pickers offer a library of predefined materials, combined with basic controls to adjust the parameters of an underlying model. However, most popular material models rely on non-intuitive, non-linear and interdependent parameters that require a lot of experience and knowledge about the underlying model in order to reach some desired effect. Section 2 includes a discussion of recent advances in perceptually uniform reparameterizations that could solve most of these problems.

Visual feedback from many current material pickers is limited to a point-lit sphere, mostly for practical reasons, such as ease of rendering or identifiable highlights. However, a sphere is not necessarily

*e-mail: peter.vangorp@cs.kuleuven.be

†e-mail: jurgen.laurijssen@cs.kuleuven.be

‡e-mail: philip.dutre@cs.kuleuven.be

the optimal shape from a perceptual point of view. We will demonstrate that the perception of materials is strongly influenced by the shape of the object, and that a sphere is not always a good choice to pre-visualize material selections.

In general, a material model defines the appearance of a surface, including spatially varying texture, angular variation, and subsurface scattering properties. We will limit the scope of this paper to spatially uniform surface properties, excluding subsurface effects. These are best described by the bidirectional reflectance-distribution function (BRDF) [Nicomemus et al. 1977], and capture effects such as glossiness and diffuse color.

2 Related Work

Visual constancy. Early work in visual perception has focused mainly on the physiological and neurological characteristics of the human visual system related to contrast and color. Examples include research into different aspects of *color constancy*: the color and lightness of an object appear remarkably constant under substantial changes in illumination (see [Brainard 2004] for a recent overview). Although limits of color constancy have been discovered [Foster 2003; Brainard and Wandell 1991], it remains one of the strongest factors in the visual perception of materials. Similar invariance of perceived glossiness under changing illumination, termed *gloss constancy*, was demonstrated by Obein et al. [2004]. While it is often assumed that color constancy under changing surface glossiness also holds [ASTM 1999; Aida 1997], this is certainly not always the case. Xiao and Brainard [2006] demonstrated that color appearance is indeed affected slightly by glossiness.

Material perception. Most of the work in this area has examined materials presented on a single shape, usually a sphere. This choice is partly justified because a sphere presents all possible surface orientations towards the viewer, without preference. It can therefore give a good idea of the full range of the reflectance function. The choice is also motivated by the sphere’s ease of use in modeling and interactive rendering. Its convexity eliminates the need for self-shadowing and interreflection.

A notable exception is the work of Nishida and Shinya [1998]. They discovered that observers had great difficulty matching Phong-shaded height fields of different amplitudes and frequencies, thus showing that gloss constancy is not guaranteed when varying the geometry. They also relate observers’ performance to information in the luminance histogram of the stimulus images.

The reflectance matching experiments of Fleming et al. [2003] indicate that people can judge material characteristics more accurately under natural environment illumination than under artificial point light sources. It is therefore probable that the human visual system uses implicit knowledge of the statistics of real-world illumination to aid in reflectance perception.

Perceptual material parameterization. Pellacini et al. [2000] examined the perceptual space of glossy materials representable by the isotropic Ward reflectance model [Ward 1992]. They showed that there are 2 perceptual gloss dimensions present in this space, namely *contrast gloss* and *distinctness-of-image gloss*. These dimensions are used as the basis for a perceptually uniform reparameterization of the Ward reflectance model. Additionally, in Ferwerda et al. [2001] the *thresholds* for perceiving differences in the perceptual gloss space are derived.

Ngan et al. [2006] suggested an L^2 metric on images of spheres rendered with different BRDFs under environment illumination, as an alternative model for the perception of distinctness-of-image gloss. The use of image-based metrics can certainly be justified

from a perceptual perspective: people perceive a material by looking at (images of) objects made out of the material, rather than consciously considering a mental model of an abstract 4D reflectance function.

The ASTM gloss appearance standards [ASTM 1999] provide widely-used definitions and measurement devices for different kinds of gloss. Westlund and Meyer [2001] simulated these measurement devices in a *virtual light meter* to derive the specular and aspecular gloss (or haze) scales corresponding to the roughness parameter range of several analytical reflectance models.

Matusik et al. [2003b] sampled the space of all isotropic surface materials by measuring over 100 BRDFs. They found that this space can be spanned by a 15-dimensional non-linear manifold. The dimensions correspond to intuitive traits, and interpolation on the manifold is guaranteed to synthesize novel physically plausible materials [Matusik et al. 2003a].

Material editing. Colbert et al. [2006] present a BRDF editor based on a multi-lobe extension of the Ward reflectance model. The editor allows an artist to paint and manipulate highlights on a point-lit sphere, chosen to show clearly distinct highlights. The resulting material is at the same time rendered on a more complex model under unshadowed environment map illumination to see if the material gives the desired result on the target shape.

The real-time BRDF editor of Ben-Artzi et al. [2006] allows the user to visualize arbitrary reflectance functions directly on the target shape, under arbitrary illumination. The viewpoint-dependent precomputation of direct lighting produces high-quality shadows and gloss. Intuitive controls are offered to reshape the curves of different reflectance functions.

The innovative BRDF navigation technique by Ngan et al. [2006] uses pre-rendered images of spheres under environment map illumination. They show the currently selected BRDF together with several variations at a specified perceptually uniform step size. This allows the user to navigate along a number of intuitive dimensions, even across different underlying analytical models. The choice of the reflectance model is made for the user, based on the desired look of the material.

3 Overview

The topic of this paper is the study of the perception of materials under different conditions. The main point of interest can be described as the study of material constancy under changing geometry. In Section 4 we present an exploratory psychophysical experiment to investigate various influences on material discrimination. Participants were shown a series of images, displaying two virtual objects in a realistic setting, and were asked whether the objects looked as if they were made of the same material. The resulting data set is analyzed using a wide range of statistical techniques in Section 5. Conclusions are drawn about meaningful characteristics like the magnitude of the influence of geometry, and material discrimination thresholds across different shapes and materials.

4 Perceptual Experiment

4.1 Stimulus images

Each stimulus image is a combination of 4 components:

1. The *scene* provides a background and illumination environment, in which objects of interest are placed.

2. Two objects of identical or different *shapes* are placed in the scene, e.g. a sphere and a bunny.
3. At least one of the objects is rendered with an original *base material*, e.g. blue plastic.
4. The second object is rendered either with the same original base material or a small *variation* thereof.

Fleming et al. [2003] suggested using natural environment illumination to improve material discrimination. In addition, we decided to place objects in a real-world scene instead of rendering them in front of the blurry environment map used as their background. Therefore, standard augmented reality techniques [Debevec 1998] were used to achieve the required level of realism. These techniques combine a background photograph of a scene and a foreground rendering of a local scene with virtual objects. The virtual objects are rendered under environment illumination, a distant light source approximation of the actual lighting conditions in the photographed scene. A standard path tracing implementation was used to render full global illumination effects, including soft shadows and interreflections [Dutr e et al. 2006]. The complete capturing, rendering and compositing pipeline was implemented using high dynamic range values and images. The final images were tone-mapped using the histogram adjustment method described by Ward Larson et al. [1997]. This mix of image-based rendering and traditional global illumination algorithms is currently the most feasible way to render large sets of realistic-looking images. Moreover, people would be able to tell the difference between a real scene and the clean and sharp look of a pristine model [Longhurst et al. 2003].

Our decision to show two objects in a single stimulus image was motivated by the following reasons:

1. Presenting one image containing both objects makes better use of the center of gaze than two separate images, each showing a single object. The objects can be close together in the middle of the image while the surrounding area can provide some context.
2. We also expect that combining both objects in a single image will improve participant performance and experiment validity by emulating a real-world scenario of material comparisons.
3. Finally, although the viewing direction might differ slightly between the objects, this only becomes apparent when two identical shapes are presented. The illumination is exactly the same for every object in the scene because of the distant illumination approximation, except for subtle interreflections and shadowing between objects.

To fully utilize participants' experience with real-world materials, we chose to use the database of real isotropic materials measured by Matusik et al. [2003b]. Ngan et al. [2005] provide the parameters for a number of analytical reflectance models that best fit the measured data, including the isotropic Ward BRDF model [Ward 1992]:

$$f_r = \frac{\rho_d}{\pi} + \frac{\rho_s}{4\pi\alpha^2\sqrt{\cos\theta_i\cos\theta_o}} \exp\left(-\frac{\tan^2\theta_h}{\alpha^2}\right)$$

where ρ_d is the diffuse reflectance, ρ_s is the specular reflectance, α is a surface roughness parameter, and θ_i , θ_o , and θ_h are the elevation angles of the incident light direction, the outgoing light direction, and the halfway vector, respectively.

The main reason for choosing the Ward model over tabulated measurement data or other analytical models is because a perceptually meaningful reparameterization exists [Pellacini et al. 2000]. It can be used to generate perceptually uniform gloss variations

on the base materials. The dimensions of contrast gloss (c) and distinctness-of-image gloss (d) are specified as independent transformations of the parameters in the Ward model:

$$c = \sqrt[3]{\rho_s + \rho_d/2} - \sqrt[3]{\rho_d/2} \quad (1)$$

$$d = 1.78(1 - \alpha) \quad (2)$$

We chose to pick variations that deviate from a base material along only one of these dimensions at a time (Δc or Δd), so we do not rely on the general Euclidean distance metric in $\{c, d\}$ -space.¹ A deviation along either one of the uniform dimensions is denoted by $\Delta\{c, d\}$. For the contrast gloss equation, the diffuse reflectance of the base material is kept constant, because even small differences would be detected immediately. Of the specular reflectance only the lightness L is changed, because gloss chromaticity is an important clue to detect metallic materials. In general, glossy non-metals produce white highlights, while the highlight color of metals is related to their diffuse color.

The Ward model works well for glossy materials but it should not be used with extreme parameter values. Any change in roughness α will be lost when specular reflectance ρ_s approaches zero. In terms of the perceptual reparameterization, if there is not enough contrast gloss to form noticeable highlights, a distinct image certainly cannot be reflected. At the other extreme, any change in specular reflectance ρ_s will lose its effect on gloss when roughness α exceeds a certain level. In other words, when the reflected image becomes too indistinct, the specular lobe acts as just another nearly-diffuse term.

Our perceptual experiment is intended to explore the interesting ranges for each variable factor in the stimuli. We selected a diverse set for each category.

Scenes: *attic, courtyard, library, tabletop*

Illumination conditions ranged from a sunny outdoor courtyard scene to an attic scene illuminated from a small window on a cloudy day. Different camera positions were used, including varying elevations of the camera above the ground plane and varying fields of view (10° – 29° vertical).

Shapes: *blob, buddha, bunny, car, cylinder, dragon, rounded icosahedron, sphere, teapot, tessellated sphere, triceratops*

A wide variety of shapes was selected based on many different characteristics that might prove interesting, including complexity, convexity, curvature, anisotropy of curvature, and presence of flat surfaces. The rounded icosahedron has subtle beveled edges.

Base materials: *acrylicblue, aluminium, chballgoldmetallic2, copper, metallicblue, metallicsilver, nickel, nylon, paintlightred, pearlpaint*

A set of base materials was chosen from the database of Matusik et al. [2003b] and Ngan et al. [2005], well-distributed within the working range of the Ward reflectance model and the $\{c, d\}$ -space. The selection includes metals, plastics and paints with varying colors and gloss.

Variations: *original*, $\Delta\{c, d\} = 0.05, 0.10, 0.15, 0.20, 0.25$

Visual inspection of a small set of test images showed that the threshold for noticing differences in materials was distributed around $\Delta\{c, d\} = 0.10$. In contrast, Ferwerda et al. [2001] found thresholds as small as $\Delta\{c, d\} = 0.03$. In Section 5 the cause of this apparent difference is explained. Variations were chosen such that materials remained within the working range of the reflectance model.

A few of the resulting stimulus images are shown in Figure 2.

¹Pellacini et al. [2000] introduced the scale factor 1.78 in the distance metric. We include it in Equation 2 instead, yielding equivalent relations.

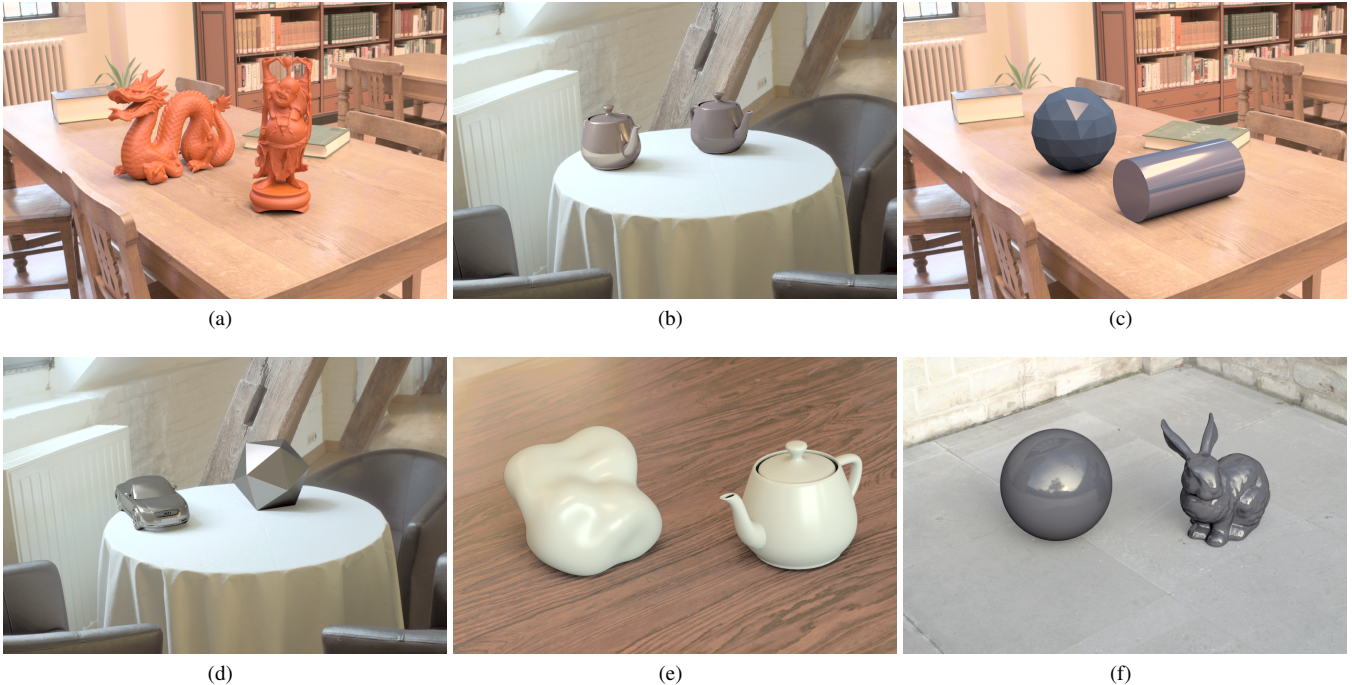


Figure 2: A selection of 6 example images, out of a total of 330 stimuli. (a) Library scene, dragon and buddha, both original light red paint. (b) Attic scene, two copper teapots (which are slightly rescaled). The teapot on the right is rendered with less contrast gloss ($\Delta c = 0.20$). (c) Library scene. The tessellated sphere is rendered with the original acrylic blue material and the cylinder is rendered with more contrast gloss ($\Delta c = 0.25$). (d) Attic scene, car and rounded icosahedron, both nickel without variations. (e) Tabletop scene, blob and teapot, both nylon without variations. (f) Courtyard scene, sphere and bunny, both aluminium without variations.

4.2 Procedure

Because of the large number of possible combinations, we used a randomized experimental design, sampling the factorial space of combinations in a well-distributed manner. Our complete set of 330 stimulus images can be partitioned into 4 groups (see Table 1).

	shape	material	number of images
Group 1	identical	identical	55
Group 2	different	identical	165
Group 3	identical	different	55
Group 4	different	different	55

Table 1: Groups of stimuli in the experiment.

More images were created for group 2 (different shape, identical materials) to obtain a better sampling of this space. This tests the perception of identical materials on different geometry, which is of most interest to us.

Participants were given written instructions before the experiment. The question asked for each stimulus image was: “Are both objects in the image made out of the same material?” This simple yes-or-no question avoids too much variability between participants. The instructions contained a few example images as a short training session to clarify what is meant by the same material, e.g. some people might call aluminium and nickel both metals, even though they see an obvious difference between both materials. Also smooth and brushed finishes should be judged as being different.

Each session consisted of 220 images: the complete groups 1, 3, and 4, and a well-distributed sampling of 55 images from group 2.

Keeping the sessions balanced between all conditions was intended to avoid bias, e.g. if a participant always selects “identical materials” as their default answer when they cannot decide with confidence, then a significantly larger proportion of identical material images in the session will artificially inflate their score.

The stimulus images were presented one at a time, in random order, on a calibrated Iiyama Vision Master Pro 19” CRT monitor [Brainard et al. 2002]. The experiment was performed in controlled office lighting conditions, without influence from daylight. Each image filled approximately half the screen area over a neutral gray background. The screen was blanked to gray for a fraction of a second between images.

Because of the large number of stimuli presented to each participant, lapses in concentration might occur. The instructions explicitly stated that participants were allowed to take a short break if they grew tired. Most participants completed a session of 220 images in under 30 minutes.

A total number of sixteen participants took part in the experiment: 6 computer graphics researchers and 10 persons who had no particular knowledge of computer graphics techniques. Analysis showed that there was no difference between these two groups (see Section 5). All participants were unpaid volunteers and had normal or corrected-to-normal visual acuity and normal color vision.

5 Results and Analysis

There are two ways in which to interpret each single observation:

Assessment: Participants indicate whether they perceive the 2 materials shown in the image as either *identical* or *different*. For

quantitative analysis, we define assessment as the percentage of “identical materials” answers.

Accuracy: Since we know which materials were used in each image, we can derive the accuracy of their assessment as either *right* or *wrong*. In other words, whether they answered the question “Are both objects made out of the same material?” correctly or not.

When a large number of factors is needed for a thorough exploratory experiment, different subsets of the resulting data are often suitable for many different data analysis techniques. In the following sections, we summarize the major trends and examine the material discrimination threshold.

No significant differences between participants: Participants’ ages ranged between 22–56, with 10 out of 16 being in their twenties. Different parts of the population with different levels of experience were represented: graphics researchers, people with various computer science backgrounds, and casual computer users. No differences were found between age groups, experience, gender or race.

In informal debriefing after the session, it became apparent that people without any experience in graphics lack the vocabulary to describe materials. In some cases, they could not describe nor even point at the differences they perceived.

5.1 Major trends in the complete data set

In this section, various hypotheses are tested with the Kruskal-Wallis analysis of variance [Kruskal and Wallis 1952]. This non-parametric test is an alternative for the well-known classical analysis of variance (ANOVA). Similarly, it verifies whether different subsets of the data have the same probability distribution, e.g. a statistic $p < 0.05$ denotes 95% certainty of a significant difference between the subsets. Unlike classical ANOVA, Kruskal-Wallis does not make the assumption of normally distributed data.

Difference in geometry diminishes accuracy: Participants’ average accuracy is rather high (79%) when observing materials on identical geometry (groups 1 and 3), but it drops significantly ($p < 0.01$) to only 62% on different geometry (groups 2 and 4). See Table 2 for a summary of these results.

group	shape	#images per person	accuracy	assessment identical
1	identical	55	87%	87%
2	different	55	71%	71%
3	identical	55	70%	30%
4	different	55	53%	47%
1 and 3	identical	110	79%	59%
2 and 4	different	110	62%	59%

Table 2: Average accuracy and assessment percentages, measured over each group of stimulus images.

However, participants do not seem to adjust their decision criterion between these subsets, because their average assessment remains constant at 59% identical. This percentage can be explained intuitively by the groups of images in each session. Exactly half of the presented images contain identical materials, and approximately 10% contain a variation of $\Delta\{c, d\} = 0.05$, which is rarely noticed. In Section 5.2 the material discrimination threshold is analyzed further.

Relative magnitude of influence from geometry and material variations: Figure 3 shows a quantitative comparison between

the assessment and accuracy when only shapes differ (group 2) and when only material variations differ (group 3 partitioned per $\Delta\{c, d\}$). Reference levels are shown for identical shapes and identical materials (group 1). Figure 3(a) illustrates that, as can be expected, the indirect influence of geometry on the assessment is smaller than the influence of direct material variations, which are the target of the experimental question. This difference is significant for all variations except $\Delta\{c, d\} = 0.05$ ($p < 0.01$). In Figure 3(b), the influence of geometry crosses the level of material variations at $\Delta\{c, d\} = 0.12$, with a 95% confidence interval of $[0.04, 0.20]$. Intuitively, this means that material picking on an unsuited shape can introduce errors equivalent to a difference of $\Delta\{c, d\} = 0.12$ on the BRDF, which is a noticeable variation (see Section 5.2).

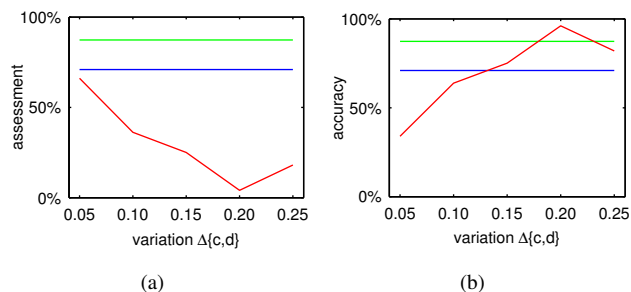


Figure 3: Comparisons of assessment and accuracy. The green line is the reference value (identical shapes and materials, group 1). The blue line has a lower value because of the influence of different geometry (group 2). The red line is the influence of material variations $\Delta\{c, d\}$. (a) Geometry always has a smaller influence on assessment than material variations. (b) The influence of different geometry on accuracy is approximately the same as the influence of material variations $\Delta\{c, d\} = 0.12$.

Tessellated geometry diminishes accuracy: Another strong effect is participants’ inability to assess the appearance of a badly tessellated mesh without shading normals. Accuracy for comparing two tessellated spheres or icosahedra (despite their rounded edges) is 69%, which is significantly lower than for other shapes (81%, $p < 0.01$). Some participants appear to fall back to a default assessment of identical materials in this case. The percentage identical assessment for these shapes is higher than average (72% vs. 56%, $p < 0.01$). When comparing tessellated shapes to different geometries this effect is less pronounced.

The well-known fact that curvature captures highlights was mathematically confirmed and quantified by Durand et al. [2005] and Ramamoorthi et al. [2007]. The facets of the tessellated sphere are uncurved, so they fail to capture any recognizable highlight. Instead, the facets look like different diffuse colors.

Attic scene influences assessment and accuracy: In the attic scene, participants’ accuracy drops from 72% to 66% ($p < 0.01$), and identical assessment increases from 56% to 66% ($p < 0.05$). Visual inspection of the images shows that the left object appears to receive more light from the nearby window at the top left. However, the objects are in fact rendered with exactly the same environment illumination. The only actual differences occur when the left object blocks the light and casts a faint shadow onto the right object, and when the dark side of the left object is reflected in the light side of the right object, creating contrast.

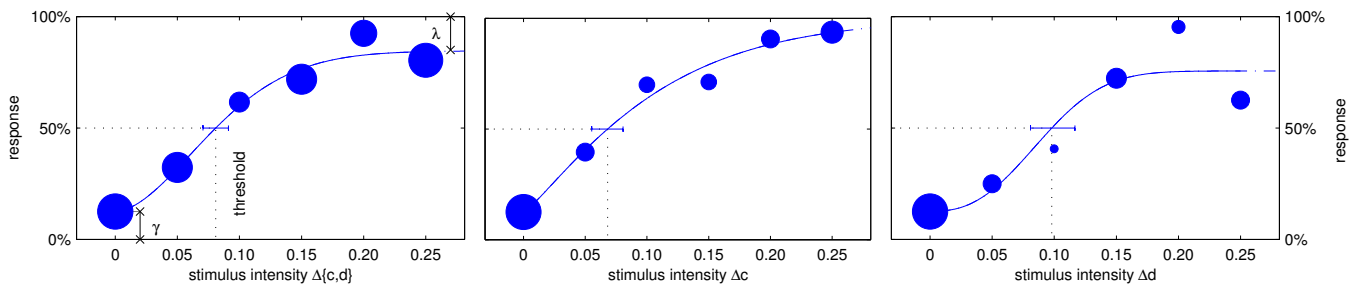


Figure 4: Psychometric functions for the entire data set of images, for material variations $\Delta\{c,d\}$, Δc and Δd respectively. Only images showing identical shapes were considered, in order to eliminate any influence of geometry on the perception of material. The locations of γ , λ , and the threshold are indicated on the first graph.

5.2 Material discrimination threshold

Data from a psychophysical experiment relates observers' responses to a physical variable of the stimuli. A psychometric function can be fitted to psychophysical data, so that useful characteristics of the underlying phenomenon can be derived. The general form of the psychometric function is [Wichmann and Hill 2001a,b]:

$$\psi(x; \alpha, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda) F(x; \alpha, \beta)$$

where x is the stimulus intensity, in this case the distance $\Delta\{c,d\}$. The *guess rate* γ is the response at zero stimulus intensity, and the *miss rate* λ indicates how often observers fail to notice a large stimulus intensity. The sigmoid function F can often be derived as the cumulative distribution function of the underlying experimental process that generates the data. We use the Weibull distribution:

$$F(x; \alpha, \beta) = 1 - \exp\left(-\left(\frac{x}{\alpha}\right)^\beta\right), \quad x \geq 0$$

This distribution is often used to model phenomena that increase in probability with increasing stimulus intensity [Weibull 1951]. One can easily see that material discrimination is related to material difference in a similar way. The Weibull distribution is indeed widely used to model discrimination and detection experiments.

In traditional, more focused experiments, only a single variable determines the stimulus intensity and thereby the response, and γ and λ are merely intended as small corrections for random guesses and lapses in concentration, with typical values up to 5%. Our fitted curves often have larger values of γ and λ , because the perceptual difference of materials is not the only variable in our stimulus images. Like in any real-world scenario, many small factors, like the surrounding scene or the specific viewpoint, can influence participants' perception significantly.

The *just noticeable difference* or *threshold*, where observers start to see the difference in the variations, is defined as:

$$\psi_p^{-1} = \alpha \beta \sqrt{\ln\left(\frac{1 - \gamma - \lambda}{1 - p - \lambda}\right)}$$

where p is the desired percentage (usually 50%) of observers that notice the difference. The abruptness of the threshold is indicated by the slope ψ_p' at that point. Note that we deviate from the common practice of using γ and λ only to improve fitting, and ignoring these small values afterwards by defining the threshold as F_p^{-1} instead, which is independent of γ and λ .

The optimal shape for discrimination tasks would have low guess and miss rates, an early threshold for discrimination of the smallest

differences, and a steep slope to minimize the dependence of the threshold on the desired response level. The optimal shape for material picking should only be as discriminative as the target shape.

The *bootstrap method*, a Monte Carlo resampling technique introduced by Efron [1979], was used to estimate the variability and confidence intervals of these statistics.

Main result: Figure 4 shows the resulting psychometric functions for the complete dataset of identical shapes (groups 1 and 3), as well as split into separate graphs for Δc and Δd variations. Sample points are indicated by dots. The size of each dot represents the number of observations for that sample. The average material discrimination threshold over all identical shapes and materials is $\Delta\{c,d\} = 0.08$. The horizontal error bar indicates the 68% confidence interval (one standard deviation) for the threshold. The separate thresholds $\Delta c = 0.07$ and $\Delta d = 0.10$ do not differ significantly from each other nor from the average. However, our thresholds are larger than those found by Ferwerda et al. [2001], likely because of the influence of geometry and color.

In Equation 2, the weighting factor of 1.78 is intended to create a perceptually uniform $\{c,d\}$ -space with a Euclidean distance metric. Because only grayscale images were used in the experiment of Pellacini et al. [2000], the reflectances $\rho_{\{d,s\}}$ in Equation 1 denote only lightness (L), not a color. The apparent glossiness is assumed to be independent of the surface chromaticity (a,b). However, our results show that the psychometric functions for different base materials do not coincide exactly, suggesting that diffuse or specular chromaticity does affect perceived gloss.

Base materials: Figure 6 shows the psychometric function for the $\Delta\{c,d\}$ -variations of the materials. Significant differences in threshold and slope are visible. Aluminium, the material that is most similar visually and numerically to the materials of Ferwerda et al. [2001], confirms an early threshold of $\psi_{50\%}^{-1} = 0.04$ or $\psi_{75\%}^{-1} = 0.06$. This is not significantly different from their reported 75%-threshold of 0.03. Figure 5 shows the different variations of aluminium spheres.

Shapes: Figure 7 shows the psychometric function for each shape. There are clearly significant differences between them. The sphere has a very small guess rate γ and miss rate λ , but it also has the latest threshold $\Delta\{c,d\} = 0.16$, meaning it is not very well-suited for material discrimination tasks. The teapot has similar γ and λ , and an earlier threshold $\Delta\{c,d\} = 0.11$, but the gradual slope means a higher desired response level (more people should perceive the difference) requires a much later threshold. A single optimal shape does not exist in our selection, although the blob comes closest. It has an early threshold $\Delta\{c,d\} = 0.03$, a steep slope, and a

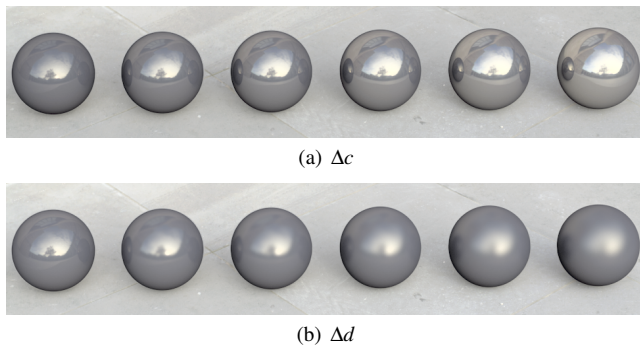


Figure 5: All variations of aluminium spheres, starting from the original material on the left and with $\Delta\{c,d\} = 0.05, 0.10, 0.15, 0.20, 0.25$ on the right. Compare each sphere to the leftmost one to identify the discrimination threshold visually.

small miss rate λ . The worst shape is the tessellated sphere, as it levels off at 50%, meaning participants performed no better than chance. The cylinder and the dragon have fits and thresholds of limited reliability because of insufficient data.

6 Conclusion

In this paper, we described an experiment in which participants had to judge whether two virtual objects, rendered and illuminated in a natural environment, were made out of the same material. The stimulus images covered a wide range of shapes, materials and illumination environments. Although our experiment is limited, we feel that this type of study can provide valuable insights into the relationship between the perception of shape and materials.

Statistical analysis shows that the accuracy of material perception is influenced by the geometrical shape of the object rendered with that particular material model. Resulting psychometric curves differ for different shapes, but also for different types of materials.

Our results suggest that more thought should be given to the design of material selection tools in modeling applications. The accepted practice of using a sphere as a base model to visualize the selected material parameters does not necessarily convey the appearance of the selected material on the target shape correctly. We found that a sphere is one of the least discriminating shapes w.r.t. the accuracy of judging materials, thus suggesting that other shapes, possibly dependent on the class of material or similar to the target shape, might be better suited for this task.

Photorealistic rendering could be another future application area. The allowable error threshold in images could be derived from the psychometric curves for materials or objects. As long as the errors remain lower than the discrimination threshold, no noticeable change in the image will be detected at the cognitive level.

For the applications sketched above, optimally, one would like to have psychometric curves available for each separate shape-material combination. Our current experiment did not contain enough images to derive such curves. However, we feel that future experiments, focusing on the influence of a single effect or particular combination would be very interesting to investigate. Future areas for testing could include the influence of texture maps or transparency, as well as more focused studies for finding better classifications of the shapes of objects.

Acknowledgements

The authors thank all the people who volunteered to participate in our experiment. We would also like to acknowledge Wojciech Matusik and Addy Ngan for providing the measured and fitted real-world materials, the Stanford 3D Scanning Repository for providing the buddha, bunny, and dragon models, RNA Studios for the car model, and Martin Newell for the teapot model.

References

- AIDA, T. 1997. Glossiness of colored papers and its application to specular glossiness measuring instruments. *Systems and Computers in Japan* 28, 1, 95–112.
- ASTM. 1999. *Annual Book of ASTM Standards, volume 06.01*. American Society for Testing and Materials.
- BEN-ARTZI, A., OVERBECK, R., AND RAMAMOORTHY, R. 2006. Real-time BRDF editing in complex lighting. *ACM Transactions on Graphics* 25, 3, 945–954.
- BRAINARD, D. H., AND WANDELL, B. A. 1991. A bilinear model of the illuminant's effect on color appearance. In *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, Eds. MIT Press, 171–186.
- BRAINARD, D. H., PELLI, D. G., AND ROBSON, T. 2002. Display characterization. In *Encyclopedia of Imaging Science and Technology*, J. P. Hornak, Ed. Wiley, New York, 172–188.
- BRAINARD, D. H. 2004. Color constancy. In *The Visual Neurosciences*, L. M. Chalupa and J. S. Werner, Eds. MIT Press, 948–961.
- COLBERT, M., PATTANAİK, S., AND KRIVANEK, J. 2006. BRDF-Shop: Creating physically correct bidirectional reflectance distribution functions. *IEEE Computer Graphics and Applications* 26, 1, 30–36.
- DEBEVEC, P. E. 1998. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of ACM SIGGRAPH 98*, ACM Press / ACM SIGGRAPH, New York, M. F. Cohen, Ed., Computer Graphics Proceedings, Annual Conference Series, ACM, 189–198.
- DURAND, F., HOLZSCHUCH, N., SOLER, C., CHAN, E., AND SILLION, F. 2005. A frequency analysis of light transport. *ACM Transactions on Graphics* 24, 3, 1115–1126.
- DUTRÉ, P., BALA, K., AND BEKAERT, P. 2006. *Advanced Global Illumination*, 2nd ed. A K Peters, Natick, MA.
- EFRON, B. 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7, 1, 1–26.
- FERWERDA, J. A., PELLACINI, F., AND GREENBERG, D. P. 2001. A psychophysically-based model of surface gloss perception. In *Proceedings of SPIE Human Vision and Electronic Imaging*, SPIE, 291–301.
- FLEMING, R. W., DROR, R. O., AND ADELSON, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision* 3, 5, 347–368.
- FOSTER, D. H. 2003. Does colour constancy exist? *Trends in Cognitive Sciences* 7, 10, 439–443.
- KRUSKAL, W. H., AND WALLIS, W. A. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* 47, 260, 583–621.

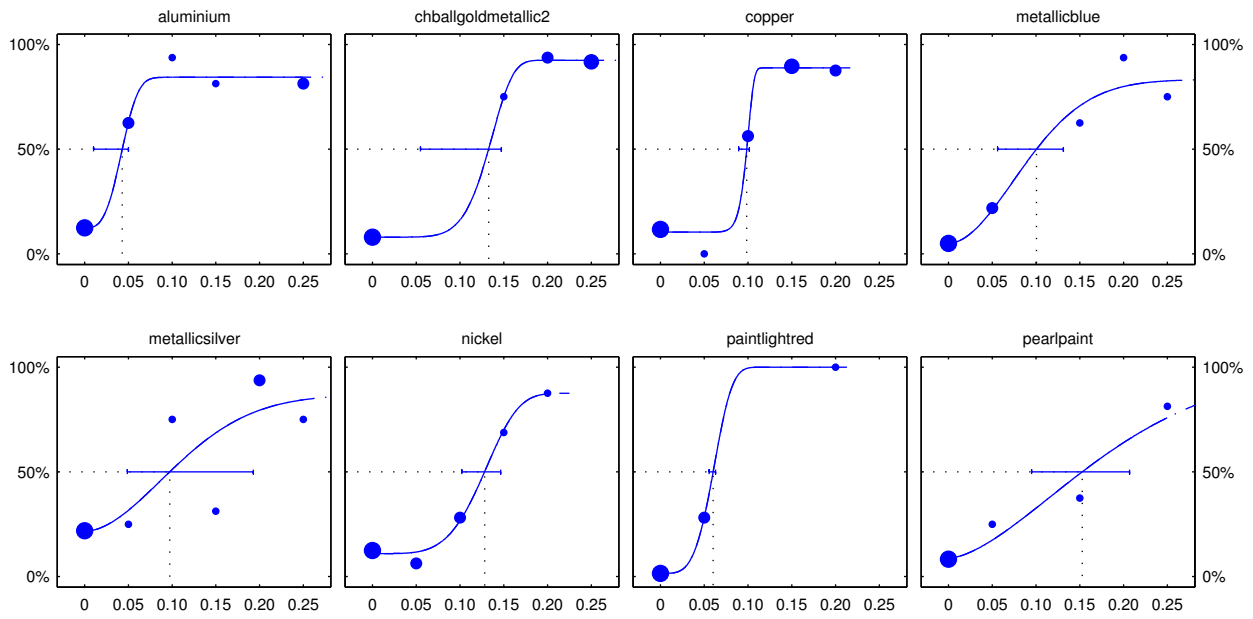


Figure 6: Psychometric functions for each material, combining $\Delta\{c, d\}$ (acrylicblue and nylon were omitted because insufficient data was available for a reliable fitting).

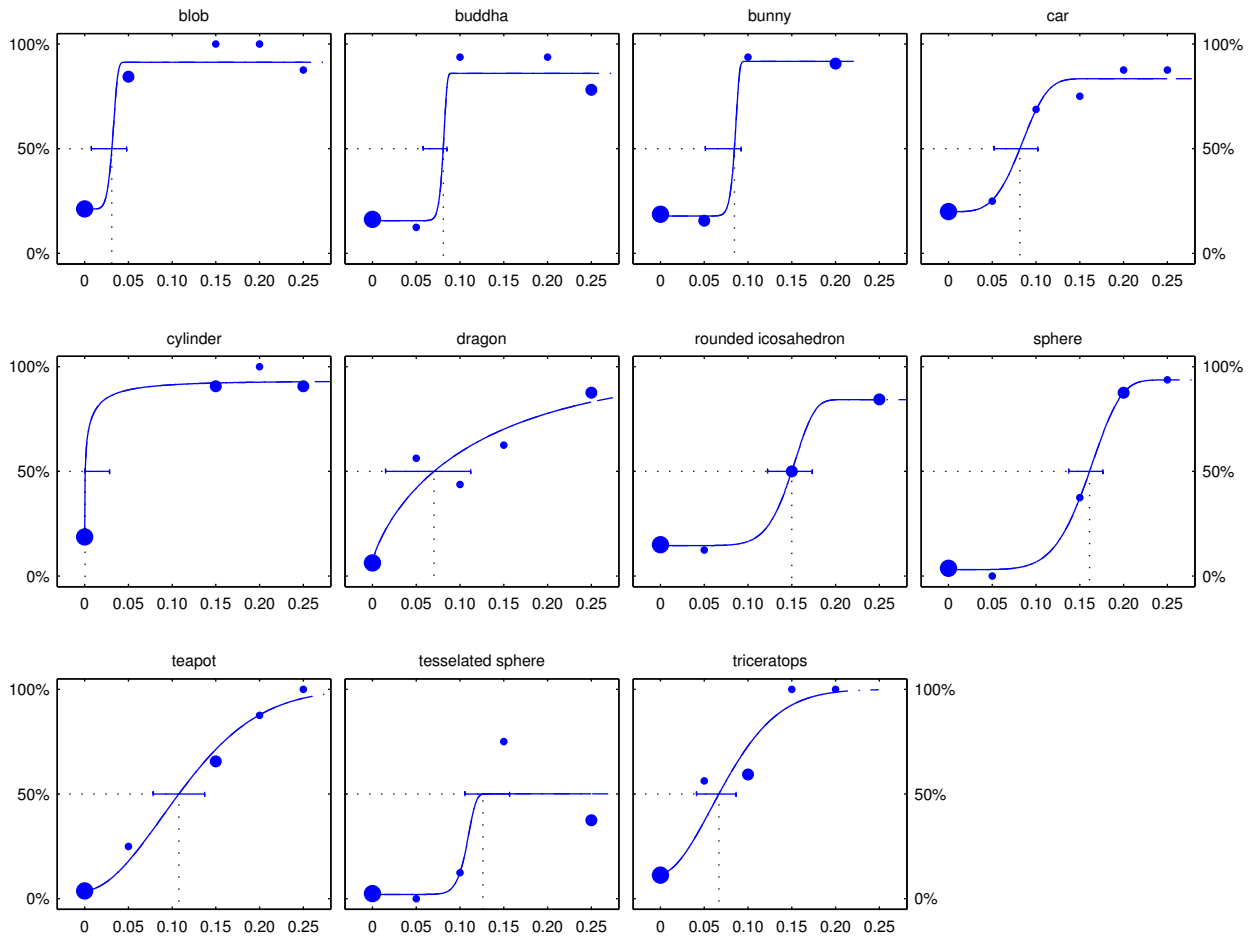


Figure 7: Psychometric functions for each shape, combining $\Delta\{c, d\}$. Only images of identical shapes were considered.

- LONGHURST, P., LEDDA, P., AND CHALMERS, A. 2003. Psychophysically based artistic techniques for increased perceived realism of virtual environments. In *Proceedings of Afrigraph*, Afrigraph, 123–132.
- MATUSIK, W., PFISTER, H., BRAND, M., AND MCMILLAN, L. 2003. A data-driven reflectance model. *ACM Transactions on Graphics* 22, 3, 759–769.
- MATUSIK, W., PFISTER, H., BRAND, M., AND MCMILLAN, L. 2003. Efficient isotropic BRDF measurement. In *Proceedings of the 14th Eurographics Symposium on Rendering*, P. Dutré, F. Suykens, P. H. Christensen, and D. Cohen-Or, Eds., Eurographics, 241–248.
- NGAN, A., DURAND, F., AND MATUSIK, W. 2005. Experimental analysis of BRDF models. In *Proceedings of the 16th Eurographics Symposium on Rendering 2005*, K. Bala and P. Dutré, Eds., Eurographics, 117–126.
- NGAN, A., DURAND, F., AND MATUSIK, W. 2006. Image-driven navigation of analytical BRDF models. In *Proceedings of the 17th Eurographics Symposium on Rendering*, T. Akenine-Möller and W. Heidrich, Eds., Eurographics, 399–407.
- NICODEMUS, F. E., RICHMOND, J. C., HSIA, J. J., GINSBERG, I. W., AND LIMPERIS, T. 1977. *Geometrical Considerations and Nomenclature for Reflectance*. Monograph 160, National Bureau of Standards.
- NISHIDA, S., AND SHINYA, M. 1998. Use of image-based information in judgments of surface-reflectance properties. *Journal of the Optical Society of America A: Optics, Image Science & Vision* 15, 12, 2951–2965.
- OBEIN, G., KNOBLAUCH, K., AND VIÉNOT, F. 2004. Difference scaling of gloss: Nonlinearity, binocularity, and constancy. *Journal of Vision* 4, 9, 711–720.
- PALMER, S. E. 1975. Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In *Explorations in cognition*, D. A. Norman and D. E. Rumelhart, Eds. Freeman, San Francisco, 279–307.
- PELLACINI, F., FERWERDA, J. A., AND GREENBERG, D. P. 2000. Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of ACM SIGGRAPH 2000*, ACM Press / ACM SIGGRAPH, New York, K. Akeley, Ed., Computer Graphics Proceedings, Annual Conference Series, ACM, 55–64.
- RAMAMOORTHI, R., MAHAJAN, D., AND BELHUMEUR, P. 2007. A first-order analysis of lighting, shading, and shadows. *ACM Transactions on Graphics* 26, 1.
- WARD LARSON, G., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics* 3, 4, 291–306.
- WARD, G. J. 1992. Measuring and modeling anisotropic reflection. *Computer Graphics (Proceedings of ACM SIGGRAPH 92)* 26, 2, 265–272.
- WEIBULL, W. 1951. A statistical distribution function of wide applicability. *Journal of Applied Mechanics* 18, 3, 293–297.
- WESTLUND, H. B., AND MEYER, G. W. 2001. Applying appearance standards to light reflection models. In *Proceedings of ACM SIGGRAPH 2001*, ACM Press / ACM SIGGRAPH, New York, E. Fiume, Ed., Computer Graphics Proceedings, Annual Conference Series, ACM, 501–510.
- WICHMANN, F. A., AND HILL, N. J. 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics* 63, 8, 1293–1313.
- WICHMANN, F. A., AND HILL, N. J. 2001. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics* 63, 8, 1314–1329.
- XIAO, B., AND BRAINARD, D. H. 2006. Color perception of 3D objects: Constancy with respect to variation of surface gloss. In *Symposium on Applied Perception in Graphics and Visualisation 2006*, ACM, 63–68.