

Perceived Glossiness: Beyond Surface Properties

Davit Gigilashvili, Jean-Baptiste Thomas, Marius Pedersen, Jon Yngve Hardeberg;
Department of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway

Abstract

Gloss is widely accepted as a surface- and illumination-based property, both by definition and by means of metrology. However, mechanisms of gloss perception are yet to be fully understood. Potential cues generating gloss perception can be a product of phenomena other than surface reflection and can vary from person to person. While human observers are less likely to be capable of inverting optics, they might also fail predicting the origin of the cues. Therefore, we hypothesize that color and translucency could also impact perceived glossiness. In order to validate our hypothesis, we conducted series of psychophysical experiments asking observers to rank objects by their glossiness. The objects had the identical surface geometry and shape but different color and translucency. The experiments have demonstrated that people do not perceive objects with identical surface equally glossy. Human subjects are usually able to rank objects of identical surface by their glossiness. However, the strategy used for ranking varies across the groups of people.

Introduction

Appearance is a complex psychovisual phenomenon that is defined as "the visual sensation through which an object is perceived to have attributes as size, shape, colour, texture, gloss, transparency, opacity, etc." [1] Due to its multiplex nature appearance is usually split into distinct attributes. According to CIE, there are four major appearance attributes: color, gloss, translucency and texture [1, 2]. Eugène [3] cites CIE definition of gloss as: "the mode of appearance by which reflected highlights of objects are perceived as superimposed on the surface due to the directionally selective properties of that surface" and adds that "gloss perception is particularly depending on the way that light is reflected from the surface of the object at and near the specular direction." [1] ASTM Standard Terminology of Appearance [4] 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." In computer graphics the Phong reflection model [5] (that is a simplification of bidirectional reflectance distribution function - BRDF) is widely used to model glossy appearance. The component responsible for this effect is the ratio of specularly reflected and incident light. However, the model does not account for transmission or sub-surface scattering and no translucency is considered. Ho *et al.* [6] have demonstrated correlation between perceived glossiness and perceived bumpiness, describing gloss as a "surface property", while Hunter [7] distinguishes six different types of gloss: 1. **Specular gloss** - "identified by shininess"; 2. **Sheen** - "identified by surface shininess at grazing angles"; 3. **Contrast gloss** - "identified by contrasts between specularly reflecting areas of surfaces and other areas"; 4. **Absence-of-bloom gloss** - "identi-

fied by the absence of reflection haze or smear adjacent to reflected high lights"; 5. **Distinctness-of-reflected-image gloss** - "identified by the distinctness of images reflected in surfaces"; 6. **Absence-of-surface-texture gloss** - "identified by the lack of surface texture and surface blemishes." He proposes that glossiness might be correlated with surface specular reflectance and concludes that reflectance distribution functions "offer the only means by which the reflectance properties of surfaces responsible for their glossiness may be completely specified." On the other hand, Motoyoshi *et al.* [8] propose that simple image statistics, like skewness of luminance histogram or similar metric of histogram asymmetry, are used by the human visual system to assess surface properties and glossiness without knowledge of the reflectance distribution function [9]. The authors explicitly mention gloss as a surface-related property without discussing the possibility that the histogram might be affected by transmission or sub-surface scattering of the light. They further conclude that average luminance has a significant impact on perceived lightness, but not on perceived glossiness and demonstrate the two images of Michelangelo's St Matthew sculpture that have identical mean luminance but substantially differ in perceived glossiness, while comparing grayscale images of the opaque surfaces. Nishida and Shin'ya [10] propose that a combination of mean luminance, luminance contrast, maximum and minimum luminance, as well as spatial structure of luminance gradients, might be cues for perception of surface properties. They also demonstrate that surface-reflectance constancy of the human visual system fails when shape is changed. Chowdhury *et al.* [11] have shown that perceived mesoscopic shape differs between translucent and opaque objects due to difference in luminance gradients.

Pellacini *et al.* [12] have explored dimensionality of gloss perception, introducing a perceptually uniform gloss space and psychophysically-based light reflection model that should enable cross-object description and matching of apparent gloss. Using multidimensional scaling the authors came up with a 2-dimensional space with orthogonal axes that are "qualitatively similar to the contrast gloss and distinctness-of-image gloss attributes". They also claim that CIELAB lightness parameter impacts apparent gloss and demonstrated that "apparent gloss is

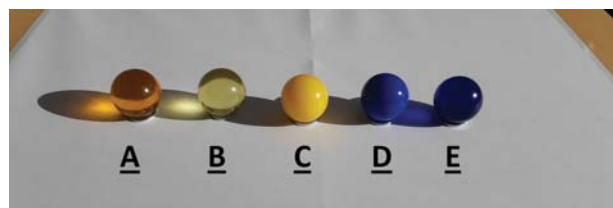


Figure 1. The objects used for the preliminary experiment.

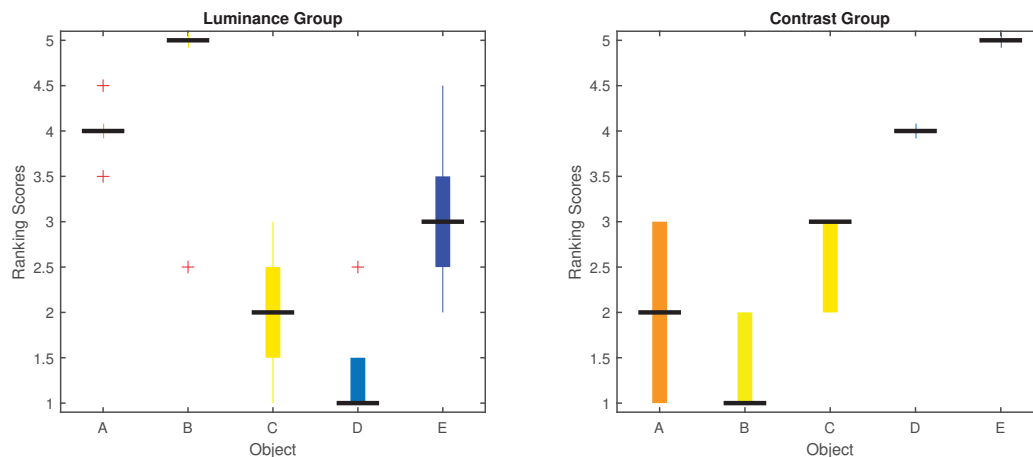


Figure 2. Boxplots for observer scores showing how observers ranked the five (A, B, C, D, and E) objects (Figure 1). 1 means least glossy, while 5 means most glossy. In case of ties, the mean score was taken. Central mark - median; bottom and top edges - 25th and 75th percentiles, respectively; Whiskers extend to the extreme data points excluding outliers; red '+' symbol - outliers. We can observe clear separation for both groups.

affected by the diffuse reflectance of a surface, with light colored surfaces appearing less glossy than dark ones having the same finish". Although the proposed framework performed well for their dataset, the study is limited to opaque spherical objects assuming that chromaticity and apparent gloss are independent, without mention of any possible impact from translucency.

The paper is organized as follows: in the next section background information is provided. Afterwards, we conduct detailed analysis of the first experiment [13] followed by the experimental setup of the new one. Subsequent section covers results and discussion. Finally, we conclude and outline the future work.

Background and Motivation

In an earlier paper [13] we summarized a psychophysical experiment where observers were asked to rank five spheres by their glossiness which had identical surface smoothness but different color and translucency (Figure 1). Aggregate frequency analysis did not show statistically significant differences in observer scores, making us hypothesize that similar gloss perception can be achieved with similar surface smoothness, but more thorough insight into the interviews of the observers has outlined three groups of people of roughly same size: 1. Subjects who considered all spheres to be equally glossy; 2. Subjects who ranked the spheres considering translucent ones more glossy. Those people mentioned shininess of the translucent spheres as the reason for their apparent glossiness. In this case brightness was the cue for them; 3. Subjects who ranked the spheres considering opaque ones more glossy. Those observers used distinctness-of-image gloss and contrast gloss (for the dark ones) as a cue. The three groups used different cues to reach the conclusion, and some of those cues may be impacted by other material properties, not only the shape and surface geometry.

In this paper we want to challenge the established opinion that gloss perception is solely surface-based quality. While translucency and color can contribute significantly to the cues like mean luminance as well as luminance contrast and luminance histogram, associated with perceived gloss in the literature [8, 10],

it has been proposed [8, 14] that the human visual system has poor ability, if any, to invert the optics. Therefore, we propose that translucency and color, particularly lightness, have significant impact on perceived glossiness. Translucency is a point of particular interest due to two reasons: first of all, light transmission and back-reflections increase overall luminance and shininess of the object that might be consciously or subconsciously associated with gloss; and secondly, caustics could play significant role too. According to Lynch [15], caustic is "three dimensional envelope of imperfectly focused rays" or "two-dimensional pattern formed when a caustic falls on a surface." Internal and external caustics and the glittering effect of the caustic highlights might be mistaken for specular highlights and thus, for gloss, considering their similarity in luminance, and proposedly poor optics inversion ability of the human visual system. We conducted series of psychophysical experiments asking people to rank objects by their glossiness. The objects had nearly identical surface smoothness but different color and translucency. As the observers were explicitly instructed that they could have ties among objects including tying all of them, if our hypothesis is false and perceived glossiness depends solely on the surface geometrical properties, the vast majority of them should have said that all objects have the same glossiness. In the previous paper [13] different cues used by subjects in opaque and translucent spheres compensated each other leading to statistically insignificant difference among perceived glossiness when analyzed the aggregated data. In order to clear up this ambiguity, we: 1. Analyzed the data from the first experiment [13] separately for different groups of people. 2. Replaced spheres with a complex object shape that decreases predictability of caustics and makes it impossible to observe distinctness-of-image gloss. As the cross-shape failure of reflectance constancy has been shown in [10], we used objects with an identical shape.

Group-based analysis of the first experiment

The first experiment using five spheres is discussed in [13]. The observers were asked to rank five spheres by their glossiness. Although the spheres had different colors and translucency, sur-

face geometry among them was nearly identical. While aggregate analysis of the overall data did not illustrate statistically significant differences in perceived glossiness, more thorough insight in the data revealed three different groups of the people using different strategies. Below we will illustrate group-based analysis of the data. The spheres used in the experiment are shown in Figure 1. 17 observers participated in the experiment. Six observers concluded that all spheres have the same glossiness; six people used luminance-based strategy (later referred as "luminance group"), and five people used distinctness-of-image gloss or contrast gloss-based strategy ("contrast group"). The boxplots for the latter two groups are illustrated in Figure 2. Due to low number of tests, it is difficult to assess statistical significance of the differences. However, the boxplots show very interesting trends. The "luminance group" has a very clear separation between shiny transparent A and B spheres, and opaque C and D spheres. The dark blue but semi-transparent sphere E has overlaps with both groups as it demonstrates characteristics of the both. On the other hand, for the "contrast group" there is a clear separation between A, B, and C spheres on the one hand, and D and E spheres, on the other hand. Dark blue and fully or significantly opaque spheres are considered more glossy, because this group of the subjects used a combination of distinctness-of-image gloss and contrast gloss that are stronger than in case of translucent or opaque but very light yellow spheres. Nevertheless, it is impossible to draw solid conclusions due to low number of subjects and test objects. We conducted a second experiment to verify the results.

Experimental Setup & Methodology

Task and Stimuli

The subjects were introduced to nine plastic female sculpture objects placed on an A3 white paper with a printed scale and two extremes: "Least Glossy" and "Most Glossy" points. Afterwards the following instruction was given: "Please, rank the objects by their glossiness: from the most glossy to the least glossy. You can have any number of ties, including the case, when all objects are tied and no ranking is possible." The observers were allowed to

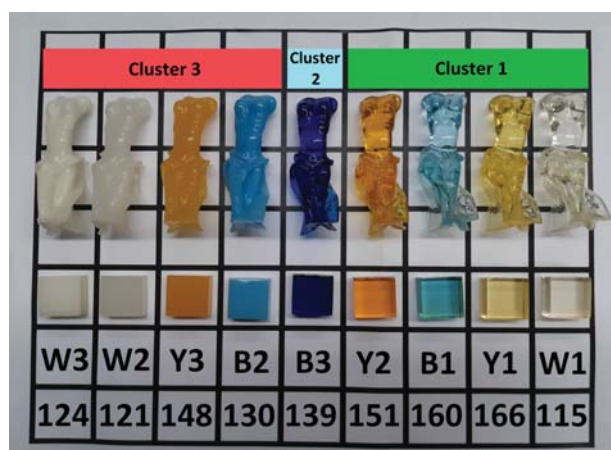


Figure 3. The female bust objects used for the the experiment. The corresponding 2-symbol codes are for reference purposes only. 3-digit codes are their IDs used by Thomas et al. Cuboid objects have been used for transmittance and relative radiance measurements discussed below. [16].

interact with the objects, touch and move them freely. No explicit definition has been given for gloss. However, they were allowed to check the definition in case of uncertainty. We used a subset of the *Plastique* artwork collection [16]. The collection has been created by **an independent artist** Aurore Deniel to support research on material appearance. The samples are illustrated on Figure 3.

Experimental Conditions

We made an assumption that impact of the illumination conditions is less than that of cross-individual differences. Psychophysical experiments have been conducted on several occasions in controlled and uncontrolled conditions, and similar trends have been revealed under all conditions. In total, 107 observers participated in the experiments. 7 experiments were conducted in uncontrolled conditions, namely: 1. 2018 Color and Imaging Conference, Demonstration Session (8 observers, attendees of the conference); 2. 2019 IS&T International Symposium on Electronic Imaging, Demonstration Session (17 observers, attendees of the conference); 3. Material Appearance 2019 Conference (8 observers, attendees of the conference); 4. Internal academic activity at the Norwegian University of Science and Technology (NTNU), Trondheim (5 observers, master and PhD students); 5. Internal academic activity at NTNU, Gjøvik (11 observers, high school students); 6. Internal academic activity at NTNU, Gjøvik (7 observers, bachelor students); 7. Internal academic activity at NTNU, Gjøvik (7 observers, bachelor, master and PhD students). In addition, two experiments took place in controlled conditions, in two different viewing booths with a distance of roughly 50 cm: 8. VeriVide Color Assessment Cabinet 60-5 under D65 illumination with 1392 lux and 6180K color temperature (30 observers of mixed backgrounds). 9. GretagMacbeth Spectralight III viewing booth under Ultralume 30 (U30) illumination with 665 lux and 2865K color temperature (14 observers of mixed backgrounds). The experiments were anonymous and no further demographic information has been collected.

Analysis of the Collected Data

The rank order of the object is recorded as a numerical value. For instance, if the object was ranked most glossy, it was assigned "1"; in case it was ranked second most glossy, the object was assigned "2", and so on. In case of ties, a mean score was assigned to all objects. For example, if the second and third objects were tied, each objects got rank equal to 2.5. If no ranking was done, each object was assigned "5". For visualization's sake, results of similar ranking strategies were grouped together, and the ranks given to the each object by different observers were plotted as a graph to visualize the variation of a position for a particular object among different trials (Figure 4). Besides, the rank scores for each object are illustrated as box-plots (Figure 5). An alternative method for analyzing the ranking could be considering each experiment a pair-comparison among all objects, where selected object gets 1, the other one gets 0, and both objects get 0.5 in case of a tie. As the both methodologies lead to nearly identical results, we report the former for consistency's sake with [13]. Afterwards, k-means clustering was conducted using MATLAB `kmeans()` function¹ to identify which objects were ranked together. That could help us to identify the right attributes that

¹MATLAB R2017b version.

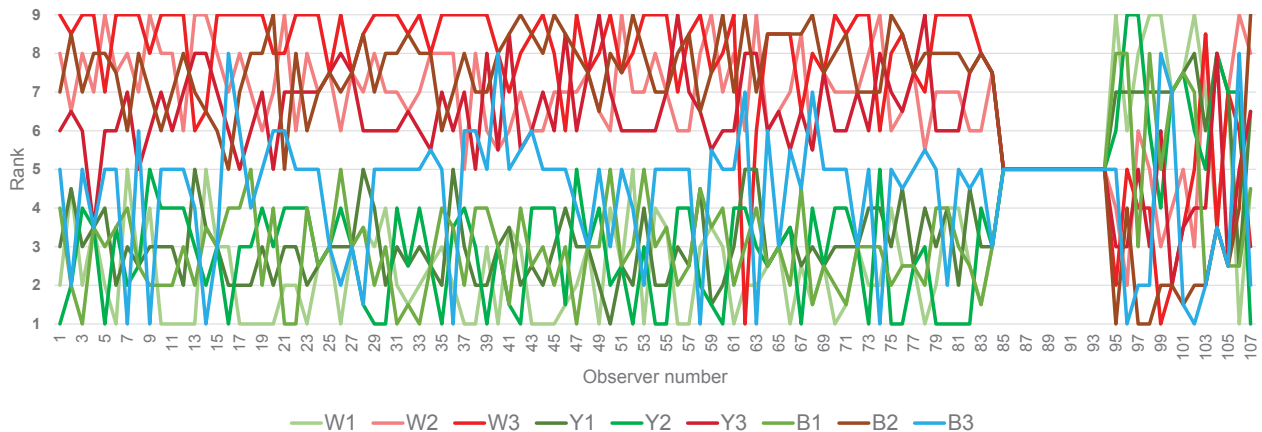


Figure 4. The aggregate results from all individual experiments. Each colored line corresponds to a particular object. For the majority of the subjects, we can see a clear separation between more transparent (marked with green hue lines), and more opaque objects (marked red hue lines).

made observers rank objects in a similar manner. Observations in this case were nine objects and variables were 107 ranks from 107 experiments. The cluster was defined as the centroid being the mean of all points in that particular cluster. Maximum number of iterations was set to 1000. Cluster centroids were initialized using *k-means++ algorithm* [17]. Finally, material luminance has been measured and correlated with mean ranking scores.

Results & Discussion

Graph Results

Identically to our previous experiment, three different ranking strategies have been observed:

1. 10 people (9.35%) mentioned that gloss was identical among objects, and thus, considered ranking impossible.
2. 84 people (78.50%) ranked more transparent objects over the ones closer to opacity.
3. 8 people (7.48%) opted for the objects closer to opacity.

The ranking of five people (4.67%) did not fit in any of the above-mentioned categories. It is worth mentioning that the trend has been similar in all illumination conditions. Clusters of the objects ranked similarly by each group of the people is further substantiated below by *kmeans clustering* results. The overall results are illustrated in Figure 4. The graphs for transparent objects are coded with the greenish hue, while the ones with more opacity are represented by reddish hue, and the dark blue translucent object that stands out from the rest of the dataset is represented by light blue graph. Each object can be identified with its two-symbol code from Figure 3. For clarity's sake, similar results are grouped across the horizontal axis. There is a very clear separation between green and red graphs for the vast majority of the cases, while blue graph oscillates between the two. In the majority of the cases, transparent objects have lower rank orders, i.e. are ranked more glossy. This group of observers is followed by the group of observers that have tied all objects. Finally, the red and green parts, still clearly separated, swap places. This part corresponds to the observers, who considered objects with more opacity being more glossy. By the right extreme of the plot, some chaotic arrangements are illustrated that did not follow

transparency-opacity cue. On the other hand, it is difficult to see patterns within transparent and opaque groups that makes us think that impact of chromatic information might be negligible.

Clustering

Clustering has been repeated 1000 times by new centroid initialization and the solution with the least sums of point-to-centroid distances was selected out of the 1000 trials. By observation of the graphs above, the number of clusters was set to 3. This lead us to the following clusters (illustrated in Figure 3):

1. Transparent and shiny objects: W1, B1, Y1, Y2.
2. Dark blue translucent object: B3.
3. Objects with more opacity and less shine: W2, W3, Y3, B2.

Rank scores and statistical properties

Rank scores have been illustrated as boxplots (Figure 5) for two major group of the observer population, and as an aggregate for all 107 observers. Objects from the same cluster are coded with the same hue. We can observe a very clear separation between transparent-shiny and more opaquish objects both for "Luminance Group" as well as for "Contrast Group" of the people, with a few outliers included, while object B3 from a separate cluster has some overlap with both clusters. In case of aggregate results, separation remains visible due to significantly higher number of observers in the "Luminance Group" and number of outliers increases due to inclusion the observers making no ranking or doing that with unique strategies. Statistical properties for each cluster of objects for each group of population are illustrated in Figure 6. For the Luminance group, as well as for the entire population, mean and median observer scores for more transparent objects are lower. Standard deviation of B3 for luminance and contrast groups is higher, as it oscillates between the two groups.

Transmittance Measurements

Transmittance spectra for each material has been measured in backlit illumination geometry and relative colorimetric values have been calculated. Due to the complexity of the surface of the female bust objects, measurements have been conducted on cuboid shapes of the identical material shown in Figure 3. The

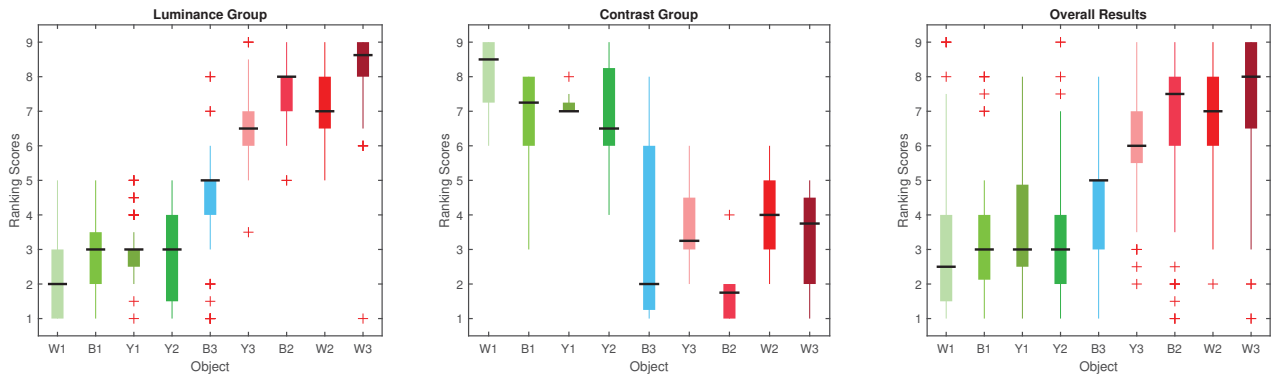


Figure 5. Boxplots for observer scores showing how each group of the observers ranked the objects. More transparent objects are given with greenish hue, objects closer to opacity are illustrated with reddish hue, while the object B3 is sky blue. 1 means most glossy, while 9 means least glossy. In case of ties, the mean score was taken. Central mark - median; bottom and top edges - 25th and 75th percentiles, respectively; Whiskers extend to the extreme data points excluding outliers; red '+' symbol - outliers. We can observe clear separation for both groups.

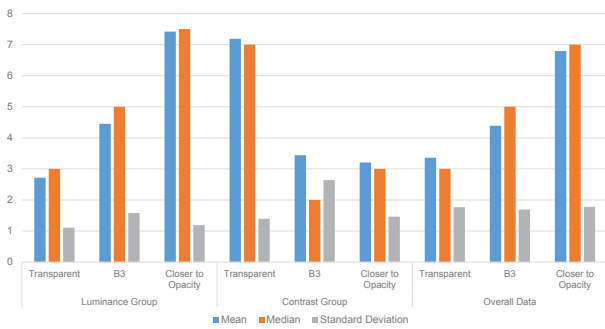


Figure 6. Statistical properties by cluster for each group of observers.

white paper seen through the object and caustics should have contributed to shiny appearance. Hence, transmitted luminance information, (Y from measured CIE XYZ), is seemingly correlated with mean rank scores for the "luminance group". This can be seen in Figure 7, where separation among high and low luminance objects is apparent, also supported by k-means clustering. Although luminance for B3 dark blue object is low, it has very high contrast gloss, observers explicitly mentioning that "highlights are more clearly visible on this object". Figure 8 illustrates mean ranking scores as a function of relative radiance expressed as a CIELAB L* value measured in reflectance setup, where cuboid objects were placed on the white background. This enables us to draw parallels with Pellacini's statement that objects with higher lightness in diffuse areas appear less glossy.

Discussion

While the impact of illumination conditions is still to be studied, cross-individual differences might have significantly affected the results. The most obvious illustration of this fact is abundance of "no ranking" scenario for Material Appearance and Electronic Imaging Demonstration Session experiments, where the majority of the subjects had expertise in color, vision, or related fields. Those who considered all objects equally glossy were explicitly asked to justify their decision. All of them defined gloss as surface-only property, limiting themselves to surface judgment.

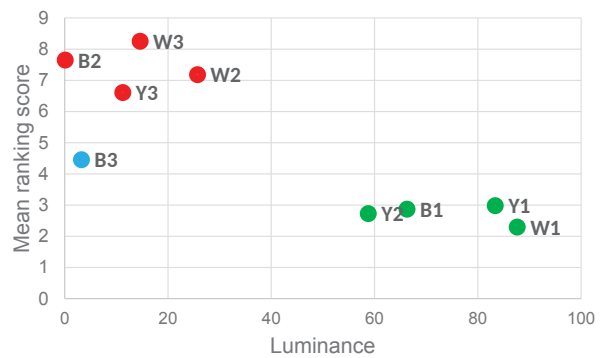


Figure 7. Mean ranking score as a function of transmittance expressed as luminance value.

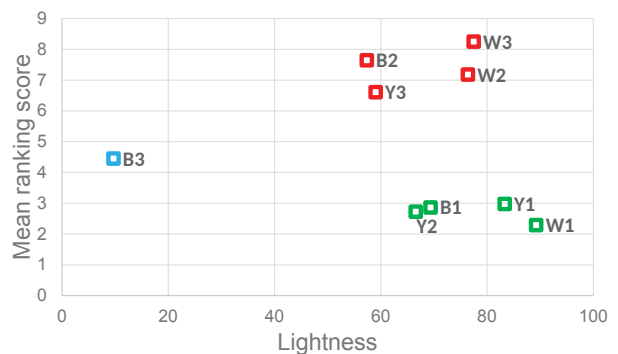


Figure 8. Mean ranking score as a function of relative radiance expressed as CIELAB L* value.

In general, still 97 out of 107 observers were able to rank the objects even though they had explicitly given a possibility not to. After analyzing the data, three groups of people pop out: the ones that judge surface only; people who consider transparent-shiny objects more glossy; and the people who considered objects with more opacity being more glossy. The justification of ranking more opaque ones more glossy were clarity of the highlights and higher

contrast gloss, while people opting for transparent ones associate gloss with overall shine and high brightness without scrupulous study of the details. In contrast with the previous experiment, where the two groups were of the equal size, here shininess-based decisions prevail significantly. This could be explained by the absence of distinctness-of-image gloss on the complex surface of the female bust objects, in contrast with a sphere. This confirms Nishida's claim [10] that perceptual surface-reflectance constancy fails when shape is changed, and challenges Pellacini's sphere-based model [12]. Clustering supports our hypothesis that translucency-related attributes as transmittance-measured luminance are common within a cluster. This leads us to hypothesize that gloss and translucency might impact each other. Several observers explicitly complained that it was impossible to isolate translucency/transparency and gloss for above-mentioned objects and thus, to judge them independently. Translucency difference between the two clusters was very large making it challenging to discard its effect. However, the ranking pattern for the B3 object was more irregular. In some cases it was ranked most glossy, justified by high contrast gloss. We can draw a parallel with the first experiment, where sphere "E", made of the similar material, also had substantial confidence interval overlaps with the both groups. This is in agreement with Pellacini's [12] finding that "for the same specular energy, contrast gloss is smaller for lighter objects". Assuming that specular reflections are identical, higher relative radiance in the diffuse part (Figure 8) leads to higher perceived lightness in non-specular areas, and thus, lower contrast gloss. Contrast and clarity of the highlights were mentioned as a cue when they came from surface reflection only, while being less reliable in case of ambiguity whether the light originated from surface reflection or from sub-surface scattering. In total, light transmission properties have impacted perceived gloss in several ways. While contributing to specular gloss by transmission and caustics, contrast gloss is impacted by lightness of the diffuse areas in opaque materials.

Conclusion and Future Work

We have observed that glossiness perception function varies among subjects. While some people try to stick to the literature definition, the vast majority of them ignore surface similarity and sort out objects by gloss using their own criteria. Whether they completely ignore the surface similarity, or they consider it but look for the additional criteria, needs to be explored in the future. There is a very clear indication that perceptual gloss cannot be estimated by surface properties only and light transmission among others might have impact on it. However, the data at hand does not enable us to analyze what is the exact way translucency strengthens glossiness perception and whether the effect comes from overall increase in luminance after light transmission, or due to internal and external caustics that are mistaken for the specular reflections. In future work we should isolate those phenomena and study their impact separately. The hypothesis needs further investigation with more dense sampling across translucency-opacity scale possibly using computer graphics. However, it comes with the compromise that tactile information - a widely-used cue for surface estimation will be lost. Although darker colors enable higher contrast-gloss and contribute to gloss perception, the role of chromatic information is still to be determined by measuring and studying reflection properties. We hypothesize that higher

transmittance will lead to stronger gloss perception in the majority of the naïve observers, but low brightness/shine for dark opaque objects could be compensated with increased contrast gloss. Particular interest will be measurement of scattering coefficient and inclusion of the multi-material objects with an eventual goal to model a correlation between material properties and perceived gloss. Development of this work will be reported in the future.

References

- [1] Christian Eugène, "Measurement of "total visual appearance": a CIE challenge of soft metrology," in *12th IMEKO TC1 & TC7 Joint Symposium on Man, Science & Measurement*, 2008, pp. 61–65.
- [2] Michael Pointer, "A framework for the measurement of visual appearance," *CIE Publication*, pp. 175–2006, 2006.
- [3] "CIE 17.4:1987 international lighting vocabulary," International Commission on Illumination, 1987.
- [4] "ASTM E284-17 standard terminology of appearance," ASTM International, West Conshohocken, PA, 2017.
- [5] Bui Tuong Phong, "Illumination for computer generated pictures," *Communications of the ACM*, vol. 18, no. 6, pp. 311–317, 1975.
- [6] Yun-Xian Ho, Michael S Landy, and Laurence T Maloney, "Conjoint measurement of gloss and surface texture," *Psychological Science*, vol. 19, no. 2, pp. 196–204, 2008.
- [7] Richard S. Hunter, "Methods of determining gloss," *NBS Research paper RP*, vol. 958, 1937.
- [8] Isamu Motoyoshi, Shin'ya Nishida, Lavanya Sharan, and Edward H Adelson, "Image statistics and the perception of surface qualities," *Nature*, vol. 447, no. 7141, pp. 206–209, 2007.
- [9] Fred E Nicodemus, "Directional reflectance and emissivity of an opaque surface," *Applied optics*, vol. 4, no. 7, pp. 767–775, 1965.
- [10] Shin'ya Nishida and Mikio Shin'ya, "Use of image-based information in judgments of surface-reflectance properties," *JOSA A*, vol. 15, no. 12, pp. 2951–2965, 1998.
- [11] Nahian S Chowdhury, Phillip J Marlow, and Juno Kim, "Translucency and the perception of shape," *Journal of vision*, vol. 17, no. 3, pp. 1–14, 2017.
- [12] Fabio Pellacini, James A Ferwerda, and Donald P Greenberg, "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., 2000, pp. 55–64.
- [13] Davit Gigilashvili, Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Marius Pedersen, "Behavioral investigation of visual appearance assessment," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2018, pp. 294–299.
- [14] Roland W Fleming and Heinrich H Bülthoff, "Low-level image cues in the perception of translucent materials," *ACM Transactions on Applied Perception (TAP)*, vol. 2, no. 3, pp. 346–382, 2005.
- [15] David K Lynch, William Charles Livingston, and William Livingston, *Color and light in nature*, Cambridge University Press, 2001.
- [16] Jean-Baptiste Thomas, Aurore Deniel, and Jon Y Hardeberg, "The plastique collection: A set of resin objects for material appearance research," *XIV Conferenza del Colore, Florence, Italy*, p. 12 pages, 2018.
- [17] David Arthur and Sergei Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.