

The Role of Subsurface Scattering in Glossiness Perception

DAVIT GIGILASHVILI, Norwegian University of Science and Technology, Norway

WEIQI SHI, Yale University, USA

ZEYU WANG, Yale University, USA

MARIUS PEDERSEN, Norwegian University of Science and Technology, Norway

JON YNGVE HARDEBERG, Norwegian University of Science and Technology, Norway

HOLLY RUSHMEIER, Yale University, USA

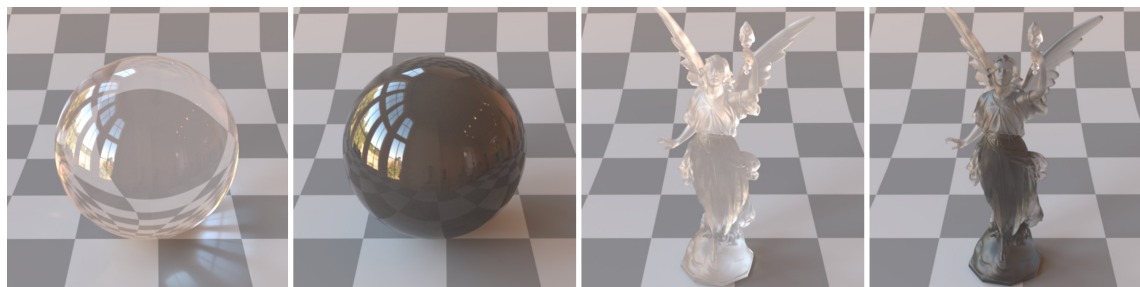


Fig. 1. Examples of materials and shapes used in the study of the impact of subsurface scattering on gloss.

This study investigates the potential impact of subsurface light transport on gloss perception for the purposes of broadening our understanding of visual appearance in computer graphics applications. Gloss is an important attribute for characterizing material appearance. We hypothesize that subsurface scattering of light impacts the glossiness perception. However, gloss has been traditionally studied as a surface-related quality and the findings in the state-of-the-art are usually based on fully opaque materials, although the visual cues of glossiness can be impacted by light transmission as well. To address this gap and to test our hypothesis, we conducted psychophysical experiments and found that subjects are able to tell the difference in terms of gloss between stimuli that differ in subsurface light transport but have identical surface qualities and object shape. This gives us a clear indication that subsurface light transport contributes to a glossy appearance. Furthermore, we conducted additional experiments and found that the contribution of subsurface scattering to gloss varies across different shapes and levels of surface roughness. We argue that future research on gloss should include transparent and translucent media and to extend the perceptual models currently limited to surface scattering to more general ones inclusive of subsurface light transport.

CCS Concepts: • **Computing methodologies** → **Perception**.

Authors' addresses: Davit Gigilashvili, davit.gigilashvili@ntnu.no, Norwegian University of Science and Technology, 22 Teknologiveien, Gjøvik, 2815, Norway; Weiqi Shi, Yale University, New Haven, Connecticut, USA; Zeyu Wang, Yale University, New Haven, Connecticut, USA; Marius Pedersen, Norwegian University of Science and Technology, Gjøvik, Norway; Jon Yngve Hardeberg, Norwegian University of Science and Technology, Gjøvik, Norway; Holly Rushmeier, Yale University, New Haven, Connecticut, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Association for Computing Machinery.

Manuscript submitted to ACM

Manuscript submitted to ACM

Additional Key Words and Phrases: material appearance, gloss perception, translucency perception, subsurface light transport, MTurk

ACM Reference Format:

Davit Gigilashvili, Weiqi Shi, Zeyu Wang, Marius Pedersen, Jon Yngve Hardeberg, and Holly Rushmeier. 2021. The Role of Subsurface Scattering in Glossiness Perception. 1, 1 (March 2021), 38 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUCTION

Humans are adept at identification of materials [2, 60] and can easily characterize their appearance [2, 19]. A typical human with normal vision does not need much effort or prior training to tell the difference between shiny and matte objects, or whether a material transmits light. Assessment of material appearance has a vital importance in our daily lives - just by visual inspection, we know whether food is edible or spoiled, whether the road is slippery or not. Tactile expectations derived from the visual appearance can guide our haptic interaction with the surrounding objects - for instance, we touch glossy, transparent crystal-looking objects with more care than we do for jelly-looking, matte objects; expecting the latter to be soft and elastic, while the former is deduced to be fragile. How the human visual system (HVS) calculates these appearance properties from the physical stimulus is far from being fully understood. Comprehending the physical processes and inverting optics [51], as well as the calculation of image statistics by our brain [44] have been named among the potential explanations, both criticized on several grounds [13, 14, 32].

Gloss is among the most important visual attributes of a material [11, 20]. It is usually associated with shininess [21] due to the specular reflection and is formally defined as an *"angular selectivity of reflectance, involving surface reflected light, responsible for the degree to which reflected highlights or images of objects may be seen as superimposed on a surface"* in the ASTM Standard Terminology of Appearance [1]. The six distinct dimensions of gloss: specular gloss, contrast gloss, distinctness-of-reflected-image gloss, absence-of-bloom gloss, absence-of-surface-texture gloss, and sheen, have been proposed by Hunter [29] back in 1937. Since then, gloss has been accepted as a surface-related quality, and perception of gloss has been studied in the context of surface scattering models [49, 62, 71]. Various image cues have been proposed to be used by the HVS for gloss perception (for instance, the total area covered by specular reflections, contrast between specular reflections and surrounding areas, the sharpness of the edges of the specular regions [38, 39]). Although it has been demonstrated that shape and illumination co-vary with the image cues proposedly used for gloss estimation [38], these cues can also be affected by the subsurface light transport.

When a light ray reaches a boundary between two media with mismatching indices of refraction, part of it is reflected specularly (i.e. the light re-emerges back towards the incidence hemisphere but on the opposite side of the surface normal) or refracted (i.e. changes the direction and continues propagation inside the new medium). The light can either get absorbed or scattered by scattering particles when propagating through a medium. An average distance a photon travels before it gets either absorbed or scattered depends on the extinction coefficient of the material. Many rendering techniques use the concept of diffuse reflectance (i.e. scattering the incident light from a surface into many different angles) for modeling opaque media. However, the optical phenomenon which is known as "diffuse reflectance" actually involves subsurface scattering of light - a photon penetrates the superficial layer of the material, where it quickly gets either absorbed by the pigments or scattered backwards towards the incidence hemisphere, defining the color of the material and generating an opaque appearance. However, if the extinction coefficient is low or the object is thin enough, a photon might re-emerge from a different side of the object - generating transparent or translucent appearance. The process is illustrated in Fig. 2. While primarily specular reflection has been thought to be responsible for glossy appearance (see [36] for a review), diffuse reflection has been shown also to be playing a role [49] - assuming negligible

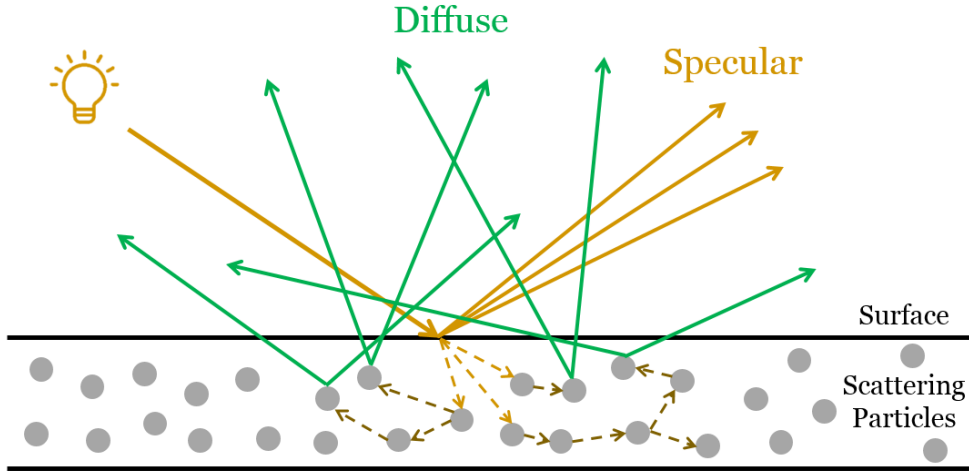


Fig. 2. Light gets either reflected specularly or refracted at the boundary of the two media with mismatching indices of refraction. What is known as diffuse reflection is actually light scattered backwards from the superficial layers of the subsurface due to high extinction coefficient. However, if the extinction coefficient is low, light can re-emerge far from the point of incidence, considerably affecting the visual appearance.

or non-existent subsurface light transport most of the time. In other words, the studies addressing gloss perception have been traditionally limited to surface reflection and fully opaque media (e.g. [15, 38, 39, 47, 49, 53, 54, 62, 64, 66, 69–71]), while a lot of materials we interact with on a daily basis, are both glossy and light-transmissive - water, glass, marble, or human skin can be named among many. The knowledge about the peculiarities of gloss perception on transparent and translucent materials is very limited.

In this paper, we hypothesize that subsurface scattering impacts glossiness perception. The hypothesis is reasoned from the following notions:

- (1) Due to the limited dynamic range and poor capability of the HVS to comprehend and invert the complex optical path of the light [14], human observers might have difficulty unmixing transmitted and surface-reflected light. Hence, caustics, direct transmission or volume scattering can be mistaken for specular reflections. Imagine a transparent crystal vase with a complex shape. It shines, has sparkles and highly luminant areas. Is it possible to tell whether the highlights are due to the reflection, direct transmission or subsurface scattering of light? Do not all these shiny parts evoke a feel of glossiness regardless of their origin?
- (2) It has been demonstrated that darker objects look glossier than lighter ones [49, 62] due to higher contrast between specular and diffusely-reflecting areas (Hunter’s contrast gloss [29]). As volume scattering and absorption can impact the contrast between specular and non-specular areas, they might also impact apparent gloss.
- (3) Observation of the mirror-like reflection image on the surface has been identified to be a strong glossiness cue [21] (Hunter’s distinctness-of-reflected-image gloss). While it has been thought to be correlated with surface roughness only [49], the distinctness of the reflected image can be dependent on light transmission properties as well. The same applies to the sharpness of the highlights, which is another glossiness cue [38, 39].

- (4) Subsurface light transport can influence the size of the highlights on complex-shaped objects. It has been demonstrated multiple times that the size of the highlights is correlated with perceived glossiness [4, 31, 38, 39].
- (5) For transparent objects, as the transmitted and reflected light integrate, overall luminance reaching the human retina is higher and the object shines more [19, 21]. Overall shine as an inherent characteristic for gloss, might evoke a perception of glossiness.
- (6) Finally, caustics and light transmission might facilitate material identification. If a stimulus is associated with a familiar, usually glossy material, the expectations about this material can impact the perception of glossiness [58].

In order to test this hypothesis, we have conducted a series of pair-comparison experiments. In the first (pilot) experiment, we studied how surface and subsurface scattering affect gloss perception on the example of spherical objects. The results of the pilot experiment have indicated that the impact of subsurface scattering on gloss varies among different levels of microfacet-scale surface roughness. This can be explained by the fact that glossiness cues vary dramatically between mirror-like and Lambertian-like surfaces [29, 53, 70]. We have interviewed several participants (members of our lab) in the pilot study. They noted that if the shape of the stimulus were different, it could have affected their answers. This correlation was deemed reasonable by us, as the macro-scale shape of the object can impact translucency and subsurface light transport [14, 19, 22]. In order to investigate further, the second experiment was arranged, studying objects with five different shapes each with five different levels of surface roughness. We analyzed the depth and curvature of object shapes and identified interesting trends in how the contribution of subsurface scattering to gloss varies among object shapes. Our contributions in this paper are the following:

- We experimentally test the hypothesis that subsurface scattering impacts gloss perception for materials with identical shape and identical surface scattering.
- We identify whether the contribution of subsurface scattering to the glossiness perception varies among different macro-scale and micro-scale (microfacet-level) shapes, and characterize this impact qualitatively.
- We discuss the need for inclusion of subsurface scattering in future studies, opening a new avenue in gloss perception research.

The paper is organized as follows: in the next section, we summarize the related work. In Sections 3-4, we present the two experiments and their results, respectively, followed by the Discussion section. Finally, we summarize the conclusions and overview the open points for future work.

2 RELATED WORK

The perception of gloss and translucency has attracted scholarly interest in vision, psychology and computer graphics alike. While substantial progress has been achieved on both topics, the two attributes have usually been studied separately from each other.

2.1 Gloss perception

One of the most widely discussed hypotheses about gloss perception is that the HVS calculates skewness of luminance histogram or a similar measure of asymmetry when assessing gloss [18, 35, 44]. Interestingly, many glossy objects have positively skewed histograms. However, it has been shown by Anderson and Kim [3] that non-glossy images can also produce similar histograms and image statistics do not fully explain the complex neurophysiological processes of gloss perception (e.g. [18, 32, 37]). Other widely studied image metrics that are proposedly related to gloss are contrast [38, 39, 49, 62], sharpness [38, 39, 49] and coverage area [4, 31, 38, 39] of the highlights. The glossiness

of a given material has been demonstrated not to be constant and can vary to a great extent, e.g. across different shapes [39, 48, 66]. In some particular cases, even Lambertian surfaces are capable of evoking gloss perception [52, 53, 70]. Gloss has been shown also to be impacted by illumination geometry [15, 48], motion [9, 56, 69] and color [46, 69]. Pellacini et al. [49] have used multidimensional scaling (MDS) and identified two perceptual dimensions of gloss that are similar to contrast and distinctness-of-image. They conclude that *"darker objects look glossier than lighter ones"*. Wills et al. [71] tried to embed bidirectional reflectance distribution functions (BRDFs) into the perceptual space. These perceptual dimensions have been modeled with physical material properties in Ward's reflectance model [68], ignoring subsurface light transport. Toscani et al. [64] have recently proposed that surface reflection has at least three perceptual dimensions: lightness, gloss, and metallicity. However, the authors did not address how these dimensions behave on highly transparent and translucent media.

2.2 Translucency perception

Translucent appearance is a result of subsurface scattering for the materials where the light can penetrate into the volume. Although Chadwick et al. [5] have reported yet imperfect still reasonable perceptual unmixing of absorption and scattering by humans in "milky tea" images, Fleming and Bühlhoff [14] argued that the HVS has poor ability to reconstruct complex processes of light and matter interaction and instead it relies on simple image cues to perceive translucency. These cues co-vary with various properties of an object. Image cues as well as the amount of light exiting the volume depend on the shape complexity and thickness of a given object. For instance, it has been shown that sharp geometric details of the object impact apparent translucency [74] and the other way round, translucency affects perception of geometric edge sharpness [6]. Sawayama et al. [57] have reported that *"sensitivity to translucent discrimination was high when the object has rugged surfaces"*. Furthermore, Gigilashvili et al. [19] have observed that objects with thin parts look more translucent and that the HVS is more sensitive to translucency differences when an object has thin parts [22]. Motoyoshi [43] observed that luminance statistics of the non-specular regions are essential for apparent translucency and that decreasing local contrast in these regions of an opaque material renders translucent appearance. Nagai et al. [45] discussed luminance statistics of potential "hot spot" image regions that are especially informative about translucency. Later, particularly edges have been proposed to contain a vital portion of the information for translucency assessment [23]. Similar to gloss, the translucency of a material is not constant either. It has been shown to be dependent on the illumination geometry [17, 73] and shape [14, 19]. Gkioulekas et al. [24] have examined translucent appearance in the context of computer graphics and found that the phase function of volume scattering affects translucent appearance.

2.3 Impact of translucency on gloss

Gigilashvili et al. [19] reported no significant differences in gloss perception of five physical spherical objects with identical surface roughness but different translucency and color. The authors revisited the study in [21] and after analyzing the observer interviews, they discovered that different people rely on different cues. The authors have identified three groups of people with different approaches to solve the gloss-based ranking task. While objects with identical surface were automatically considered equally glossy by some subjects, two other groups used different cues for ranking, either overall shininess of the object - mostly present in transparent and translucent spheres, or distinctness-of-image and contrast - that were higher for more opaque ones. When the experiment was conducted using complex-shaped objects instead of spherical ones [21], the majority of the observers considered translucent objects glossier than their opaque counterparts. The authors hypothesize that this happens due to the complex shape, which

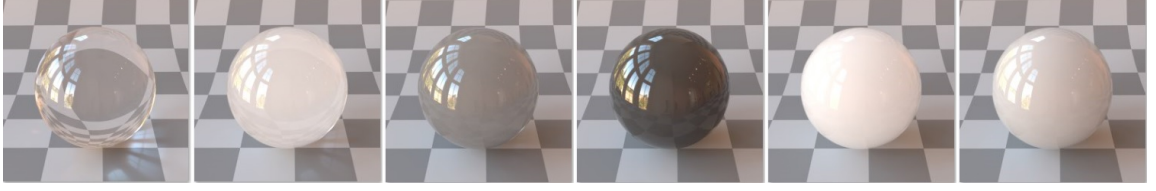


Fig. 3. Spheres with the same surface roughness ($\alpha=0$) but different subsurface scattering properties. $[\sigma_t, \text{albedo}]$ parameters of these spheres are equal to $[0.10, 0.50]$; $[1.00, 0.90]$; $[2.00, 0.60]$; $[3.00, 0.30]$; $[3.00, 0.95]$; $[4.00, 0.90]$, from left to right, respectively.

generated more caustics and back-reflections for translucent and transparent materials, while lacking distinctness-of-image for the opaque ones. They refer to the reasoning by Fleming and Bülthoff [14] about poor optics inversion ability of the HVS and propose that subjects might have mistaken caustics for specular reflections. If that is possible for physical objects during direct interaction, confusion can be even larger in computer graphics, where haptic interaction is impossible and tactile information is absent. It is worth mentioning that these works have been primarily of a qualitative nature. To the best of our knowledge, this is the first work quantitatively evaluating the impact of translucency on gloss.

3 EXPERIMENT 1: PILOT STUDY

3.1 Methodology

3.1.1 Objectives. The objectives of this experiment are two-fold: firstly, we test a hypothesis that subsurface scattering impacts gloss perception when surface scattering and object shape are identical; secondly, we observe how surface and subsurface scattering impact perceived gloss together.

3.1.2 Stimuli. We began our study by considering different scenes to use for our experiments. For illumination, we followed the previous work [24] using the side-lighting by rotating Bernhard Vogl’s museum environment map provided by Mitsuba [30] to a proper angle. We created synthetic images of spherical objects using a physically-based rendering in Mitsuba. Spheres have been widely used in the past for studying gloss perception (e.g. [15, 19, 49, 62, 72]). For surface reflectance, we used an isotropic rough dielectric microfacet model with the Beckmann distribution [30]. The model is defined by roughness α (the root mean square slope of microfacets) and an index of refraction IOR . As we restrict our attention to subsurface scattering effects, we use a fixed IOR of 1.5 which is typical for translucent media such as glass, wax and polymeric materials [42, 59]. All objects were placed on a Lambertian checkerboard. It is important to highlight that the rendering technique we used [30, 67] has accounted for Fresnel effects. Fresnel effects imply that the amount of observed reflectance varies with the observation angle, which have been shown to be important for gloss perception [12] and for appearance of dielectric materials, in general [26]. The experiment was conducted in two rounds: since our primary goal was to explore whether subsurface light transport influences gloss perception, in the first round we compared objects with an identical surface roughness parameter (also referred to as α) and different parameters of subsurface scattering. In order to explore how the impact of volume scattering on gloss perception varies among the different levels of surface roughness, we have repeated the experiment for the different α s separately. In the second round, we compared the stimuli with different α s. We select roughness from the set $\{0, 0.05, 0.1, 0.25, 0.5\}$ to cover a wide range of surface reflectance behavior. Some of the stimuli are illustrated in Fig. 3-4.

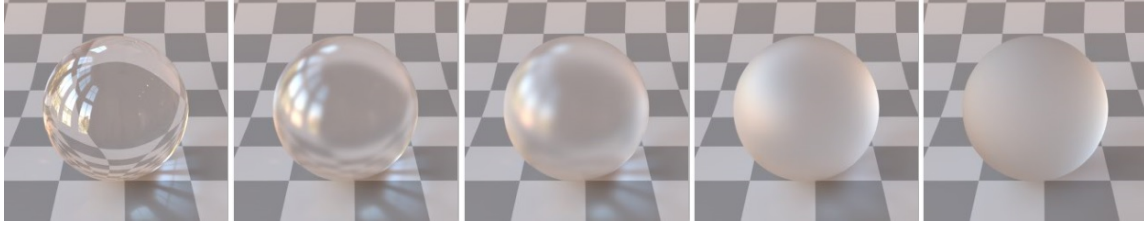


Fig. 4. Spheres with the same subsurface scattering properties ($\sigma_t=0.10$ and $\text{albedo}=0.50$) but different surface roughness, with α equal to 0.00, 0.05, 0.10, 0.25, 0.50, from left to right, respectively.

We used a homogeneous isotropic subsurface scattering model to simulate the translucent appearances. For this pilot we assume an isotropic phase function and wavelength-independent scattering and absorption for subsurface light transport. The subsurface scattering parameters are the extinction coefficient σ_t and *albedo*. For the extinction coefficient, we found through experimentation that increasing σ_t over 10 does not yield significant differences in appearance for our shape because the material becomes opaque. Therefore, we selected $\sigma_t \in \{0, 0.1, 0.5, 1, 2, 3, 4, 5, 10\}$. For albedo we selected $\text{albedo} \in \{0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95\}$. Such a dense sampling of parameters covers a wide range of appearance but would require an enormous number of comparisons to be evaluated. Although the parameters have been selected based on visual inspection in a trial-and-error manner, many pairs of parameter values still lead to indistinguishable appearances, which are redundant for the user study. To select a smaller set of parameter combinations for stimuli with the same surface reflectance, we used the K-means clustering algorithm to find six distinctive clusters based on different subsurface scattering parameters. We used the averaged Euclidean distance of pixels from the rendered images as a metric to perform K-means clustering. We have explored other clustering algorithms, such as affinity propagation [16], but K-means has provided the best clustering results according to the silhouette coefficient. We used the cluster center as our stimulus for the user study. Since the K-means has been conducted separately on different groups of surface roughness, the cluster centers were not identical for all surface roughness levels. The variation in the cluster centers was small, however, and so we selected identical subsurface scattering parameters for all levels of surface roughness. 30 different stimuli were used in total (5 different levels of surface roughness $\{0.00, 0.05, 0.10, 0.25, 0.50\}$ and 6 different combinations of σ_t and *albedo*, where $[\sigma_t, \text{albedo}] \in \{[0.10, 0.50]; [1.00, 0.90]; [2.00, 0.60]; [3.00, 0.30]; [3.00, 0.95]; [4.00, 0.90]\}$). We used the volumetric path tracing integrator of Mitsuba to render the stimuli with 512×512 pixel resolution and 16384 samples per pixel. The tonemapped (clipped) low dynamic range images have been used to ensure the compatibility with the user displays. All images can be found in *supplementary materials* (Fig. 23).

3.1.3 Experimental Design. We considered two different designs of 2 alternative forced-choice task: either displaying two stimuli and asking the subjects (also referred to as users) to select a glossier stimulus, or displaying three stimuli and asking to select two stimuli closer to each other in terms of gloss (a setup similar to Wills et al. [71]). We ran a preliminary study with both designs. 8 members of our lab completed the tasks and participated in informal post-experiment interviews. 7 subjects out of 8 mentioned that selecting a glossier stimulus between the two was an easier task than comparing the three by similarity. They also admitted that oftentimes they had found it difficult to isolate gloss from total appearance and were tempted to judge similarity by overall appearance or lightness. Therefore, we selected the former option for the task design.



Fig. 5. The user interface identical to this one has been used to conduct the experiments on the Amazon Mechanical Turk.

First, we conducted separate paired-comparison experiments for each level of *alpha*. The users were shown two spherical objects with the same surface roughness and different subsurface scattering parameters. They were asked to select the one with a glossier appearance. The user interface is illustrated in Fig. 5. The proper command of English among subjects was ensured with the Amazon Mechanical Turk average approval rate filter (see section 3.1.6). Only the users with a positive track record of similar tasks were allowed to participate. The following instruction was given to them: **"Click on the image that contains the glossier object. You can click after taking two seconds to look at the images."** No further definition or guidance was provided. The reason for abstaining from a definition is the following: any particular definition for *gloss* could have biased subjects' decisions. For instance, as mentioned above, the ASTM Standard Terminology of Appearance [1] defines *gloss* as "*angular selectivity of reflectance, involving surface reflected light, responsible for the degree to which reflected highlights or images of objects may be seen as superimposed on a surface.*" Reference to the definition that highlights gloss as a *reflectance* property might have had an implication for some subjects that *subsurface scattering effects* should be ignored. This contradicts the objective of this experiment. The research objective of this study was the identification of the factors impacting the overall sensation of gloss, not the psychometric measurement of an internal function for a given visual cue. It is worth mentioning that seminal works on gloss perception (e.g. [49, 71]) usually have no mention that the term was defined for the subjects, unless the objective is a psychophysical measurement of a particular, explicit cue (such as *specular contrast* and *specular sharpness* in [38]).

There was no time limit for each trial. Each user was asked to complete 100 trials in random order, of which 75 were unique trials (6 different materials yield 15 trials for each roughness level, totalling to 15×5) and 25 were repeated trials with images in reverse order. We used the repeated trials to assess *intra-rater reliability* by counting the number of pairs (out of 25) the subject selected the same stimulus on both trials. We designed our system with a delay mechanism: the users could only select the candidate image two seconds after the pair was displayed. This mechanism makes sure that users take time to examine the images. The users, on average, took about 5 minutes to assess 100 comparisons.

To understand how surface reflectance and volume scattering influence gloss perception together, we conducted a second round of paired-comparison experiments, where the two candidate images had different surface roughness. Instead of dividing the 30 stimuli into 5 groups and conducting experiments separately for each roughness level, this time the users had to compare the stimuli from different roughness groups, yielding 360 unique pairs in total (each of the 30 stimuli was compared with other 24 stimuli of different *alpha*; from the first round of the experiment we

already had the data for the objects with the same α). 25% of the pairs were shown twice for controlling *intra-rater reliability*.

3.1.4 Analysis: Hypothesis testing. We formulate a null hypothesis that subsurface light transport has no impact on gloss perception. In order to test the null hypothesis, we conducted Binomial exact statistical significance tests, as our outcome is binary. Under the null hypothesis, the expected probability of each stimulus being considered glossier is 0.50. We assess observed frequencies and calculate the probability of observing those frequency values when the null hypothesis is true. As it is not important at this stage which of the two stimuli is glossier (we just want to show that subsurface scattering makes them look different in terms of gloss), we conduct a two-tailed test - i.e. it does not matter whether the observed frequency is larger or smaller than the expected one. If the probability of observing given frequencies is less than 0.05 under the null hypothesis, the difference is deemed significant and the null hypothesis is rejected. In order to avoid falsely rejecting the null hypothesis due to multiple testing (type I error), we applied Holm-Bonferroni [28] correction to the data.

3.1.5 Analysis: Z-scores. A further method to analyze the pair-comparison data is Z-scores (Standard scores) [10, 65]. It is based on Thurstone’s law of comparative judgment [63] - assuming that each sample has a quality that is being assessed by a subject and these qualities are Gaussian random variables. Each time a subject compares the two samples, realizations from both random variables are drawn and compared, selecting the one with higher quality. The probability of selecting a given option is found using the standard normal cumulative distribution function (CDF). The inverse CDF of the standard normal is a Z-score showing how many standard deviations away is a given option from the mean. Usually, Thurstone’s simplified Case V model is used assuming that all samples are independent and have equal variance [65]. For all samples we present the mean Z-scores and their 95% confidence intervals as error bars (calculated using MATLAB Colour Engineering Toolbox [25]). The mean Z-score shows how far a given stimulus is from the mean of the set of stimuli being assessed. If the 95% confidence intervals of the Z-scores do not overlap, we can tell with 95% confidence that the qualities of the two stimuli are significantly different.

3.1.6 Subjects. The sample size is found by desired statistical power, significance level and effect size for the Binomial null hypothesis testing. The desired statistical power was set to 0.8 (the probability of rejecting the null hypothesis when the alternative hypothesis is true) and the significance level was set to 0.05 (the probability of falsely rejecting the null hypothesis when it is actually true). As per the null hypothesis two stimuli are equally glossy, the expected probability is 0.5. In order to decide on alternative proportion, two different effect size metrics [55] were used: Cohen’s g - usually used for the cases where the expected proportion is 0.5 and simply found as a difference between the proportions, and Cohen’s h - that is found as:

$$h = 2 (\arcsin \sqrt{p_1} - \arcsin \sqrt{p_2}) \quad (1)$$

where h is Cohen’s h (sometimes reported as an absolute value) and p_1 and p_2 are the two proportions. Under an alternative proportion of 0.75, $g=0.25$ and $h=0.52$, being interpreted by Cohen [7] (cited in [55]) as large and medium effect sizes, respectively. Thus, we set an alternative proportion to 0.75. Considering these values, the needed sample size was approximated as 29.

We conducted our experiments on Amazon Mechanical Turk (MTurk) and collected responses from 50 users per pair. In total, around 250 subjects participated in both rounds. The users were compensated for participation. The compensation varied from experiment to experiment and was within the range of 2-3 USD per 100 comparisons. In order to ensure the reliability of the users, two filters were applied: firstly, only the MTurk users with an average

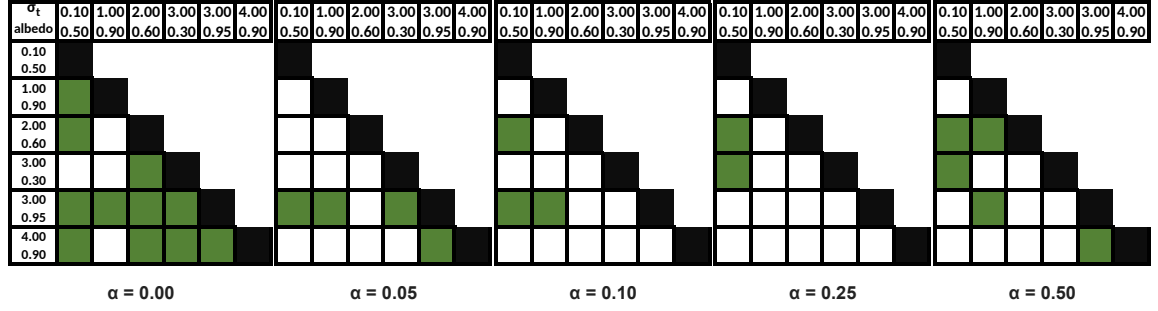


Fig. 6. Significance tables for each roughness level. Each lower triangular matrix shows which of the stimuli pairs are significantly different. Green cells - statistically significant difference; white cells - no statistically significant difference. The number of significantly different pairs is larger for smooth objects (alpha equal to zero).

approval rate above 50% were allowed to participate; and secondly, the participants were ranked by their performance in the *intra-rated reliability* test, i.e. by the consistency of their responses on the validation set (how many times they selected the same stimulus in the pairs shown twice). Eventually, 30 most consistent subjects were considered per stimuli pair, around 150 subjects in total. The reason for users' inconsistency can be not only their inattentiveness but also the stimuli which are visually indistinguishable. The number of such pairs is unknown before the experiment and hence, it is not possible to set a threshold for "acceptable consistency" in advance. For this reason, we had to rely on ranking instead of absolute values of consistency. Interestingly, the top 30 users turned out to be consistent in at least 70% of the cases. Finally, it is worth mentioning that the results with concurrent clicks from the same IP address were discarded, because it was impossible to calculate their intra-rater reliability and to identify how many unique subjects were responding.

3.2 Results

The results for the fixed roughness experiment are shown in Fig. 6-7. Fig. 6 shows that the difference is significant and the null hypothesis can be rejected for a substantial number of image pairs. This is especially true for smooth objects. The number of pairs that are significantly different gradually decreases, but for $\alpha = 0.50$ it starts increasing again. While the two-tailed Binomial tests can just tell whether the difference is significant, the Z-score plot in Fig. 7 illustrates which stimuli have been deemed glossier. If the null hypothesis were true, all stimuli were expected to end up with similar Z-scores. However, the observed trend is consistent with the Binomial tests - the difference among some stimuli is significant and it is large for smooth objects while the difference gradually diminishes but starts increasing again for the highest α . The materials either with low σ_t or albedo were considered glossiest, while the ones with high albedo turned out less glossy. The results including all comparisons among the 30 stimuli are shown in Fig. 8-9. The significance table shows that the vast majority of the differences between different roughness levels are significant, while no significant differences are usually observed among the objects with the same roughness. However, there are a few exceptional instances - the materials with high albedo (0.95) are not significantly glossier than some other objects with a rougher surface (Fig. 8). Examples of the objects with different surface roughness but equivalent (not significantly different) apparent gloss are illustrated in Fig. 10.

A clear trend is visible in Z-score plots (Fig. 9) - with the increase of surface roughness, the perception of glossiness is decreasing monotonically, being consistent with the prior works [27, 54]. It is worth noting that although it is the

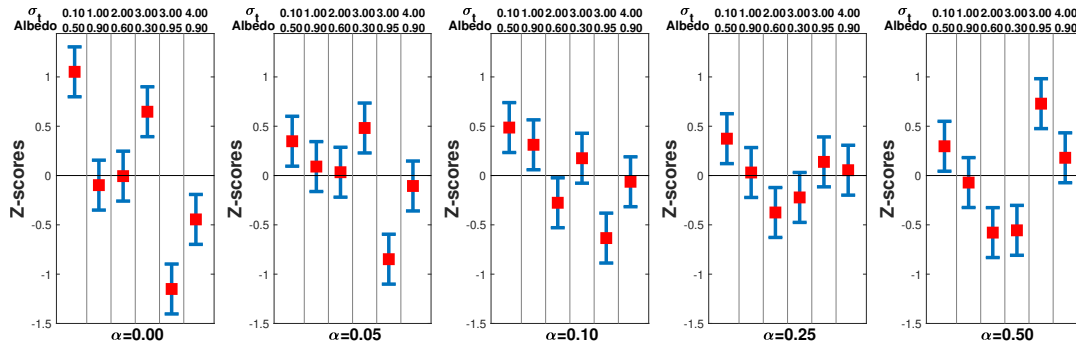


Fig. 7. Z-scores for fixed roughness experiments. A red cube corresponds to the mean Z-score for a given object, while the error bar corresponds to a 95% confidence interval. The variation among Z-scores decreases with the increase of roughness, i.e. Z-scores of five different materials are more equal when α is high. However, this trend is not monotonic and it does not hold for $\alpha=0.50$.

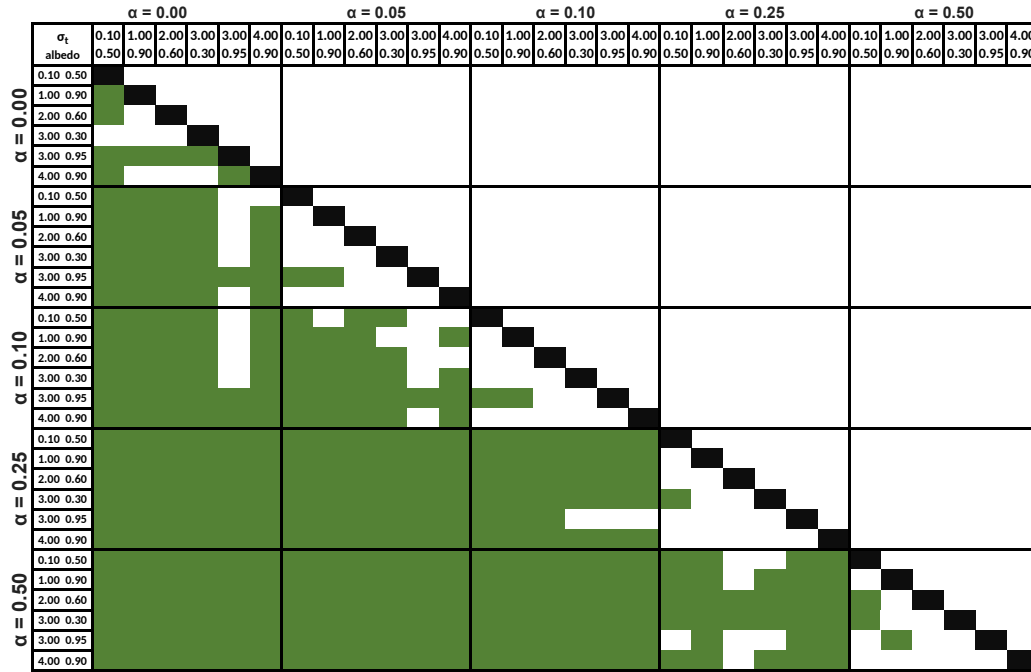


Fig. 8. The significance table for all 30 stimuli. The lower triangular matrix marks the stimulus pairs with statistically significant difference. Green cells - statistically significant difference; white cells in the lower triangle - no statistically significant difference.

identical data, the Z-score differences among the stimuli within each roughness group decreases when considered together with all other stimuli (compare Fig. 7 and 9). This can be explained by the fact that a Z-score for a given stimulus is relative and depends on the judgment against all other stimuli in the set. Within a larger pool of stimuli and various α s, the subjects tend to focus more on the surface reflectance instead of the subtle effects of subsurface light transport. All these observations demonstrate that even though the subsurface light transport has an impact, the surface reflectance still plays a major role in the perception of glossiness.

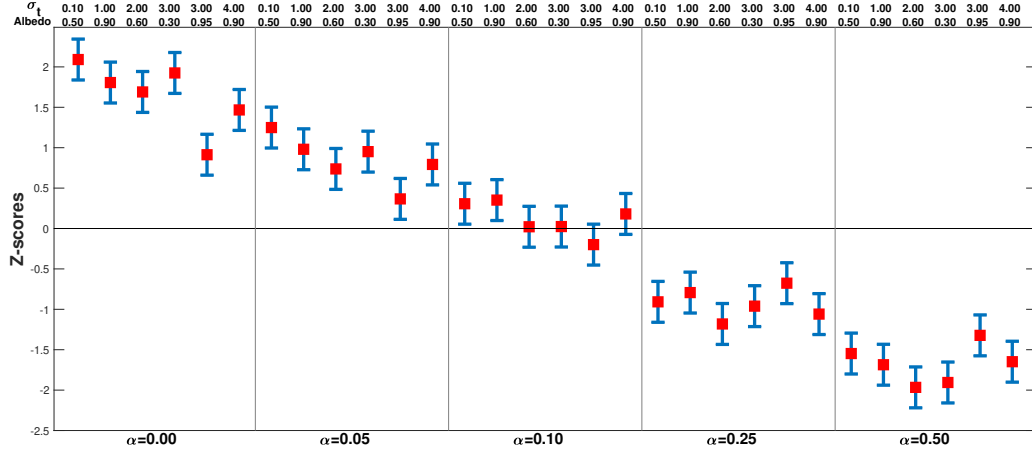


Fig. 9. Z-scores for the comparisons of all 30 stimuli. As we observe, surface scattering is dominant over subsurface scattering and smoother objects usually look glossier. However, in some cases, high albedo makes objects no glossier than some of the rougher ones.

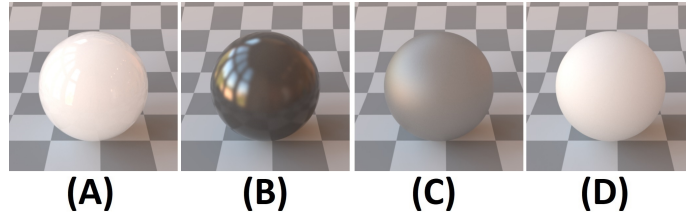


Fig. 10. The difference between apparent gloss of the objects A and B, as well as between C and D, has been shown not to be significant. We can consider them having equivalent apparent gloss. Even though A has smoother surface ($\alpha=0.00$) than B ($\alpha=0.05$), low albedo of the latter compensates for the difference in surface scattering. Similarly, C has relatively smoother surface ($\alpha=0.25$) than D ($\alpha=0.50$), but in this case, it is the high albedo of the latter that is responsible for the equivalent apparent gloss despite substantial difference in surface scattering.

3.3 Discussion

While surface roughness has a strong negative impact on gloss (being consistent with [49, 62]), for numerous pairs of the stimuli with identical surface roughness, we have rejected the null hypothesis and observed a significant gloss difference induced by subsurface scattering of light. The way subsurface scattering impacts gloss perception differs among different levels of surface roughness and changes non-monotonically.

When α is low and σ_t is high, gloss increases as the albedo decreases. This phenomenon is illustrated in Fig. 11 (also supported by the plot in Fig. 16). With a high extinction coefficient, the subsurface light penetration is reduced, yielding appearance closer to diffuse reflectance. This scenario can be paralleled with a diffuse component in Ward's surface reflectance model: decreasing the diffuse reflectance leads to glossier appearance - proposedly, due to increased contrast, making our observations consistent with that of Pellacini et al. [49].

When the stimuli are rough (high α) and do not have strong glossiness cues (such as specular highlights), caustics or the overall shinier look created by high volume scattering could potentially be considered a glossiness cue. This might explain why people can still tell the difference between the stimuli with high α in our experiments, and

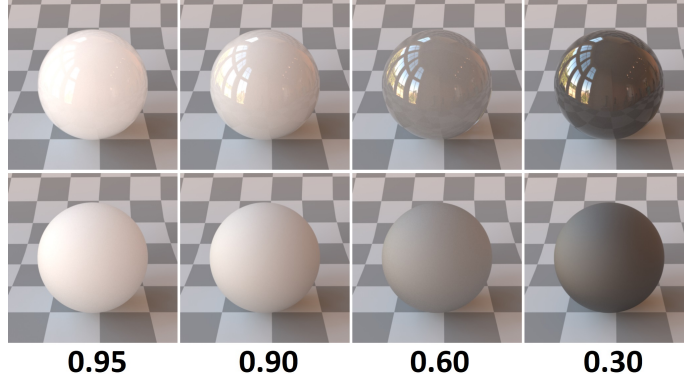


Fig. 11. The number below the images corresponds to their albedo. Although all of the objects have identical surface roughness in each row ($\alpha=0.00$ in the top row; $\alpha=0.50$ in the bottom one), the users have distinguished them in terms of glossiness. According to user responses, the top row can be ranked in terms of apparent gloss, from left to right, the rightmost one being the glossiest. The bottom row can be ranked in the opposite way - the leftmost one being glossiest (but the difference between the two rightmost ones is not significant).

why Lambertian surfaces are capable of evoking perception of glossiness [53, 54]. In general, the stimuli with low σ_t and smooth surface ($\alpha=0$) were selected as the glossiest (see the leftmost image in Fig. 3). The caustics and back-reflections from the background might be reasons for this (a similar trend has been observed for some subjects in Gigilashvili et al. [19, 22]). Furthermore, the glass-like appearance can also evoke a stronger perception of glossiness due to material identification and the association with the properties of a familiar material, as proposed by Schmid et al. [58]. Several important points have been learned from this experiment that guided the subsequent experiments:

- Since the way subsurface light transport contributes to gloss depends on the surface scattering, we decided to study this contribution for each surface roughness level individually.
- If the change in surface scattering induced by subtle changes in microfacet slopes has a dramatic impact on the behavior of subsurface scattering, we believe the same will be true for macro-scale changes of the object shape. Therefore, we decided to study the contribution of subsurface scattering for multiple different shapes individually and to compare the trends among them.

4 EXPERIMENT 2: IMPACT OF SHAPE

4.1 Methodology

4.1.1 Objectives. Experiment 1 provides evidence that subsurface scattering can impact gloss perception for spherical objects, and this impact depends on the amount of surface scattering. The objective of Experiment 2 is to quantitatively study whether subsurface scattering impacts glossiness perception in shapes other than a sphere, and to explore qualitatively how these effects vary with the shape complexity expressed in depth and curvature.

4.1.2 Stimuli. The same scene and rendering technique was used as in Experiment 1. In order to study a broad spectrum of stimuli, we varied the same three parameters as in Experiment 1 and also the shape of the object, where $shape \in \{sphere, spiky\ sphere, Stanford\ Lucy, low\ resolution\ Lucy, cylinder\}$ and $\alpha \in \{0, 0.05, 0.1, 0.25, 0.5\}$.

The sphere had already been studied in Experiment 1, while Experiment 2 was conducted on four new shapes. Several factors were considered when selecting the shapes: we need a shape that differs from a sphere by surface complexity

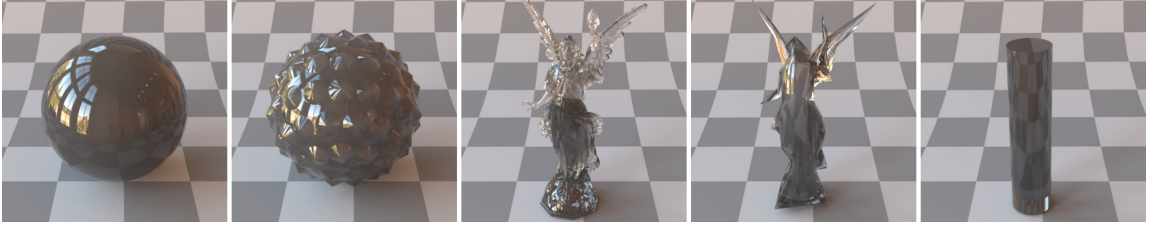


Fig. 12. Five different shapes have been studied throughout the experiment. Left to right: sphere (3.00; 0.30), spiky sphere (3.00; 0.30), Stanford Lucy (5.00; 0.10), low-resolution Lucy (5.00; 0.10) and cylinder (5.00; 0.10). The numbers given in the parentheses are σ_t and albedo, respectively. $\alpha=0.00$ for all of them.

and curvature, i.e. does not have large curved areas and does not reflect the mirror image of the environment (if you pick it up, you cannot see yourself); has many fine details; is not compact, has thin parts that transmit light well; we selected the *Lucy* from the Stanford 3D Scanning Repository [33], as it satisfies these conditions and has been used in other works for studying the appearance of translucent materials (e.g. [24]). Afterwards, we wanted to isolate several features and selected the following objects: is as thick as a sphere but has more complex surface geometry - *spiky (bumpy) sphere*; has little thickness, similar to *Lucy*, has thin parts, but lacks fine details, has relatively simple surface geometry and lower curvature - the *low-resolution Lucy*; the main body is as thick as that of *Lucy*, but lacks thin parts and has very simple surface geometry and a very low curvature - a *cylinder*. The objects are illustrated in Fig. 12.

We defined the initial pool of subsurface scattering properties as $\sigma_t \in \{0, 0.1, 0.5, 1, 2, 3, 4, 5, 10\}$ and $\text{albedo} \in \{0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95\}$. We performed a clustering process similar to that used in Experiment 1 (described in section 3.1.2). As the clustering was conducted for each individual shape and surface roughness, the cluster centers were not identical among them. Although the difference was negligible among the surface roughness levels, it was substantial between the sphere and the *Lucy*. Therefore, we selected two sets of $[\sigma_t\text{-albedo}]$ pairs, $\{[0.1, 0.5]; [1.0, 0.9]; [2.0, 0.6]; [3.0, 0.3]; [3.0, 0.95]; [4.0, 0.9]\}$ for spiky sphere (identical parameters had already been used for a sphere in Experiment 1), and $\{[0.5, 0.8]; [1.0, 0.4]; [3.0, 0.4]; [3.0, 0.7]; [3.0, 0.9]; [5.0, 0.1]\}$ for the *Lucy*, low-resolution *Lucy* and the cylinder. All images can be found in *supplementary materials (Fig. 23-27)*.

4.1.3 Experimental Design. The experimental design was identical to the first round of Experiment 1. The objects were compared only with the objects of similar shape and α .

4.1.4 Analysis. Similarly to Experiment 1, Binomial tests were conducted to test the null hypotheses for each pair, and Z-scores were calculated to assess the big picture. In addition to this, a scatter plot of Z-scores as a function σ_t and albedo was plotted to identify how these individual parameters of subsurface light transport affect gloss. Finally, we used the variance of the Z-scores and the number of significantly different pairs for a given shape and α , to compare the magnitude of the subsurface scattering impact on perceptual gloss. The shapes have been quantified in terms of *depth* (thickness) and *surface curvature*. The 3D models were presented in dimensionless units - the radius of a sphere was considered 1, and all other shapes were quantified relative to that. *Depth* was defined as a range of coordinates in all three dimensions separately, covered by the point cloud of a given object. *Local surface curvature* (Gaussian and mean) has been calculated for all points on the object surface [8, 41] and average values have been reported.

4.1.5 Subjects. The procedure was identical to Experiment 1.

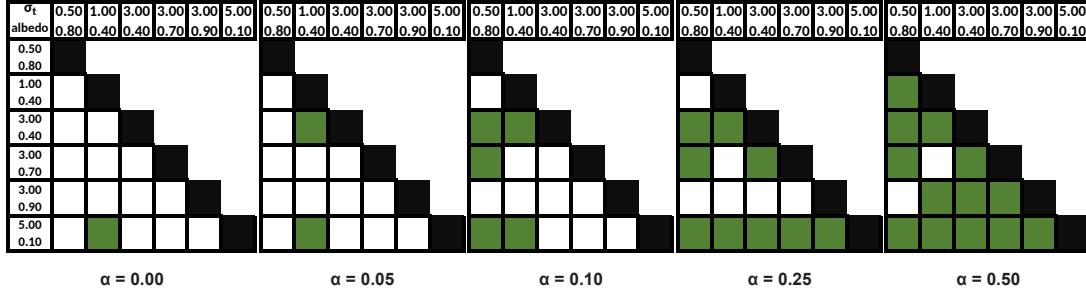


Fig. 13. The results for Lucy. Significance tables for each roughness level. Each lower triangular matrix shows which of the stimuli pairs are significantly different. Green cells - statistically significant difference; white cells - no statistically significant difference. The number of significantly different pairs is larger for rough objects.

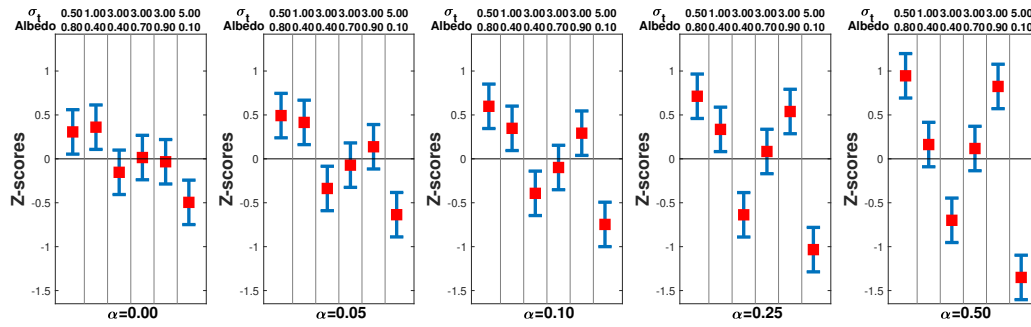


Fig. 14. Z-scores for Lucy. A red cube corresponds to the mean Z-score for a given object, while the error bar corresponds to 95% confidence interval. The difference among Z-scores grows with the increase of roughness.

4.2 Results

With this experiment, we wanted to answer three questions:

- (1) Does subsurface scattering affect gloss for object shapes other than a sphere?
- (2) How does the impact of subsurface scattering on gloss co-vary with surface roughness for object shapes other than a sphere?
- (3) How do σ_t and albedo relate with the perceived glossiness and how does this differ across the shapes?

4.2.1 Does subsurface scattering impact gloss? In Experiment 1 we demonstrated with spherical objects that subsurface scattering impacts gloss perception. The results for the Lucy are shown in Fig. 13-14. Although the results are not one-to-one comparable with that of a sphere due to the differences in subsurface scattering parameters, the following contradiction in the overall trends still stands out (compare with Fig. 6-7): the impact is subtle for smooth Lucy objects and the contribution of subsurface scattering increases with *alpha*, while the opposite is true for spherical objects. The null hypothesis was rejected for 13 out of 15 pairs when *alpha*=0.5, while it was rejected for one pair only when *alpha*=0. The results for the spiky sphere and low-resolution Lucy closely follow the trends of a sphere and Lucy, respectively. Interestingly, a cylinder was the least affected object by the change in subsurface scattering. The detailed results for those shapes can be found in the *supplementary materials* (refer to Fig. 28-36 for all results).

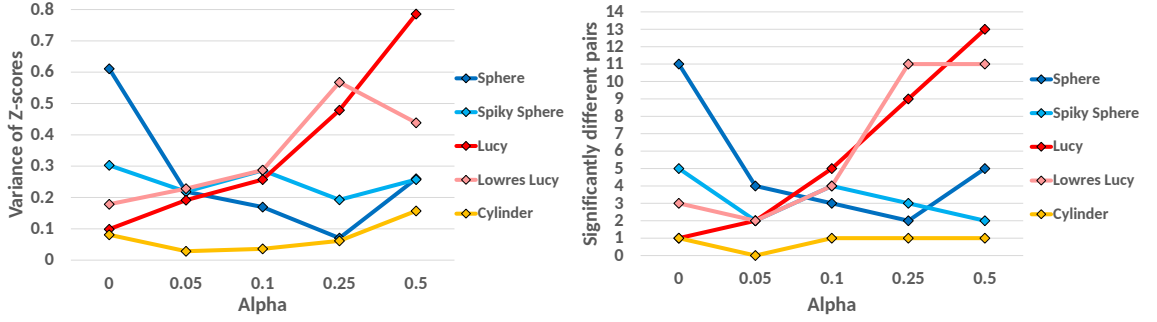


Fig. 15. The variance (left) of the mean Z-scores and the number of significantly different pairs (right). The two metrics are consistent.

4.2.2 Impact of α across different shapes. We compared the variance of the mean Z-scores, as well as the number of statistically significantly different pairs (out of 15) for each shape and α . The results are shown in Fig. 15. As expected, the results are very consistent between the two metrics. The large variance of the Z-scores or the higher number of significantly different pairs means that the variation in subsurface scattering leads to larger gloss differences. The σ_t and albedo parameters used for rendering, although subtly, still differ between a sphere and spiky sphere, on the one hand, and the Lucy, the low-resolution Lucy and the cylinder, on the other hand. This makes it challenging to directly compare the results between the two groups. However, we can still observe how the variance changes with α for a given shape. For spherical objects, the impact of subsurface scattering on gloss is larger when $\alpha=0$. The impact gradually diminishes as α increases, but interestingly, the impact starts climbing again when $\alpha=0.5$. Conversely, the impact of subsurface scattering on Lucy-shaped objects increases with the α . It is also worth noting that the cylinder remains the least affected object for all α s.

4.2.3 Gloss, σ_t and albedo. Till now the impact of subsurface scattering on gloss perception was discussed as a whole, single phenomenon. However, for modeling purposes in the future, it is of vital importance to identify how each particular physical attribute relates to the perceived gloss. Mean Z-score as a function of σ_t and albedo is shown in Fig. 16, and the mean Z-scores in the σ_t -albedo space are shown in Fig. 17- 18. Interestingly, for Lucy, there is a negative linear correlation between Z-scores and σ_t , and a positive linear correlation between Z-scores and albedo (refer to Fig. 19). As for the sphere, the albedo is negatively correlated with Z-scores when α is low, but it becomes positive for large α s (refer to Fig. 11). Fig. 17-18 show that for both shapes the increase in α has a negative impact on low albedo materials and a positive impact on high albedo ones. The results for all other shapes are reported in the **supplementary materials**.

4.3 Discussion

The object shapes come in different surface curvature and thickness (depth). The thickness of the objects is normalized to a unit sphere radius and is shown in Table 1 (columns 1-3). It is an important parameter, because the extinction coefficient is meaningful in terms of object size - the larger the distance light needs to travel within the medium, the larger the probability of absorption and scattering is. In other words, object depth directly impacts the appearance of the dielectric materials. This explains why the trends are similar between a sphere and a spiky sphere, as well as Lucy and low-resolution Lucy. Only subtle differences have been observed between a sphere and a spiky sphere, and between Lucy and low-resolution Lucy. However, an essentially different trend has been observed in cylinders, even though

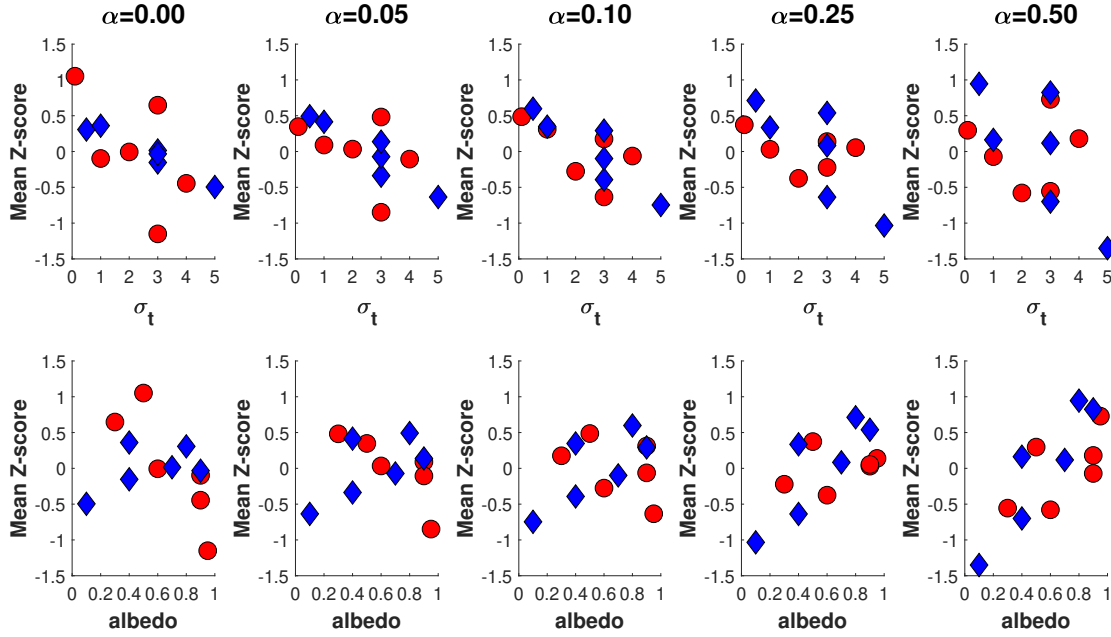


Fig. 16. Z-score as a function of the extinction coefficient (top row) and albedo (bottom). Sphere (red circles) and Lucy (blue diamonds). Linear correlations are apparent for Lucy.

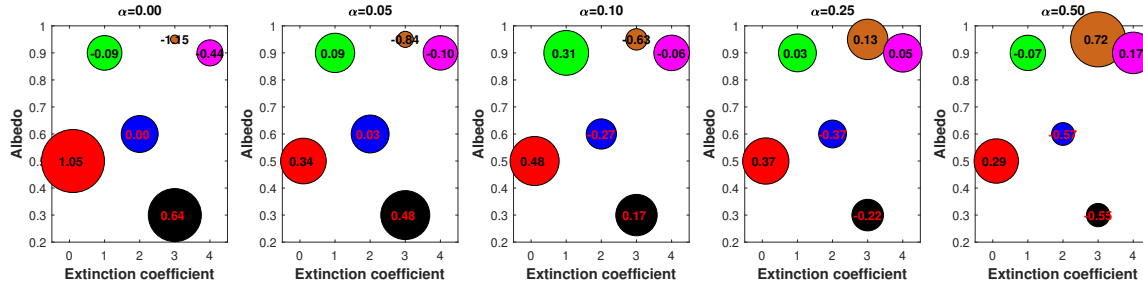


Fig. 17. The results for a sphere. Larger circle diameters represent a higher mean Z-score. Lower albedo and σ_t lead to a glossier look for smoother objects, while the trend changes as the roughness increases. Note that Z-scores are relative to the objects of the same roughness and circles of the same color are not directly comparable among the five plots.

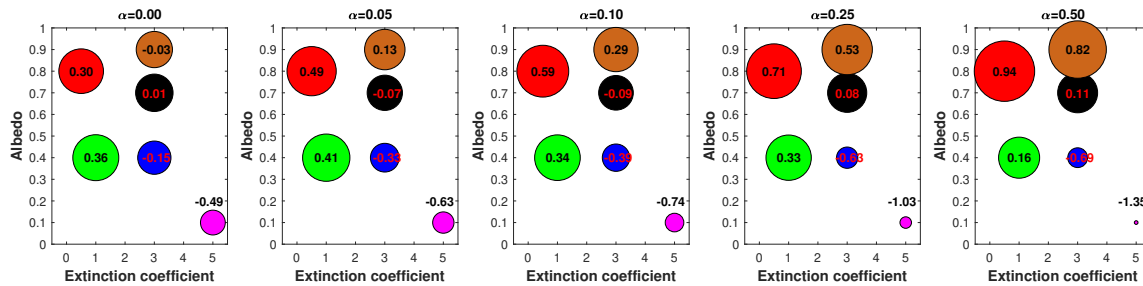


Fig. 18. The results for Lucy. High albedo and low extinction coefficient usually yield glossier stimuli.

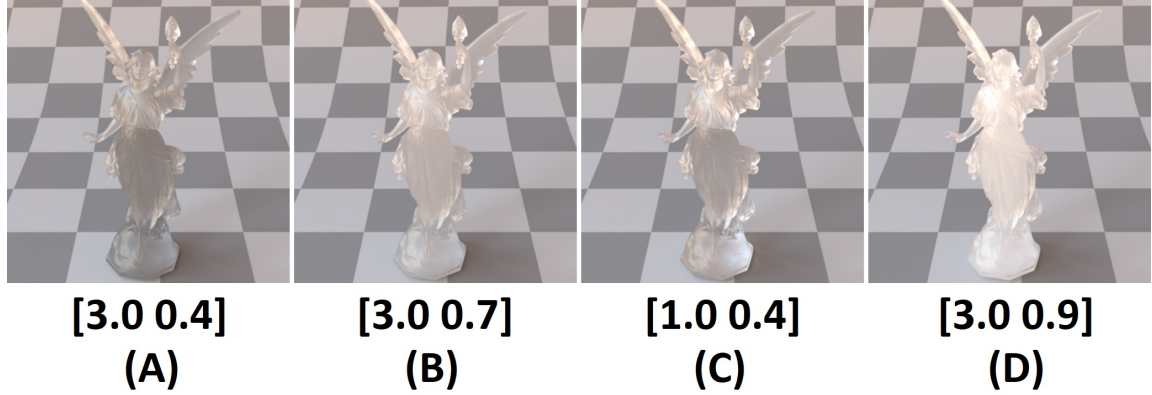


Fig. 19. The numbers in the brackets correspond to σ_t and albedo. $\text{Alpha}=0.25$ for all objects. They can be ranked by glossiness from left to right, the rightmost one being the glossiest (difference between B and C is not significant though). We can observe that although A, B and D have identical σ_t , higher albedo makes them look glossier, because it generates more highlights which apparently are mistaken for specular reflections. On the other hand, A and C have identical albedo, but differ in σ_t . Low σ_t of C generates more caustics, which are also mistaken for specular reflections.

its thickness is nearly identical to the body of Lucy. This observation indicates that thickness does not account for all differences caused by shape and surface complexity - thus, curvature should also be considered.

Local surface curvature has been found on all points of the 3D object and an average value has been calculated. The curvature at a given point can have a positive or a negative sign. However, we are primarily interested in how rugged the overall surface is, and not in the directionality of the curvature, neither in convexity or concavity of the shape. Therefore, the average has been calculated among absolute values. The curvature measure is summarized in Table 1 (columns 4-5). Note that both Gaussian and mean curvatures are equal to 1 for a unit sphere, and Gaussian curvature is equal to 0 for a cylinder. Marlow and Anderson [38] demonstrate that the weighted average of sharpness, contrast, and size of the highlights account for most of the variance in gloss judgements. The authors argue that these cues are constrained by the macro-, meso-, and microscale shape of the object. For instance, specular sharpness can vary as a function of curvature, as *"specular reflections will be sharpest in image regions that run parallel to local directions of high curvature, and will be most shallow (stretched) along directions of low curvature."* Their experiments have shown that higher curvature leads to higher specular sharpness and contrast, thus - higher glossiness, albeit the correlation

	X	Y	Z	GC	MC
Sphere	2	2	2	1	1
Spiky Sphere	2.09	2.10	2.10	742.81	22.48
Lucy	0.94	1.48	2.73	22691.61	58.44
Lowres. Lucy	0.88	1.48	2.68	89.11	7.61
Cylinder	0.45	0.45	1.90	0	2.48

Table 1. The depth of the objects in **X**, **Y** and **Z** dimensions and their curvature. A sphere and a spiky sphere are larger than the rest. Lucy is the tallest. Although dimensions for Lucy and low-resolution Lucy look substantially larger than that of a cylinder, this is due to the span of Lucy's wings. The approximate size of its body is 0.45 both in X and Y dimensions. The cylinder was designed after the torso of Lucy. Gaussian curvature (**GC**) and mean curvature (**MC**) are found locally for each point of the 3D object. The average of the absolute values is reported. Lucy is the shape with the highest curvature that is no surprise considering its level of fine details.

with specular coverage is subtle. However, their findings are based on fully opaque media. Sharpness and contrast will certainly be dependent on the light exiting the volume after subsurface light transport. The curvature of the surface can also influence the coverage area (size of the highlights) due to subsurface scattering, as it has been the case for high albedo Lucy in our experiment (image C in Fig. 20). This indicates that their findings are not directly transferable to translucent materials. In the future work, cross-shape comparisons are needed (e.g. sphere with Lucy) in order to identify whether objects with higher curvature look glossier for translucent objects as well.

Interestingly, for low curvature objects, low σ_t materials (transparent) and materials with high σ_t and low albedo (dark opaque) are considered glossiest (refer to the first and fourth images from the left in Fig. 3). We conducted an additional experiment with 15 smooth spherical objects and applied the nonclassical nonmetric multidimensional scaling analysis (MDS) with raw user response frequency as a distance matrix. From the extracted features, we can see that transparent low σ_t and dark opaque materials were placed close to each other in 2D embedding (refer to Fig. 37 in *supplementary materials*). The same trend holds for higher dimensions. Marlow and Anderson [38] also see similarities between the two types of materials and propose that similar mechanisms might be used in both cases, as the clear image of the surrounding "inside or behind the depth" of the object body is visible in both cases - although one is the result of direct transmission, while the other is a mirror reflection image. The mirror reflections on dark opaque objects are intuitively associated with perceived gloss, but the link between the background image seen-through the transparent media and gloss certainly deserves further study.

Curvature could however explain the primary difference, as well as similarities in trends between a sphere and a spiky sphere (although we have not compared them directly). For low alpha, a low albedo dark opaque sphere (image A Fig. 21) is among the glossiest, while that is not that case for a smooth spiky sphere made of the same material (image B Fig. 21). This is because the high curvature of the spiky sphere does not permit a clear mirror reflection to be observed. On the other hand, the transparent object is the glossiest for both shapes (images C and D Fig. 21). However, the image cues differ dramatically between the two. The transmission image is not visible for a transparent spiky sphere (image D Fig. 21), but the curvature of spikes produces shiny highlights due to internal scattering (the resulting image is also affected by the limited dynamic range). Similarly, the lower curvature of low-resolution Lucy makes transparent one glossiest for all alphas, while that is not the case for Lucy, as its curvature does not permit clear transmission.

5 GENERAL DISCUSSION

The results of the two psychometric experiments have enabled us make the following observations:

- Subsurface scattering can impact apparent gloss. This impact depends on micro-scale surface roughness and macro-scale shape of the object.
- Subsurface scattering had larger impact on apparent gloss of smooth spherical objects than on that of rough spherical objects; for complex Lucy shape, the opposite was true - rough Lucy objects being more impacted than smoother ones; the impact of subsurface scattering on apparent gloss was subtle for cylindrical objects.
- For smooth spherical objects, apparent gloss is negatively correlated with albedo, but the correlation is positive for rough spherical objects. For Lucy, apparent gloss is negatively correlated with the extinction coefficient and positively correlated with albedo, regardless of roughness.
- Surface scattering has generally stronger effect on apparent gloss than subsurface scattering. However, in some particular instances, subsurface scattering could compensate for surface scattering effects, yielding equivalent gloss appearance on the objects with different surface roughness.

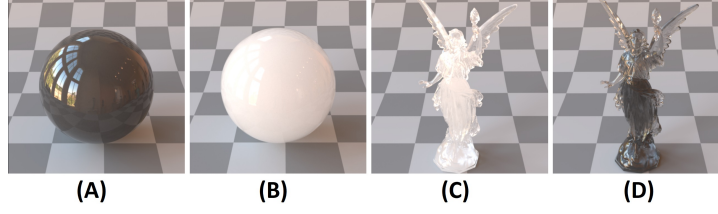


Fig. 20. Although objects A and B have identical shape and surface roughness, the lower albedo of subsurface scattering makes object A more mirror-like. Although spectral reflectance is identical, object A looks darker due to higher absorption inside the volume. Lucies in C and D have identical shape and surface roughness, but higher albedo of C generates more highlights. It is difficult to tell whether the highlights on C are specular reflections, caustics or result of volume scattering, while specular reflections are easier to isolate on low albedo object D.

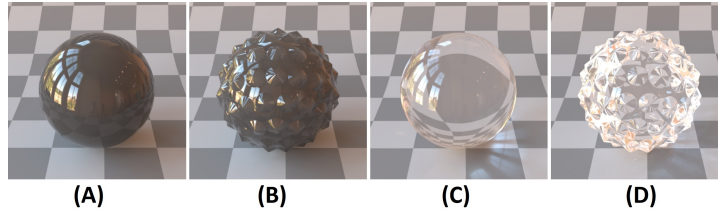


Fig. 21. Curvature influences glossiness cues - thus, the perceived relative glossiness of the objects.

5.1 The impact of subsurface scattering and its dependence on roughness

The effect of subsurface scattering was statistically significant for numerous material pairs. This is a clear indication that subsurface scattering is a contributing factor to perceived gloss and should be considered in future studies on gloss perception. However, this impact differs among the object shapes. We hypothesize that this difference comes from different image cues present in objects of different shapes and surface roughness. For more opaque smooth spherical objects lower albedo led to a glossier appearance. As the lower curvature of a spherical object produces a distinct reflected image of the environment, we believe that this is a widely used cue by the HVS for glossiness perception. The darker the object, the more distinct the reflected mirror image is. Besides, the contrast between specular and non-specular areas is also large and the reflections stand out more. This phenomenon is demonstrated in Fig. 20 - object A and B have an identical shape and surface roughness, but the subsurface scattering albedo of A is substantially lower, which makes it easier to observe the mirror reflection of the environment on it. This is consistent with the previous findings [49, 62]. As the sphere becomes rougher, the reflection of the environment, as well as specular reflections, disappear and the cues used for judgment of glossiness change. As rough objects look all Lambertian and non-glossy, the difference among them decreases. However, objects with higher albedo look lighter and shinier, which could potentially become a cue for glossiness [27, 52–54]. While the impact of *alpha* on gloss is monotonic, the impact of subsurface scattering is not. Qi *et al.* [54] have demonstrated the monotonic relationship between *alpha* and gloss, while they showed that the contribution of meso-scale roughness is non-monotonic. Further study is needed to explain why the impact is non-monotonic for spheres and why it starts increasing for *alpha*=0.5. It is interesting that for smooth spheres, the materials with the lowest extinction coefficient looked glossiest. We have speculated above that the presence of the transmission image inside the object can be reminiscent of mirror reflection, while the association with familiar material (e.g. glass), as well as caustics could have also played the role.

5.2 Shape-dependence of the effect

For Lucy-shaped objects, the opposite trend was observed. Usually, the albedo was positively correlated with gloss, the extinction coefficient was negatively correlated, and the overall impact was increasing with the roughness. If we inspect the Lucy-shaped images, we will see that the surface geometry does not allow to observe a clear reflection image, neither clear specular reflections. Subjects seemingly rely on highlighted areas that result not only from the specular reflections, but from internal scattering and caustics as well. It is difficult to tell which highlight is a specular reflection, which one is caustic, and which ones are produced by subsurface scattering - especially in low dynamic range scenarios. Naturally, high albedo objects with lower extinction coefficient produce more highlights. Refer to images C and D in Fig. 20. High albedo and limited dynamic range make it challenging to tell whether the highlights of image C were produced by specular reflections or subsurface scattering. The same task is a lot easier when the albedo is low (image D). The size of the highlights has been shown by Marlow and Anderson [38] to be positively correlated with perceived gloss. The curvature of the surface (as in the case of Lucy) can lead to large highlight areas due to high subsurface scattering. Interestingly, all smooth objects were considered equally shiny, while the differences between highlights start to prevail when the roughness is increased, producing a broader range of gloss perception.

These observations are consistent with Gigilashvili et al. [21]. They observed that the impact of translucency on gloss was different between spheres and complex female bust objects, qualitatively similar to Lucy. They interviewed the subjects and learned that the cues used for gloss estimation were different for different shapes, but they were also subject to individual interpretations. Further study is needed to investigate the reasons for the dramatic difference between sphere and Lucy results. Interestingly, the trends were similar between a sphere and a spiky sphere, as well as between Lucy and low-resolution Lucy. We believe this is correlated with the size of the objects. First of all, spheres and spiky spheres cover larger field-of-view, having a more apparent reflection of the environment than a low-resolution Lucy, which has simple surface geometry itself, but still occupies too little space of the field of view to reflect clear images of the environment. Secondly, translucency varies with the thickness of the object [14, 19] and the path light travels inside the volume is indeed more similar between a sphere and a spiky sphere than between a thick sphere and thin Lucy. However, these speculations need concrete experimental evidence. On the other hand, a cylinder is the least affected shape by subsurface scattering. The reason for this could be the fact that its curved surface enables a clear reflection image for all smooth ones, while the rough ones resemble in highlight coverage cues - in the end yielding little difference among the cylinders with the same *alpha*.

5.3 Surface versus subsurface scattering

We have observed that surface roughness usually has a stronger impact on material glossiness than subsurface scattering. However, we have also demonstrated notable examples when subsurface scattering effects compensated for surface roughness and smoother objects did not appear glossier. Interestingly, both surface roughness and subsurface scattering blur non-specular areas - both generating similar image-level measurements in these regions. If the surface is smooth and sharp specular reflections are visible, the two cases can be effortlessly distinguished (because surface roughness, unlike subsurface scattering, blurs specular highlights too). However, estimating the contribution of subsurface scattering becomes increasingly difficult with the rougher surfaces (see B in Fig. 22). It would be an interesting future direction to study, how adept the HVS is to estimate the contribution of the subsurface scattering when surface scattering is high, or when specular highlights are superimposed on the rendering of a rough object.

6 LIMITATIONS AND FUTURE WORK

This work has been the first attempt to explore how subsurface scattering contributes to apparent gloss. The materials addressed in this study represent a tiny subset of all possible materials that can exist around us. In order to keep the number of experimental stimuli within the manageable limits, we had to fix multiple intrinsic and extrinsic parameters, which also implies that our findings come with particular limitations which need to be addressed in future works:

- We used isotropic phase function and wavelength-independent σ_t and albedo. Subsurface scattering in most real materials has large spectral and spatial variation. Materials with wavelength-dependent subsurface scattering (chromatic effects) and non-isotropic phase functions should be studied in the future. The phase function has been shown to be important for material appearance [24]. The authors provided 2-dimensional perceptual embedding of the phase functions, where the dimensions modulate diffuse translucent and sharp, detailed, glass-like appearances, respectively. We hypothesize that the latter could be correlated with apparent gloss.
- While the index of refraction has been fixed to 1.5 in our experiments, we believe other indices of refraction also deserve attention in the future.
- We used Beckmann microfacet normal distribution to modulate surface scattering parameter. It is interesting to explore, whether our findings hold if the surface roughness is modeled with other distributions, such as Phong [30] or GGX [67]. We hypothesize that the impact will be negligible, as the clustering of a large pool of parameters will converge to relatively similar appearances. However, this needs further study and experimental evidence.
- Although we plot Z-scores as a function of σ_t and albedo, the effects of the two parameters need to be studied separately and more in depth. The future experiments could include comparisons for each σ_t and albedo, separately. It is also important to explore the potential interaction between these two parameters. We believe that there is a significant interaction between the effects of the two parameters. For example, the impact of albedo can be large for high σ_t , but it becomes negligible when σ_t is very low. We believe a mixed effects statistical model is needed to describe the correlation between gloss and subsurface light transport, while σ_t , albedo and *alpha* can be treated as fixed effects, random effects, such as user physiological and display characteristics, should be also included.
- Illumination conditions have been fixed throughout the experiment. It has been shown before that illumination geometry affects both translucency [14, 73] and gloss [15, 48]. Therefore, the study should be extended to other illumination geometries.
- As a metric for clustering, Euclidean distance could be substituted with more perception-aware metrics, such as L^4 -norm [50], the cubic root metric used by Gkioulekas *et al.* [24] or the appearance similarity metric proposed by Lagunas *et al.* [34]. Additionally, the perceptual accuracy could be improved if the comparisons were done in the CIELAB space instead of RGB [50, 61]. However, using RGB usually biases chromatic information [50, 61]. As our stimuli have been mostly achromatic, we believe the comparison in the RGB space has not introduced any significant bias in the clustering process.

Besides, addressing the research question from the perspective of image-based measurements has been beyond the scope of this work. However, we believe that future works should investigate how subsurface light transport affects image structure and statistics which proposedly are glossiness cues. This could bring to light *why* and *how* subsurface scattering contributes to apparent gloss.

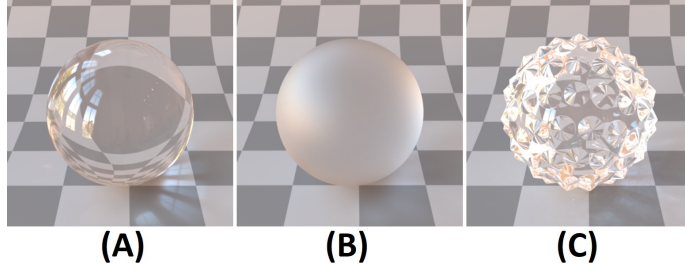


Fig. 22. The structure of the image A provides more cues on how to segment reflection and transmission components, while the task looks considerably more difficult for images B and C.

First of all, our results once again illustrate that no one-to-one correspondence between physical and perceptual properties exists and that our ability to segment specular reflections from image structure is limited [40]. This is why users might have mistaken caustics for specular reflections. The image-level intensities result from a combination of reflection and transmission. Unmixing those is an ill-posed problem and the HVS uses different constraints for this task, such as, apparent object shape [40]. While smooth spherical and cylindrical shapes facilitate separation of specular and non-specular components, the task becomes increasingly difficult for complex geometries. For instance, in Fig. 22, it is easier to separate reflection and transmission components in image A than it is for images B and C. We hypothesize that additional factors that usually facilitate this segmentation, such as motion, binocular vision or surface texture [9, 56, 69] could decrease the impact of subsurface scattering on apparent gloss. This could explain why many users tied all physical objects in previous works when interaction was permitted [19, 21].

On the other hand, the users still saw a glossiness difference, even when segmenting specular and non-specular components should have been relatively simple - particularly, in case of smooth spherical objects. We believe this happened because apparent gloss is not a function of apparent specular reflection only, but it also depends on extrinsic factors that are independent from specular reflections, such as lightness of the non-specular areas [49].

It remains an open question exactly which image cues and which psycho-visual mechanisms of gloss perception are affected by the subsurface scattering, and rigorous future work is needed to answer it. Similarly to Marlow and Anderson [38], psychophysical studies should be conducted in the future to measure how different image-level measurements, such as perceived coverage, sharpness and contrast of the highlights co-vary with the perceived glossiness of the materials of different shapes and light transport properties. This will help us understand the differences observed in this paper, and the robustness of the state-of-the-art will also be tested in the context of light-transmissive media. Moreover, particular image statistics should be studied to quantify and model the impact of subsurface scattering on the gloss cues in the image space. Additional interviews with the subjects could potentially help with the identification of the most salient cues and interpreting the results. Particularly, eye tracking experiments in the controlled conditions could provide deeper insight into the actual image cues used for glossiness assessment. And last but not least, we believe that perceived gloss is *at least* two dimensional - distinctness and contrast, as proposed by Pellacini et al. [49], being the major perceptual dimensions of gloss, even for translucent objects. However, the model quantifying these perceptual dimensions should include σ_t and albedo along with other physical parameters, to enable accurate placement of the translucent stimuli in the perceptual gloss space. We have observed in Experiment 1 that for high σ_t , when the light does not penetrate deep into the volume, the processes and findings are phenomenologically similar to Ward's model used by Pellacini et al. [49]. Multidimensional scaling analysis (MDS) similar to [49] could reveal how σ_t and albedo

contribute to distinctness and contrast, given that the stimuli are sampled densely enough in σ_t -albedo space. With that being said, we believe a separate embedding might be needed for each *alpha*, as the HVS might apply different internal perceptual functions to the stimuli with different roughnesses (i.e. with different gloss cues).

Our findings have practical implications for computer graphics, perception, as well as material appearance measurement and reproduction research. They show that material appearance modelling should be done on the shape we are particularly interested in and generalization of the findings based on one shape or surface roughness should be taken with extreme care. We also propose that future gloss perception research should include materials that permit subsurface light transport and the perceptual models of gloss should be updated so that they could account for potential contribution from subsurface scattering. Finally, gloss measurement protocols should accommodate translucent materials.

7 CONCLUSION

We have conducted psychophysical experiments to test whether subsurface scattering of light contributes to gloss perception and to characterize this impact qualitatively and quantitatively. The results support our hypothesis and provide ample evidence that gloss perception is impacted by subsurface scattering. The impact varies across shapes and surface roughness levels; this we believe is the result of different low- and high-level image cues being used (by the HVS) for different shapes to assess gloss. Our findings propose that modelling appearance should be taken with care and findings should not be generalized to other shapes and surface scattering models. Moreover, the state-of-the-art findings based on fully opaque materials might not be valid for transparent and translucent media. Understanding *why* subsurface light transport contributes to apparent gloss and *how* it is used by the HVS would be an important future direction. Eventually, a higher number of stimuli (ideally in HDR) will be needed to build a complete perceptual space of gloss. We believe the future work addressing gloss perception should not be limited to fully opaque materials and the perceptual models should account for subsurface scattering. Rigorous work is needed in the future to identify the exact mechanisms for predicting perceptual gloss from materials' surface and subsurface light transport properties.

8 ACKNOWLEDGMENTS

This work was supported in part by NSF grant IIS-2007283. The work was also supported by MUVApp (#250293) and MANER (#288187) projects of the Research Council of Norway.

REFERENCES

- [1] 2017. ASTM E284-17 Standard Terminology of Appearance. ASTM International, West Conshohocken, PA. <https://doi.org/10.1520/E0284-17>
- [2] Edward H Adelson. 2001. On seeing stuff: the perception of materials by humans and machines. In *Human vision and electronic imaging VI*, Vol. 4299. International Society for Optics and Photonics, 1–12.
- [3] Barton L Anderson and Juno Kim. 2009. Image statistics do not explain the perception of gloss and lightness. *Journal of Vision* 9, 11:10 (2009), 1–17.
- [4] Jacob Beck and Slava Prazdny. 1981. Highlights and the perception of glossiness. *Perception & Psychophysics* (1981).
- [5] Alice C Chadwick, George Cox, Hannah E Smithson, and Robert W Kentridge. 2018. Beyond scattering and absorption: Perceptual unmixing of translucent liquids. *Journal of vision* 18, 11:18 (2018), 1–15.
- [6] Nahian S Chowdhury, Phillip J Marlow, and Juno Kim. 2017. Translucency and the perception of shape. *Journal of vision* 17, 3:17 (2017), 1–14.
- [7] Jacob Cohen. 1988. *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Erlbaum.
- [8] Alireza Dastan. 2020. Gaussian and mean curvatures calculation on a triangulated 3D surface. <https://www.mathworks.com/matlabcentral/fileexchange/61136-gaussian-and-mean-curvatures-calculation-on-a-triangulated-3d-surface>. Retrieved: 2020-07-30 from MATLAB Central File Exchange.
- [9] Katja Doerschner, Roland W Fleming, Ozgur Yilmaz, Paul R Schrater, Bruce Hartung, and Daniel Kersten. 2011. Visual motion and the perception of surface material. *Current Biology* 21, 23 (2011), 2010–2016.
- [10] Peter G Engeldrum. 2000. *Psychometric scaling: a toolkit for imaging systems development*. Imcotek.

- [11] Christian Eugène. 2008. Measurement of "total visual appearance": a CIE challenge of soft metrology. In *12th IMEKO TC1 & TC7 Joint Symposium on Man, Science & Measurement*. 61–65.
- [12] Franz Faul. 2019. The influence of Fresnel effects on gloss perception. *Journal of Vision* 19, 13:1 (2019), 1–39.
- [13] Roland W Fleming. 2014. Visual perception of materials and their properties. *Vision research* 94 (2014), 62–75.
- [14] Roland W Fleming and Heinrich H Bülthoff. 2005. Low-level image cues in the perception of translucent materials. *ACM Transactions on Applied Perception (TAP)* 2, 3 (2005), 346–382.
- [15] Roland W Fleming, Ron O Dror, and Edward H Adelson. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision* 3 (2003), 347–368.
- [16] Brendan J Frey and Delbert Dueck. 2007. Clustering by passing messages between data points. *Science* 315, 5814 (2007), 972–976.
- [17] Davit Gigilashvili, Fereshteh Mirjalili, and Jon Yngve Hardeberg. 2019. Illuminance Impacts Opacity Perception of Textile Materials. In *Color and Imaging Conference*. Society for Imaging Science and Technology, 126–131.
- [18] Davit Gigilashvili, Midori Tanaka, Marius Pedersen, and Jon Yngve Hardeberg. 2020. Image Statistics as Glossiness and Translucency Predictor in Photographs of Real-world Objects. In *10th Colour and Visual Computing Symposium (CVCS)*. CEUR Workshop Proceedings, 15 pages.
- [19] Davit Gigilashvili, Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Marius Pedersen. 2018. Behavioral investigation of visual appearance assessment. In *Color and Imaging Conference*. Society for Imaging Science and Technology, 294–299.
- [20] Davit Gigilashvili, Jean-Baptiste Thomas, Marius Pedersen, and Jon Yngve Hardeberg. 2019. Material appearance: ordering and clustering. In *Material Appearance 2019, IS&T International Symposium on Electronic Imaging*. Society for Imaging Science and Technology, 202:1–202:6.
- [21] Davit Gigilashvili, Jean-Baptiste Thomas, Marius Pedersen, and Jon Yngve Hardeberg. 2019. Perceived Glossiness: Beyond Surface Properties. In *Color and Imaging Conference*. Society for Imaging Science and Technology, 37–42.
- [22] Davit Gigilashvili, Philipp Urban, Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Marius Pedersen. 2019. Impact of Shape on Apparent Translucency Differences. In *Color and Imaging Conference*. Society for Imaging Science and Technology, 132–137.
- [23] Ioannis Gkioulekas, Bruce Walter, Edward H Adelson, Kavita Bala, and Todd Zickler. 2015. On the appearance of translucent edges. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5528–5536.
- [24] Ioannis Gkioulekas, Bei Xiao, Shuang Zhao, Edward H Adelson, Todd Zickler, and Kavita Bala. 2013. Understanding the role of phase function in translucent appearance. *ACM Transactions on graphics (TOG)* 32, 5 (2013), 1–19.
- [25] Phil J. Green. 2003. A Colour Engineering Toolbox. <http://www.color.org/resources/ColourEngineeringToolbox.zip>. Accessed on 25.08.2020.
- [26] Dar'ya Guarnera, Giuseppe Claudio Guarnera, Matteo Toscani, Mashhuda Glencross, Baihua Li, Jon Yngve Hardeberg, and Karl R Gegenfurtner. 2018. Perceptually validated analytical BRDFs parameters remapping. In *ACM SIGGRAPH 2018 Talks*. 1–2.
- [27] Yun-Xian Ho, Michael S Landy, and Laurence T Maloney. 2008. Conjoint measurement of gloss and surface texture. *Psychological Science* 19, 2 (2008), 196–204.
- [28] Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), 65–70.
- [29] Richard S Hunter. 1937. Methods of determining gloss. *NBS Research paper RP 958* (1937), 19–39.
- [30] Wenzel Jakob. 2010. Mitsuba renderer. <http://www.mitsuba-renderer.org>.
- [31] Iona S Kerrigan and Wendy J Adams. 2013. Highlights, disparity, and perceived gloss with convex and concave surfaces. *Journal of Vision* 13, 1:9 (2013), 1–10.
- [32] Juno Kim, Phillip Marlow, and Barton L Anderson. 2011. The perception of gloss depends on highlight congruence with surface shading. *Journal of Vision* 11(9), 4 (2011), 1–19.
- [33] Stanford University Computer Graphics Laboratory. 1994. The Stanford 3D Scanning Repository. <http://graphics.stanford.edu/data/3Dscanrep/>.
- [34] Manuel Lagunas, Sandra Malpica, Ana Serrano, Elena Garces, Diego Gutierrez, and Belen Masia. 2019. A similarity measure for material appearance. *arXiv preprint arXiv:1905.01562* (2019), 12 pages.
- [35] Michael S Landy. 2007. A gloss on surface properties. *Nature* 447, 7141 (2007), 158–159.
- [36] Frédéric B Leloup, Gael Obein, Michael R Pointer, and Peter Hanselaer. 2014. Toward the soft metrology of surface gloss: A review. *Color Research & Application* 39, 6 (2014), 559–570.
- [37] Phillip Marlow, Juno Kim, and Barton L Anderson. 2011. The role of brightness and orientation congruence in the perception of surface gloss. *Journal of Vision* 11(9), 16 (2011), 1–12.
- [38] Phillip J Marlow and Barton L Anderson. 2013. Generative constraints on image cues for perceived gloss. *Journal of Vision* 13, 14:2 (2013), 1–23.
- [39] Phillip J Marlow, Juno Kim, and Barton L Anderson. 2012. The perception and misperception of specular surface reflectance. *Current Biology* 22, 20 (2012), 1909–1913.
- [40] Phillip J Marlow, Dejan Todorović, and Barton L Anderson. 2015. Coupled computations of three-dimensional shape and material. *Current Biology* 25, 6 (2015), R221–R222.
- [41] Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H Barr. 2003. Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and mathematics III*. Springer, 35–57.
- [42] mfa Boston CAMEO. 2020. Paraffin Wax. http://cameo.mfa.org/wiki/Paraffin_wax. Accessed on 25.08.2020.
- [43] Isamu Motoyoshi. 2010. Highlight–shading relationship as a cue for the perception of translucent and transparent materials. *Journal of Vision* 10, 9:6 (2010), 1–11.

- [44] Isamu Motoyoshi, Shin'ya Nishida, Lavanya Sharan, and Edward H Adelson. 2007. Image statistics and the perception of surface qualities. *Nature* 447, 7141 (2007), 206–209.
- [45] Takehiro Nagai, Yuki Ono, Yusuke Tani, Kowa Koida, Michiteru Kitazaki, and Shigeki Nakauchi. 2013. Image regions contributing to perceptual translucency: A psychophysical reverse-correlation study. *i-Perception* 4, 6 (2013), 407–428.
- [46] Shin'ya Nishida, Isamu Motoyoshi, Lisa Nakano, Yuanzhen Li, Lavanya Sharan, and Edward Adelson. 2008. Do colored highlights look like highlights? *Journal of Vision* 8, 6 (2008), 339.
- [47] Gaël Obein, Kenneth Knoblauch, and Françoise Viénot. 2004. Difference scaling of gloss: Nonlinearity, binocularity, and constancy. *Journal of vision* 4 (2004), 711–720.
- [48] Maria Olkkonen and David H Brainard. 2011. Joint effects of illumination geometry and object shape in the perception of surface reflectance. *i-Perception* 2, 9 (2011), 1014–1034.
- [49] Fabio Pellacini, James A Ferwerda, and Donald P Greenberg. 2000. Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 55–64.
- [50] Thiago Pereira and Szymon Rusinkiewicz. 2012. Gamut mapping spatially varying reflectance with an improved BRDF similarity metric. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 1557–1566.
- [51] Zygmunt Pizlo. 2001. Perception viewed as an inverse problem. *Vision research* 41, 24 (2001), 3145–3161.
- [52] Lin Qi, Mike J Chantler, J Paul Siebert, and Junyu Dong. 2012. How mesoscale and microscale roughness affect perceived gloss. In *Predicting Perceptions: Proceedings of the 3rd International Conference on Appearance*. Lulu Press, Inc. Edinburgh, Scotland, 48–51.
- [53] Lin Qi, Mike J Chantler, J Paul Siebert, and Junyu Dong. 2014. Why do rough surfaces appear glossy? *Journal of the Optical Society of America A* 31, 5 (2014), 935–943.
- [54] Lin Qi, Mike J Chantler, J Paul Siebert, and Junyu Dong. 2015. The joint effect of mesoscale and microscale roughness on perceived gloss. *Vision research* 115 (2015), 209–217.
- [55] Ralph L Rosnow and Robert Rosenthal. 2003. Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 57, 3 (2003), 221.
- [56] Yuichi Sakano and Hiroshi Ando. 2010. Effects of head motion and stereo viewing on perceived glossiness. *Journal of Vision* 10, 9:15 (2010), 1–14.
- [57] Masataka Sawayama, Yoshinori Dobashi, Makoto Okabe, Kenchi Hosokawa, Takuya Koumura, Toni Saarela, Maria Olkkonen, and Shin'ya Nishida. 2019. Visual discrimination of optical material properties: a large-scale study. *BioRxiv* (2019), 800870.
- [58] Alexandra C Schmid, Pascal Barla, and Katja Doerschner. 2020. Material category determined by specular reflection structure mediates the processing of image features for perceived gloss. *bioRxiv* (2020). bioRxiv preprint DOI: <https://doi.org/10.1101/2019.12.31.892083>.
- [59] Scientific Polymer Products, Inc. 2020. Refractive Index of Polymers. <https://scientificpolymer.com/technical-library/refractive-index-of-polymers-by-index/>. Accessed on 25.08.2020.
- [60] Lavanya Sharan, Ruth Rosenholtz, and Edward H Adelson. 2014. Accuracy and speed of material categorization in real-world images. *Journal of Vision* 14, 9:12 (2014), 1–24.
- [61] Tiancheng Sun, Ana Serrano, Diego Gutierrez, and Belen Masia. 2017. Attribute-preserving gamut mapping of measured BRDFs. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 47–54.
- [62] Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Gabriele Simone. 2017. Image contrast measure as a gloss material descriptor. In *International Workshop on Computational Color Imaging*. Springer, 233–245.
- [63] Louis L Thurstone. 1927. A law of comparative judgment. *Psychological review* 34, 4 (1927), 273–286.
- [64] Matteo Toscani, Dar'ya Guarnera, Giuseppe Claudio Guarnera, Jon Yngve Hardeberg, and Karl R Gegenfurtner. 2020. Three perceptual dimensions for specular and diffuse reflection. *ACM Transactions on Applied Perception (TAP)* 17, 2 (2020), 1–26.
- [65] Kristi Tsukida and Maya R Gupta. 2011. *How to analyze paired comparison data*. Technical Report. University of Washington, Department of Electrical Engineering, Seattle, WA.
- [66] Peter Vangorp, Jurgen Laurijssen, and Philip Dutré. 2007. The influence of shape on the perception of material reflectance. In *ACM Transactions on graphics (TOG)*, Vol. 26. ACM, 77:1–77:10.
- [67] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. 2007. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*. Eurographics Association, 195–206.
- [68] Gregory J Ward. 1992. Measuring and modeling anisotropic reflection. In *Proceedings of the 19th annual conference on Computer graphics and interactive techniques*. 265–272.
- [69] Gunnar Wendt, Franz Faul, Vebjørn Ekroll, and Rainer Mausfeld. 2010. Disparity, motion, and color information improve gloss constancy performance. *Journal of vision* 10, 9:7 (2010), 1–17.
- [70] Maarten WA Wijntjes and Sylvia C Pont. 2010. Illusory gloss on Lambertian surfaces. *Journal of Vision* 10, 9:13 (2010), 1–12.
- [71] Josh Wills, Sameer Agarwal, David Kriegman, and Serge Belongie. 2009. Toward a perceptual space for gloss. *ACM Transactions on Graphics (TOG)* 28, 4 (2009), 1–15.
- [72] Bei Xiao and David H Brainard. 2008. Surface gloss and color perception of 3D objects. *Visual neuroscience* 25, 3 (2008), 371–385.
- [73] Bei Xiao, Bruce Walter, Ioannis Gkioulekas, Todd Zickler, Edward Adelson, and Kavita Bala. 2014. Looking against the light: How perception of translucency depends on lighting direction. *Journal of Vision* 14, 3:17 (2014), 1–22.

- [74] Bei Xiao, Shuang Zhao, Ioannis Gkioulekas, Wenyan Bi, and Kavita Bala. 2019. Effect of Geometric Sharpness on Translucent Material Perception. *bioRxiv* (2019), 795294.

9 SUPPLEMENTARY MATERIALS

9.1 Stimuli

Figures 23-27 illustrate all stimuli used in the experiments.

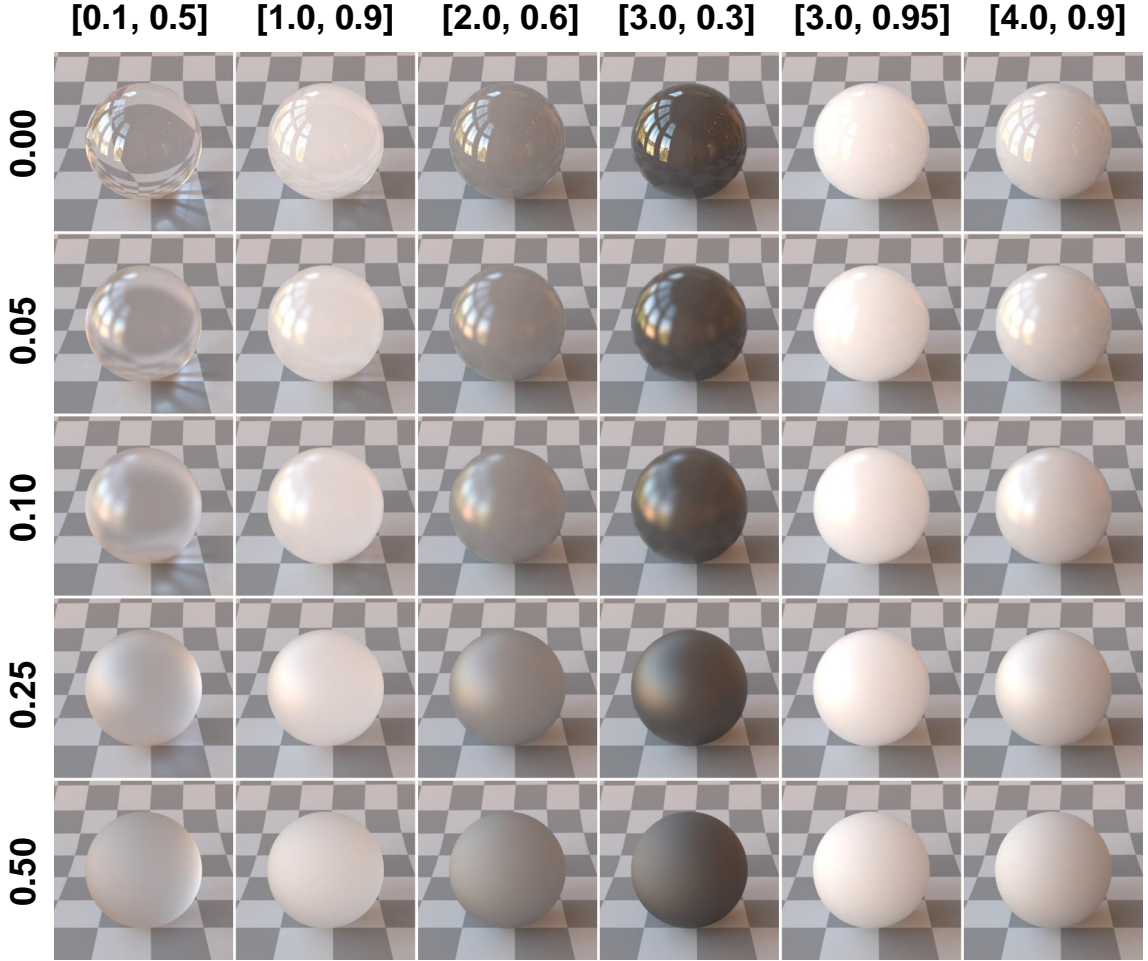


Fig. 23. All spherical objects used in the experiments. Columns correspond to σ_t and albedo, while rows correspond to different *alphas*, respectively.

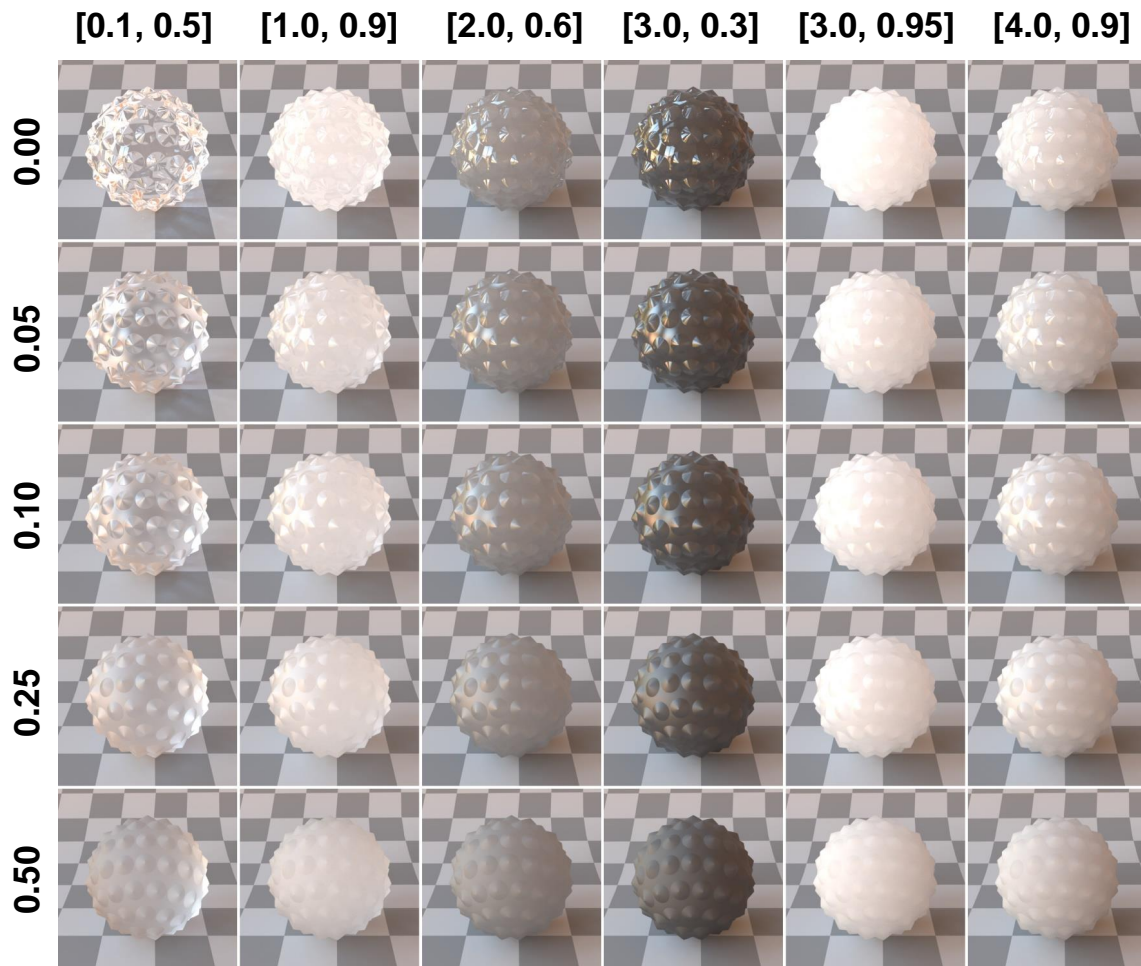


Fig. 24. All spiky sphere objects used in the experiments. Columns correspond to σ_t and albedo, while rows correspond to different α values, respectively.

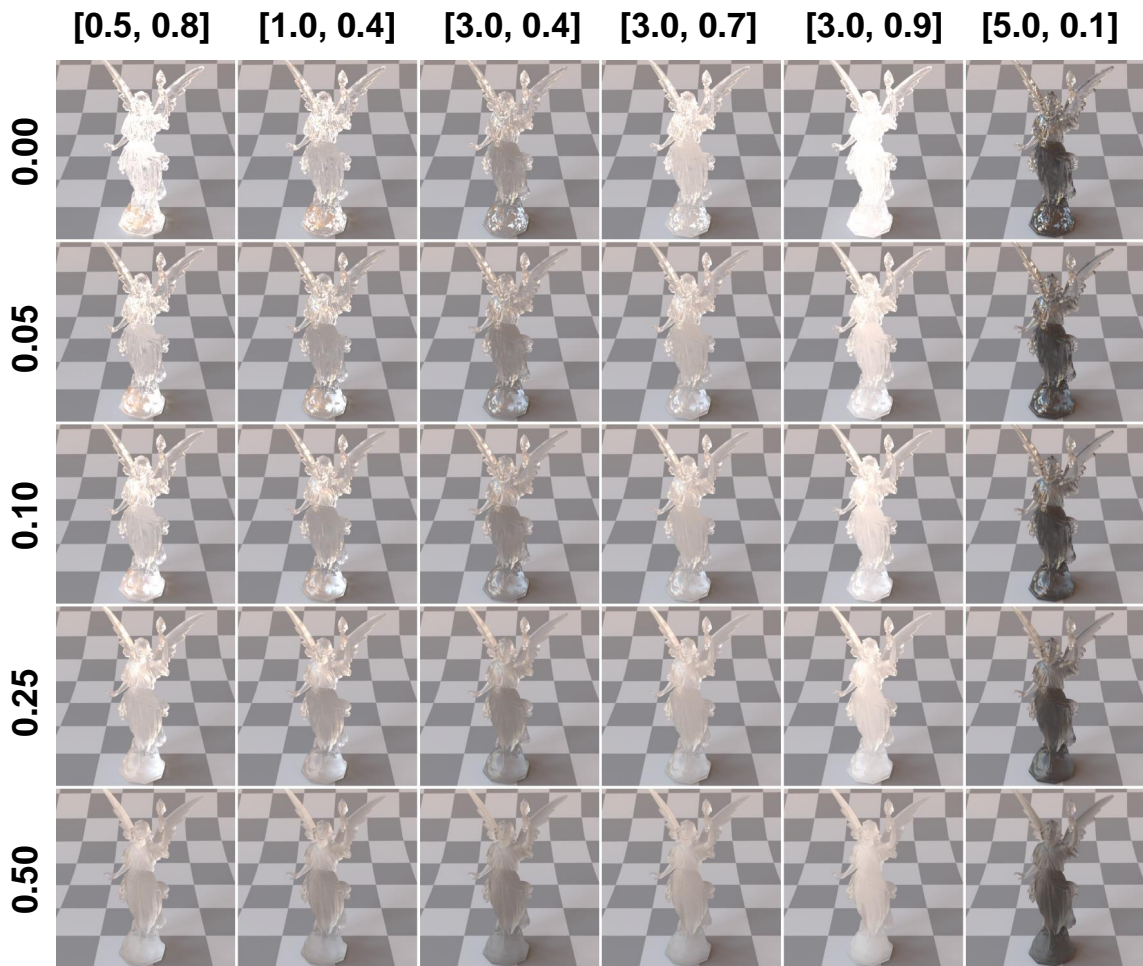


Fig. 25. All Stanford Lucy objects used in the experiments. Columns correspond to σ_t and albedo, while rows correspond to different α s, respectively.

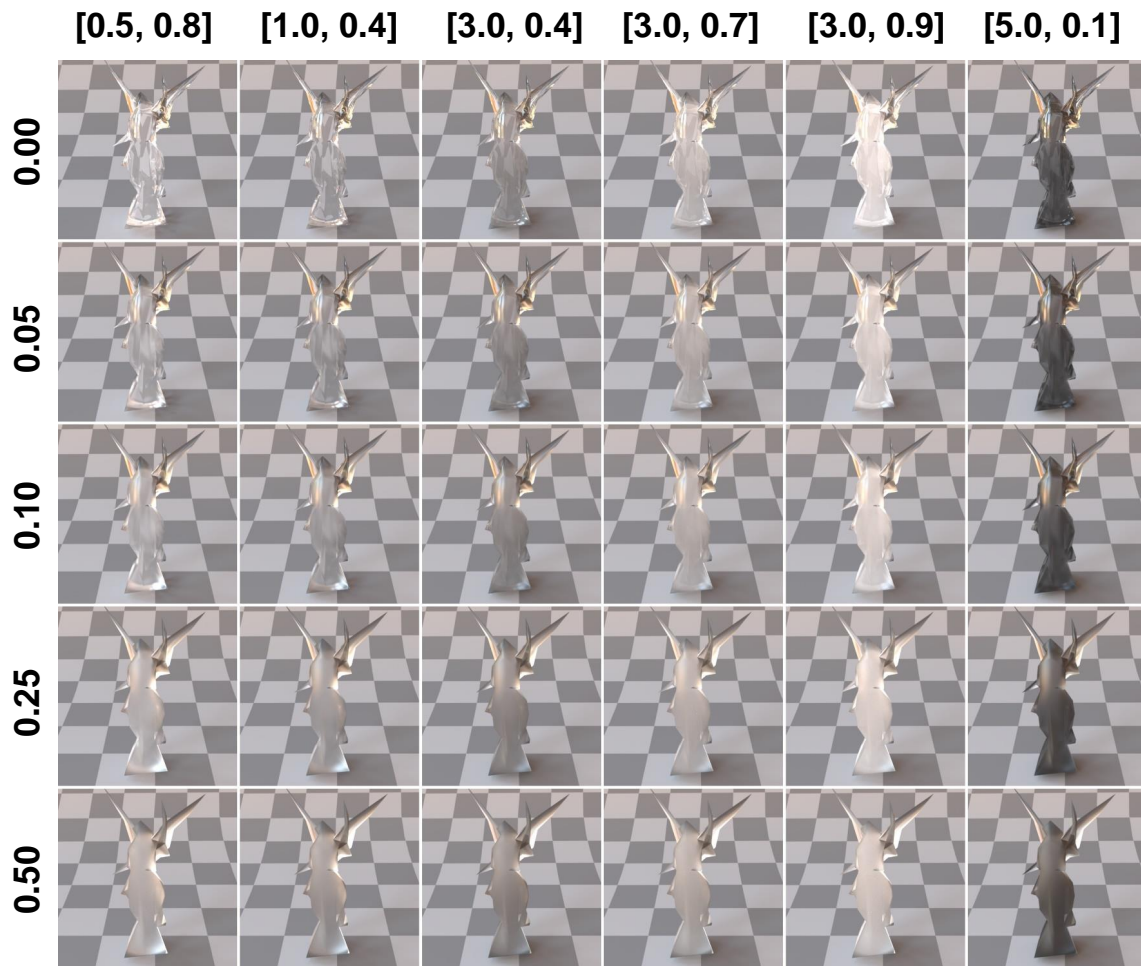


Fig. 26. All low resolution Lucy objects used in the experiments. Columns correspond to σ_t and albedo, while rows correspond to different *alphas*, respectively.

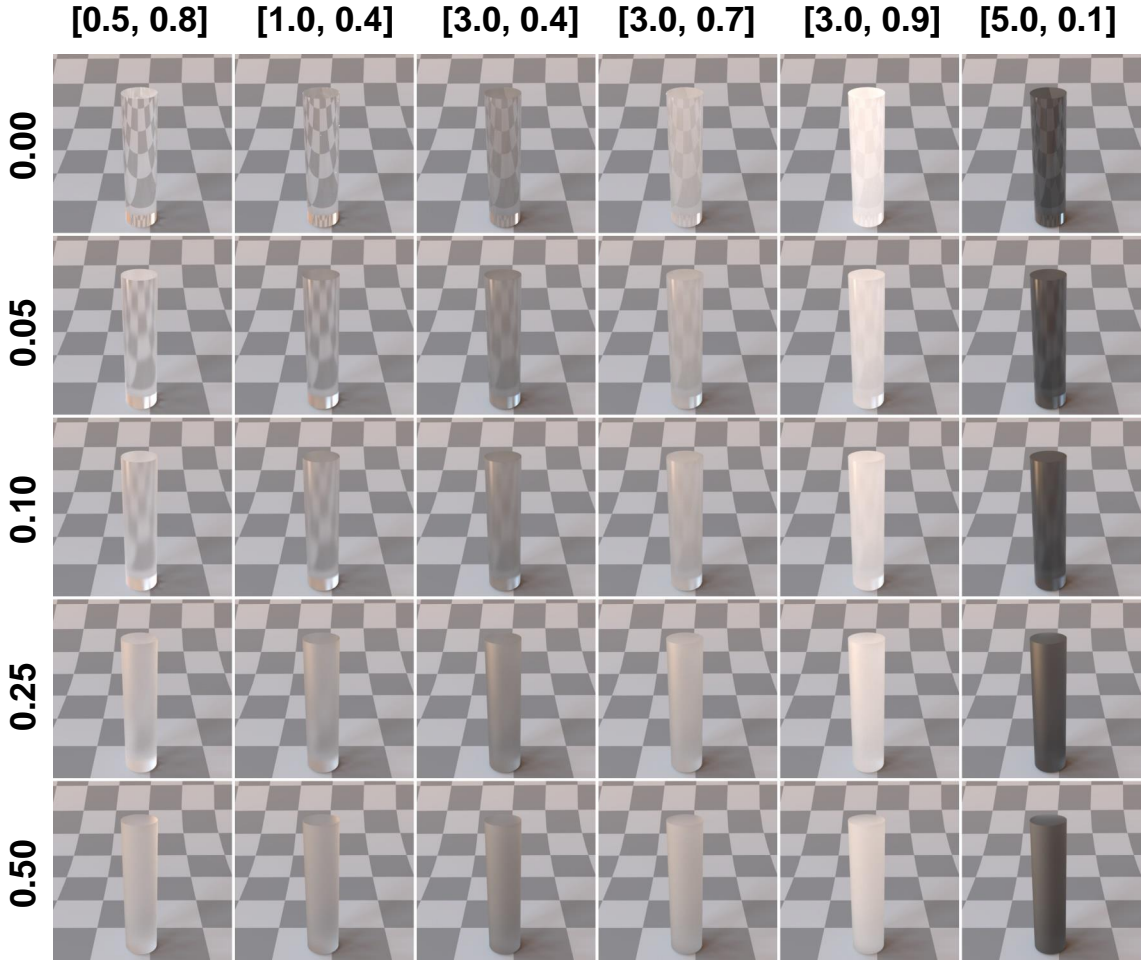


Fig. 27. All cylindrical objects used in the experiments. Columns correspond to σ_t and albedo, while rows correspond to different α s, respectively.

9.2 Z-scores

Z-score plots have been reported for spherical and the Lucy shapes in the main body of the manuscript. Below Z-scores are reported for three remaining shapes: spiky sphere, low-resolution Lucy and a cylinder, in Fig. 28-30, respectively. The overall trends between a sphere and spiky sphere, as well as between Lucy and low-resolution Lucy are largely similar. However, there are still some subtle differences. For instance, a low albedo spherical object is among the glossiest for lower α s. However, that is not true for a spiky sphere. The absence of the mirror-like reflection of the environment in spiky spheres due to surface curvature could be the explanation for this fact. There is a difference between Lucy and low-resolution Lucy as well. For low α s, the transparent low σ_t low-resolution Lucy is the glossiest one, while that is not the case for the Lucy made of the identical material. The transparent low-resolution Lucy, similarly to a sphere, permits seeing the background through the object, while that is not possible for Lucy due to high surface curvature. This can be an indication that background plays a role in gloss perception and this similarity between a sphere and the low-resolution Lucy should be scrutinized in the future studies. Finally, the impact of subsurface scattering on the glossiness of cylindrical objects is minimal in comparison with spherical and Lucy-shaped objects.

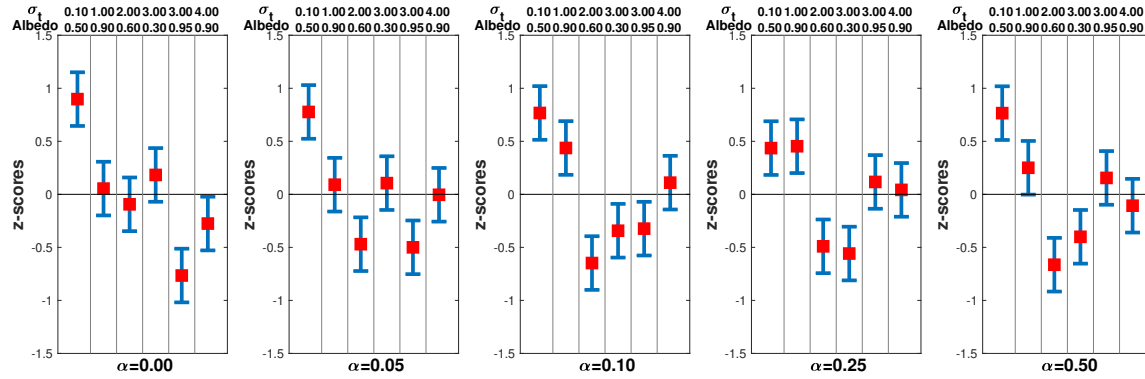


Fig. 28. Results for the spiky sphere. A red cube corresponds to the mean Z-score for a given object, while the error bar corresponds to 95% confidence interval. The difference among Z-scores diminishes with the increase in roughness.

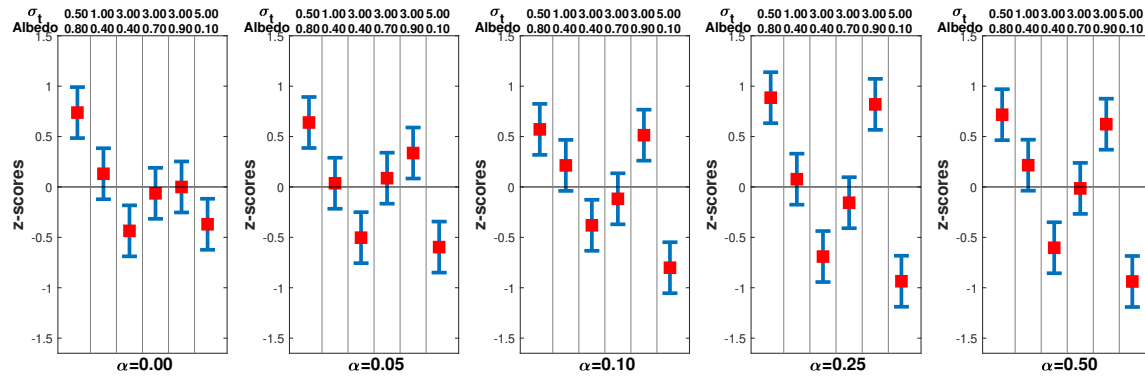


Fig. 29. Results for low-resolution Lucy. The difference among Z-scores climbs with the increase in roughness.

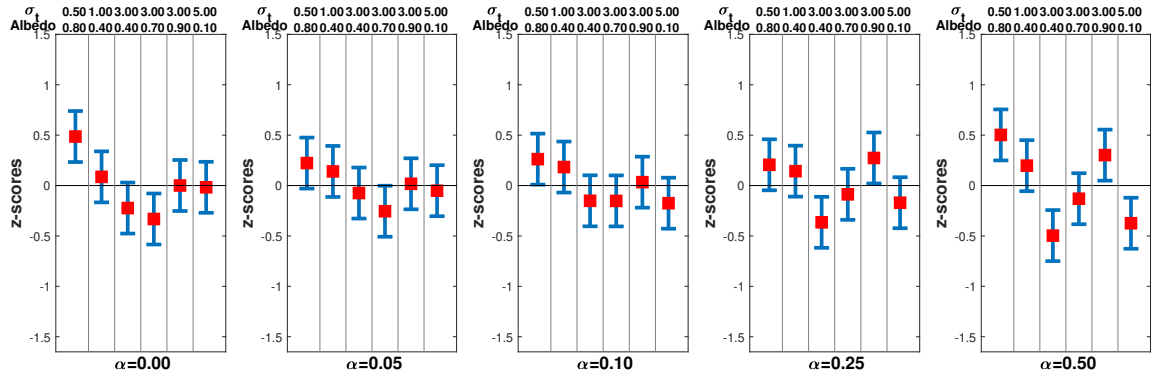


Fig. 30. Z-scores for a cylinder. Subsurface scattering has a subtle effect on the glossiness of cylinders and the majority of the objects with equal alpha looks equally glossy, except for very high alphas.

9.3 Z-scores in σ_t -albedo space

The variation of the Z-scores in σ_t -albedo space shows how these individual parameters affect gloss (Fig. 31-33). The trends are similar between a sphere and a spiky sphere, as well as between the Lucy and low-resolution version of it. For spiky sphere, the larger circles are concentrated on the lower end of the σ_t axis, meaning that materials with lower σ_t usually look glossier. The tendency of black and brown circles in Fig. 31 manifests how the negative correlation between albedo and Z-scores gradually changes into a positive one as the α increases. In the case of low-resolution Lucy, all large-diameter circles are concentrated in the high albedo part of the space (Fig. 32), while the diameters do not differ largely in the case of cylinders (Fig. 33), being less affected by σ_t -albedo variations.

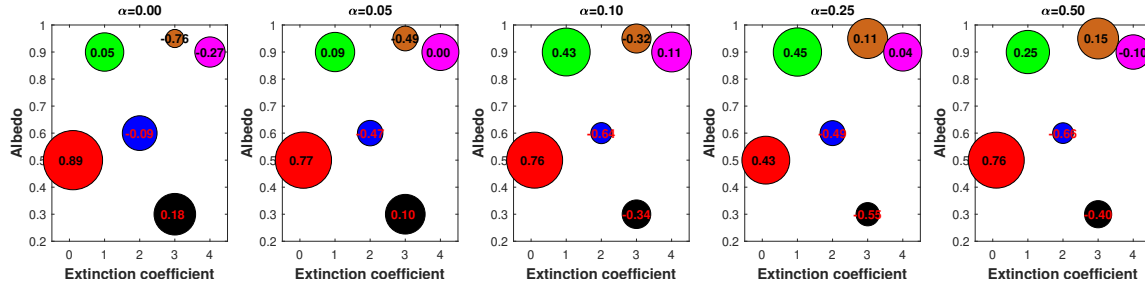


Fig. 31. Results for the spiky sphere. Larger circle diameters represent a higher mean Z-score. Similarly to a sphere, lower albedo and σ_t lead to a glossier look for smoother objects, while the trend changes as the roughness increases. Note that Z-scores are relative to the objects of the same roughness and circles of the same color are not directly comparable among the five plots.

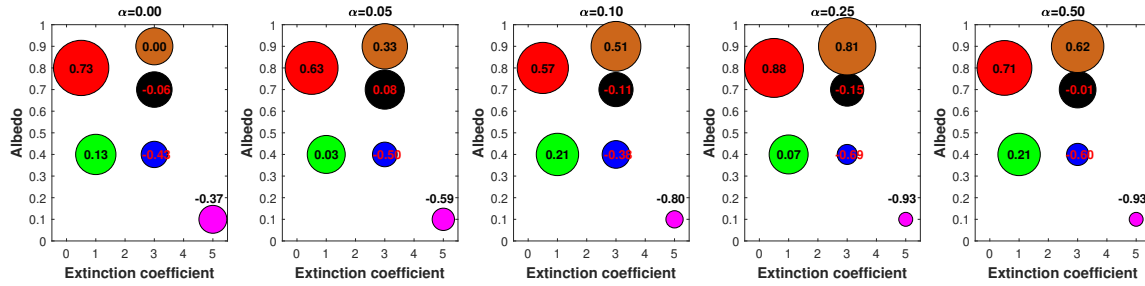


Fig. 32. Results for low-resolution Lucy. Similarly to Lucy, high albedo usually yields glossier stimuli.

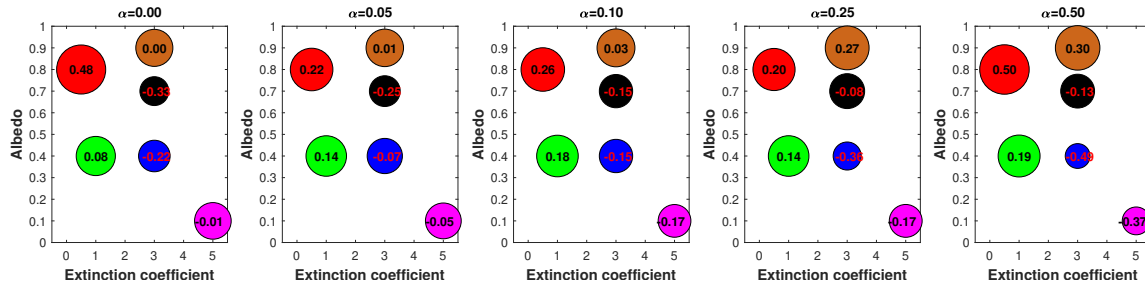


Fig. 33. Results for cylinder.

9.4 Z-scores as a function of σ_t and albedo

Z-scores are shown as a function of σ_t and albedo in Fig. 34-36. In the case of a spiky sphere (Fig. 34), the negative correlation can be seen between Z-scores and both physical parameters, only when *alpha* is low. In the plots of low-resolution Lucy (Fig. 35), similarly to Lucy, the negative correlation can be seen with the extinction coefficient, and the positive correlation is apparent with an albedo that becomes even stronger as the *alpha* increases. Being consistent with the prior reasoning, neither parameters affect the Z-scores of a cylinder.

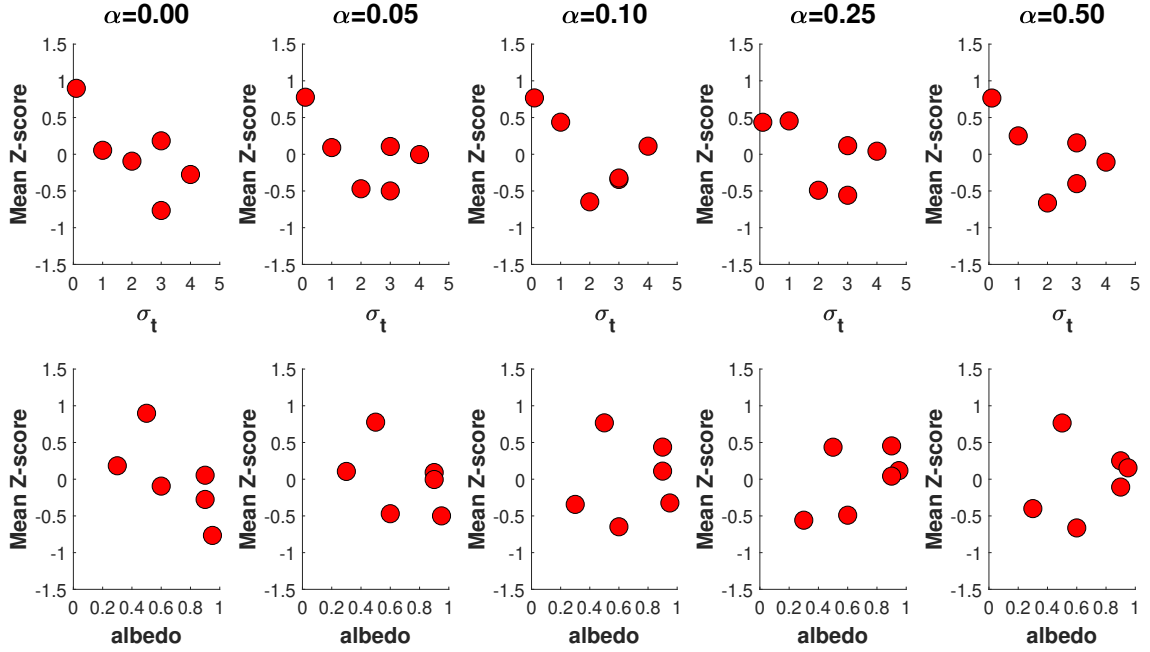


Fig. 34. Z-score as a function of extinction coefficient (top row) and albedo (bottom) for the spiky sphere.

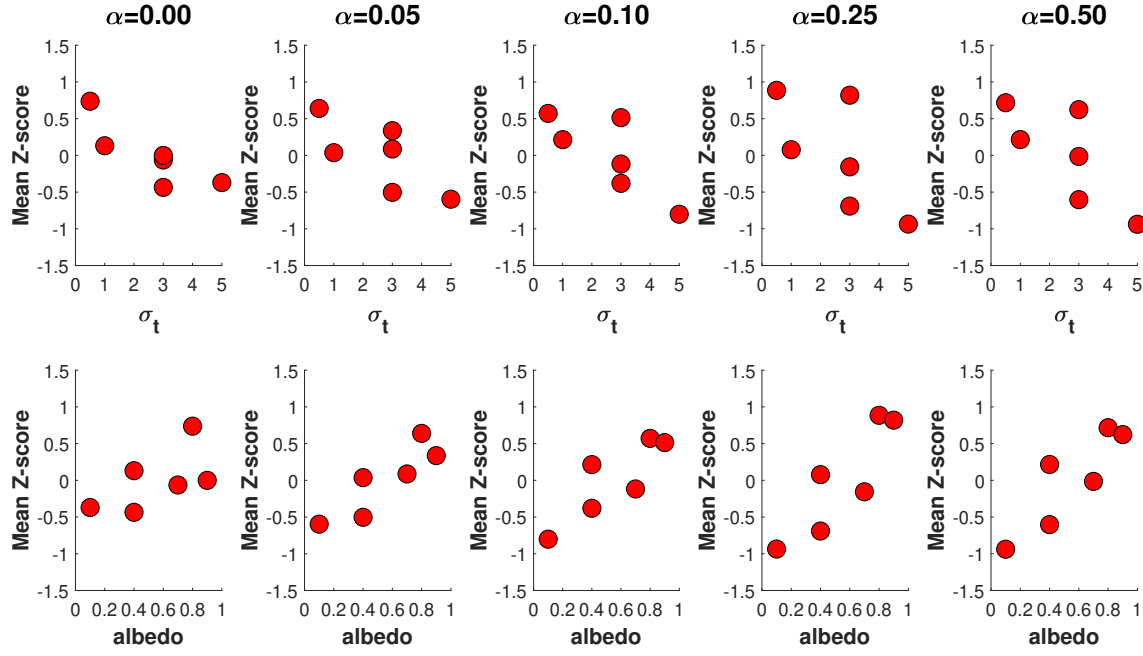


Fig. 35. Z-score as a function of extinction coefficient (top row) and albedo (bottom) for low-resolution Lucy.

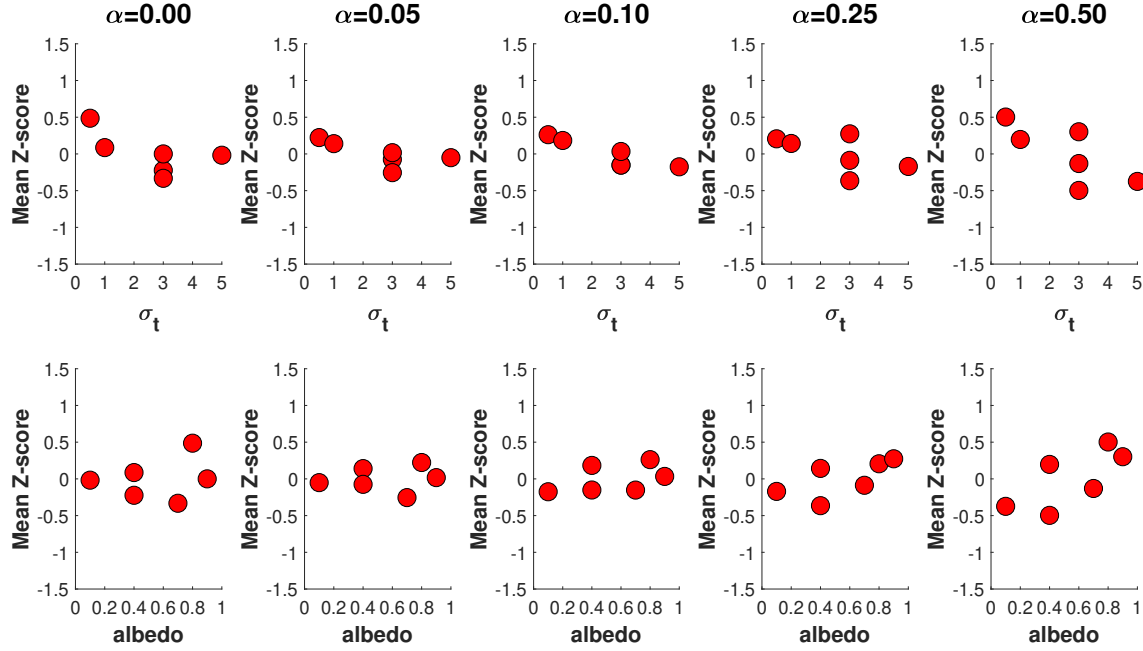


Fig. 36. Z-score as a function of extinction coefficient (top row) and albedo (bottom) for a cylinder.

9.5 Multidimensional scaling

An additional experiment has been conducted using 15 smooth-surfaced spherical stimuli sampled in σ_t -albedo space. Interesting proximity between transparent low σ_t and dark opaque (high σ_t and low albedo) objects has been observed in 2-dimensional, 3-dimensional, as well as in the higher dimensional embeddings.

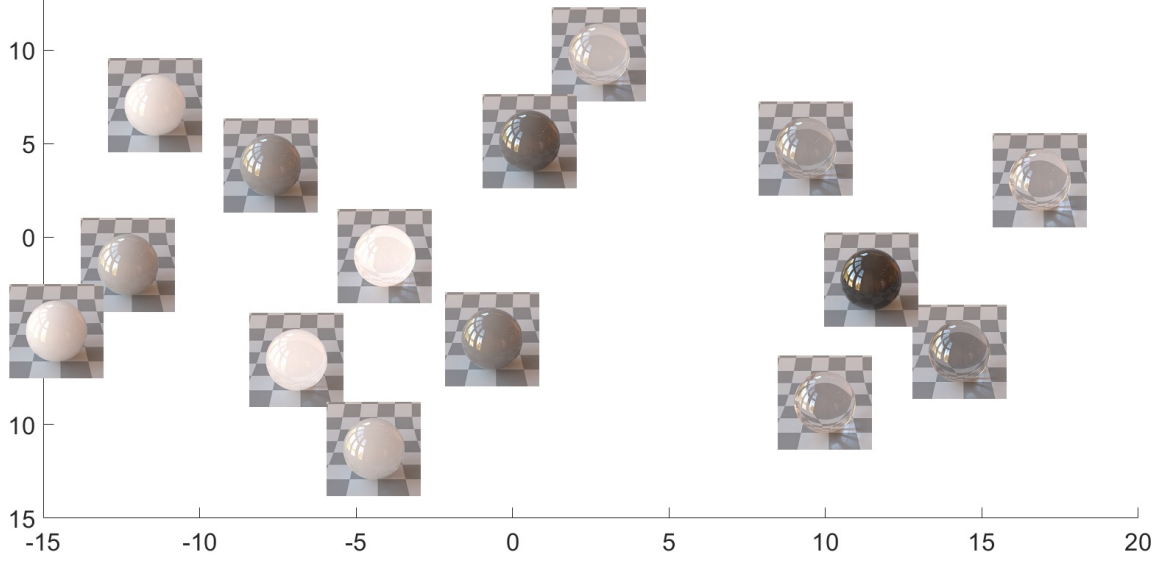


Fig. 37. 2D embedding of the 15 stimuli. It is apparent that transparent and dark opaque materials are located near each other. The same trend holds for higher dimensional embeddings.