Gloss management for consistent reproduction of real and virtual objects

Supplementary Material

Bin Chen
binchen@mpi-in.mpg.de
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Michal Piovarči
michal.piovarci@gmail.com
Institute of Science and Technology Austria
Klosterneuburg, Austria

Chao Wang
chaowang@mpi-inf.mpg.de
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Hans-Peter Seidel
hpseidel@mpi-sb.mpg.de
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Piotr Didyk
piotr.didyk@usi.ch
Universitȧ della Svizzera Italiana
Lugano, Switzerland

Karol Myszkowski
karol@mpi-inf.mpg.de
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Ana Serrano
anase@unizar.es
Universidad de Zaragoza, I3A
Zaragoza, Spain

ABSTRACT

This document supplements more details in the topics as follows:

- S1: Detailed Stimuli Description
- S2: Photographs of Fabricated Geometries
- S3: Preliminary Study: Ordering Experiment
- S4: Main Study: Procedure
- S5: Main Study: Results
- S6: Additional Study: Lightness experiment
- S7: Gloss Management: Correction
- S8: Applications

CCS CONCEPTS

• **Computing methodologies** → *Appearance and texture representations*; *Perception.*

KEYWORDS

glossiness perception, glossiness reproduction

S1 DETAILED STIMULI DESCRIPTION

Here we describe additional details about the selection and fabrication of our stimuli.

S1.0.1 Geometry. In the work of Serrano et al. [2021], the authors study gloss perception as a function of surface geometry. Informed by this work, we select the two geometries that led to the most extreme differences in gloss judgments. In particular, we select the *dragon* as an example of a highly tessellated, complex mesh, as well as the *ghost*, whose shape is a combination of a simple sphere and a wavy surface [Havran et al. 2016]. As examples of more "neutral" shapes, we select the *blob* with a smoothly changing, low curvature surface that shows a good performance in gloss discrimination experiments [Vangorp et al. 2007], and the *bunny* that features more high-frequency details.

S1.0.2 Illumination. In order to create a large scale diffuse illumination, we hang two layers of transmissive diffusers over the experimental area to construct the diffuse illuminant. The diffusers even redistribute ceiling mounted 4000 K light sources. We estimated the final illumination to 67 ± 5 lux over the entire experimental table. To construct the other illumination conditions, we build a gantry frame above the observer. The gantry contains five additional 4000 K spotlight emitters for creating more-complex illumination. Refer to Fig. S8.13 for more details on the specific light source placement.

S1.0.3 Gloss Fabrication. To physically realize the selected geometries we reproduce them on a stereolithography printer. The geometries are hand-polished and evenly coated with matte black (Pantone Hexachrome Black U) lacquer. In order to produce gloss variations, we rely on off-the-shelf varnishes. We mix Schmincke 610 glossy varnish and Schmincke 611 matte varnish in different proportions. For clarity, we will refer to the mixtures by the percentage of glossy varnish. To uniformly coat the geometries, we dip them into a pre-prepared varnish mixture. We repeat this process three times to ensure a uniform coating. Figure S8.1 demonstrates the enlarged view of details on the object surface after 3D printing (left), after polishing and coating (middle), and after varnishing with 100% glossy varnish (right). We include in our experiment *ten* samples with the percentage of glossy varnish ranging from 10% to 100% in 10% increments. Figs. S8.2, S8.3, S8.4, and S8.5 show all glossiness samples for four different shapes that were used in our main experiment under the *diffuse*, *one spotlight*, *three spotlights*, and *five spotlights* illuminations, respectively.

S1.0.4 Display settings and calibration. To create a virtual depiction of our fabricated samples we capture HDR photographs. Each photograph is color adapted to match our 4000 K light sources. To match the luminance of the HDR image to our scene we use the

gamma-offset-gain display model [Berns 1996]. We manually measure the peak brightness, black level, and surface reflection of the display using luminance meter Minolta LS-100 to restore the luminance of the real world as much as possible. The monitor has an elevated black level. We repeat this procedure for our three display brightness levels (*Bright*-75%, *Medium*-29%, and *Dark*-8%). The measured data and fitted curves are shown in Fig. S8.6. Moreover, we show the luminance range coverage of our three display brightness under our four illuminations using bunny with 100% glossiness level, as shown in Fig. S8.18. An exact peak luminance match is possible only for diffuse illumination. For the remaining illuminations, the main discrepancy lies in the over-saturated glossy highlights. In all display brightness scenarios the black luminance levels in the display are elevated with respect to the real world counterparts.

S2 PHOTOGRAPHS OF FABRICATED GEOMETRIES

To display the physical sample we follow a standardized approach. We start by taking a full dynamic range image of the physical stimuli from the head position of the participant. We collect a total of 13 brackets to ensure we capture the glossy highlights and combine them into a single image using the method of Hanji et al. [2020]. Then, the white balance of merged HDR image is changed to 4000 K to match our physical light sources.

For the purpose of this work we fabricated four geometries *dragon*, *ghost*, *blob*, and *bunny* with 10 different varnish finishes. Here we visualize tonemapped HDR photos of our samples under *diffuse* (Fig. S8.2), *one spotlight* (Fig. S8.3), *three spotlights* (Fig. S8.4), and *five spotlights* (Fig. S8.5) illumination.

S3 PRELIMINARY STUDY: ORDERING EXPERIMENT

The goal of this preliminary experiment is to validate our selected varnish mixtures. We aim at verifying that the selected mixtures uniformly span a large range of perceived glossiness values, i.e., that our samples do not cluster perceptually in some region of the glossiness scale, and that our samples are different enough so that adjacent levels can be distinguished while being close to their discrimination threshold. For this purpose, we design an ordering experiment both for the real varnished objects and their displayed counterparts. This experiment is conducted on all four geometries (*ghost*, *blob*, *bunny*, and *dragon*) under two illuminations (*diffuse* and *one spotlight*) with the *bright* display condition, and repeated both on the display and the real world. The participants' task is to order the ten gloss samples (initially set at random order) by increasing perceived gloss. In total, 14 volunteers (< 60 years old, 6 females, 8 males, no intersex/other) with normal or corrected to normal vision participated in this study.

Results. First, following Thurstone's Law of comparative judgment [Thurstone 1927], from the mean rank of the samples (ordinal scale) we compute the corresponding z-scores (interval scale) and visualize the scale of perceived glossiness intensity as a function of the physical gloss of our samples (refer to Figs. S8.7 and S8.8). This is computed for each combination of geometry and illumination. We confirm that the perceived glossiness of our samples increases monotonically and it is well distributed across the range. Second, we use Friendman Rank Sum test to analyze whether the rank of each sample (dependent variable) is significantly different between the different samples (independent variable). For post-hoc analysis we use Durbin-Conover tests (Figs. S8.9, S8.10 and S8.11). For all combinations of the factors all adjacent samples are significantly different ($p < 0.001$). This indicates that users are able to distinguish all the samples both in the real setup and in the display. Our data also shows that the participants' ordering is not always perfect, suggesting that adjacent samples are close to the discrimination threshold while still being distinguishable.

S4 MAIN STUDY: PROCEDURE

Before the main experiment each participant underwent a training session that illustrated the concept of gloss was given to each participant. This was followed by a test session to confirm their understanding of the concept. All the 42 participants passed the test. For the main experiment participants were sitting in front of a table. The chair height was adjusted so that the head of each participant was at approximately the same height. On the left of the participant we positioned the display, which was directly facing the observer and was positioned in such a way that it avoided direct reflection from the light sources. In front of the participant we showed the physical sample, which was attached to a holder that maintained a consistent position and orientation of each sample with respect to the observer. Both the physical and virtual samples were placed at 80 cm with 5.725° viewing angle from the observer and were adjusted to be of the same size and at the same height. Participants' task was to match the image in the display (one of the ten photos corresponding to our ten real samples) to that of the real object in terms of perceived gloss. We developed a dedicated interactive interface that allowed participants to switch between the ten images with the mouse wheel or the keyboard. The glossiness starting point was randomly reset after each trial to avoid cross-impact between different trials. Each participant saw a total of 160 trials (10 samples x 4 geometries x 4 illuminations under a random display brightness) in 4 sessions, one for each illumination. The participants were required to have 3-minutes break to adapt their vision after shifting to new illumination. The whole experiment lasts around 75 minutes.

S5 MAIN STUDY: RESULTS

In this section we include additional figures and details related to the main statistical analysis of the data. In Figure S8.12 we show all the data collected in our main experiment by each level of each factor. Then, in Tables 4 and 5 we include respectively the coefficients and the p-values for each coefficient for the multinomial logistic regression, in which we base our analysis in the main paper.

S6 ADDITIONAL STUDY: LIGHTNESS EXPERIMENT

The lightness experiments follows the same procedure as our main study but we consider only one geometry: *bunny*, manufactured using two achromatic colors: *gray* and *white*. There were 12 volunteers that participate in this user study (age 22 ~ 32, 5 females, 7

males, no intersex/other). All participants had normal or corrected to normal eyesight.

S7 GLOSS MANAGEMENT CORRECTION

To capture the reflectance of our varnishes we manufacture 10 cylindrical samples onto which we apply our 10 varnish mixtures, (Fig. S8.14). Next, to get raw reflectance data, we follow the measurement procedure described by Piovarči et al. [2020]. Finally, we fit a Cook-Torrance model with GGX microfacet distribution [Trowbridge and Reitz 1975] into the measured data. To form a mapping from varnish mixture to measured roughness we use piece-wise linear interpolation of the fitted values, (Fig. S8.15).

Table 1: Our per illumination gloss correction has the form of $y = p_1x^2 + p_2x + p_3$. In the following table we present the fitted values based on our main experiment. To interpolate the curves to different brightness levels we rely on a piece-wise linear fit of the coefficients.

Brightness	Illumination	p_1	p_2	p_3
Bright	diffuse	-0.003471	1.027	32.04
	one spotlight	0.005098	0.3963	12.02
	three spotlight	-0.0008342	1.028	5.961
	five spotlight	-0.0006012	1.015	4.926
Medium	diffuse	-0.001752	0.6643	52.99
	one spotlight	0.00268	0.6248	12.2
	three spotlight	0.001046	0.8028	10.39
	five spotlight	0.0005564	0.845	11.33
Dark	diffuse	0.001194	0.1627	73.82
	one spotlight	0.003167	0.5537	14.49
	three spotlight	-0.0001339	0.9006	10.57
	five spotlight	-0.0004012	0.9275	11.77

S8 APPLICATIONS

S8.1 Rendering for applications

To be able to reproduce the scene with high reality we capture 8K environment maps in high dynamic range (HDR) for our four illuminations as shown in Fig. 3 in the main paper. Specifically, originate from the point at which real objects are placed, we capture 5 Zenith angles ($-60°, -30°, 0°, 30°, 60°$) and 12 azimuth angles that evenly cover $[0°, 360°]$, results in 60 views in total with the help of Panorama head. Each view is fused from 13 exposures (1/500, 1/250, 1/125, 1/60, 1/30, 1/15, 1/8, 1/4, 1/2, 1, 2, 4, 8) using algorithm proposed by [Hanji et al. 2020] (considering noise reduction) to fully cover the dynamic range of both illuminations.

Rendering. With the HDR environment maps, Cook-Torrance model, and the geometries all ready, we first map the real size of whole setup into the unit of Mitsuba2 renderer based on same proportion. We apply the thin lens camera model with focal length set to 50 mm instead of pinhole model to better simulate the real camera that was used for real object capture.

S8.2 Digital Product Design: Results

Fig. S8.16 shows the results of the evaluation experiment for *Digital Product Design* application broken up by display brightness, illumination and geometry.

Table 2: This table shows how much our corrections modifies the varnish mixture to achieve a consistent match between displayed version and a manufactured artifact.

Illumination	Glossiness	Corrected glossiness
diffuse	20%	0%
diffuse	60%	30.3%
one spotlight	20%	16.6%
one spotlight	60%	65.6%

S8.3 Digitizing Physical Artifacts: Results

We include in Fig. S8.17 the results of the evaluation experiment corresponding to the *Digitizing Physical Artifacts* application broken up by illumination and geometry.

Table 3: This table shows the effect of our correction on the roughness parameter of a fitted Cook-Torrance model to achieve consistent display of digitized physical artifacts.

Brightness	Illumination	50%	62%	87%	100%
Bright	diffuse	0.0747	0.0589	0.0352	0.0324
	one spotlight	0.1205	0.0910	0.0749	0.0735
	three spotlight	0.1050	0.0788	0.0639	0.0580
	five spotlight	0.1079	0.0798	0.0654	0.0594
Dark	diffuse	0.0401	0.0344	0.0278	0.0263
	one spotlight	0.1075	0.0828	0.0726	0.0665
	three spotlight	0.1008	0.0783	0.0648	0.0592
	five spotlight	0.0967	0.0768	0.0619	0.0559

REFERENCES

Roy S Berns. 1996. Methods for characterizing CRT displays. *Displays* 16, 4 (1996), 173–182.

Param Hanji, Fangcheng Zhong, and Rafał K Mantiuk. 2020. Noise-aware merging of high dynamic range image stacks without camera calibration. In *European Conference on Computer Vision*. Springer, 376–391.

Vlastimil Havran, Jiri Filip, and Karol Myszkowski. 2016. Perceptually motivated BRDF comparison using single image. In *Comput. Graph. Forum*, Vol. 35. 1–12.

Michal Piovarči, Michael Foshey, Vahid Babaei, Szymon Rusinkiewicz, Wojciech Matusik, and Piotr Didyk. 2020. Towards Spatially Varying Gloss Reproduction for 3D Printing. *ACM Trans. Graph.* 39, 6, Article 206 (nov 2020), 13 pages. https://doi.org/10.1145/3414685.3417850

Ana Serrano, Bin Chen, Chao Wang, Michal Piovarci, Hans-Peter Seidel, Piotr Didyk, and Karol Myszkowski. 2021. The effect of shape and illumination on material perception: model and applications. *ACM Trans. on Graph.* 40, 4 (2021).

Louis L Thurstone. 1927. A law of comparative judgment. *Psychol. Rev.* 34, 4 (1927), 273.

TS Trowbridge and Karl P Reitz. 1975. Average irregularity representation of a rough surface for ray reflection. *J. Opt. Soc. Am. A* 65, 5 (1975), 531–536.

Peter Vangorp, Jurgen Laurijssen, and Philip Dutré. 2007. The influence of shape on the perception of material reflectance. In *Proc. ACM SIGGRAPH*. 77:1–77:9.

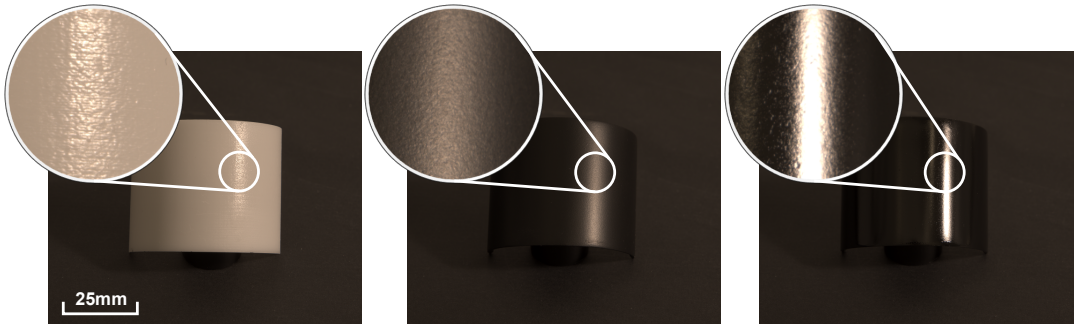

Figure S8.1: Fabrication demonstration. From left to right: object surface after 3D printing, after polishing and coating, and after varnishing using 100% glossy varnish.

Table 4: Statistical analysis of the main experiment described in Sec. 4 in the main paper. Model coefficients corresponding as obtained from the multinomial logistic regression

Gloss	X.Intercept.	Illum1P	Illum3P	IllumD	GeomBlob	GeomDragon	GeomGhost	DispBright29	DispBright75
20	0.082473392	-0.081791484	0.20247418	0.592862438	-0.299997416	0.053021681	-0.840488963	-0.471701548	-0.082954675
30	-0.353935033	-0.121277657	0.023275547	1.331636021	-0.033057145	0.681120037	-0.51776405	-0.4502114	-0.456993236
40	-1.085066114	-0.944998574	-0.232723664	2.30586456	-0.431553431	0.198804725	-0.710511224	-0.464519018	-0.434413211
50	-2.687472411	-1.33096533	-0.102317144	3.540931444	-0.002954862	1.393608162	-0.03026577	-0.588938858	-0.79456742
60	-3.251554959	-1.673627767	-0.302629542	4.872558135	-0.553704181	1.080533233	-0.157352838	-0.683979167	-0.920499761
70	-3.33021821	-1.864614826	-0.406755244	5.408736564	-0.430435375	1.853039571	-0.619879382	-0.656253092	-1.258355236
80	-5.90892343	-2.101561438	0.023043221	7.461932773	-0.073113957	2.444712992	-0.257812553	-0.933985786	-1.820664836
90	-8.288171169	-2.470459782	-0.65770454	9.885541253	-0.879517258	2.297634044	0.085357322	-1.238114592	-2.727506232
100	-12.09966967	-2.912454556	-0.571147699	13.19561593	-1.477285981	2.73417088	0.437482509	-1.358610054	-3.432049318

Table 5: Statistical analysis of the main experiment described in Sec. 4 in the main paper. P-values corresponding to each of the coefficients as obtained from the multinomial logistic regression

Gloss	X.Intercept	Illum - 1S	Illum - 3S	Illum - Diff	Geom - Blob	Geom- Dragon	Geom- Ghost	Disp - Medium	Disp - Bright
20	0.722895861	0.659745508	0.290957743	0.134382834	0.13978839	0.813579241	3.29497E-05	0.011633169	0.651249826
30	0.131521263	0.497746899	0.901350016	0.000244496	0.86996673	0.001915419	0.009053658	0.01099357	0.011247266
40	9.90148E-05	3.85827E-06	0.25566488	1.84521E-10	0.048157414	0.414869932	0.000892651	0.018414702	0.028788698
50	6.03961E-14	1.15971E-08	0.657809577	0	0.990478004	2.64838E-07	0.901131486	0.00679211	0.000298979
60	0	8.13372E-12	0.21315802	0	0.027936214	0.000101775	0.516807219	0.002381474	4.49195E-05
70	0	6.43929E-15	0.087044151	0	0.077756564	2.76068E-12	0.011124186	0.002460768	1.21296E-08
80	0	1.11022E-15	0.927561676	0	0.775396726	0	0.314996106	3.2062E-05	3.10862E-15
90	0	2.57572E-14	0.04100352	0	0.002279958	3.26406E-14	0.758601482	4.72377E-07	0
100	0	1.33227E-15	0.112849306	0	2.33888E-06	0	0.13904202	1.65129E-07	0

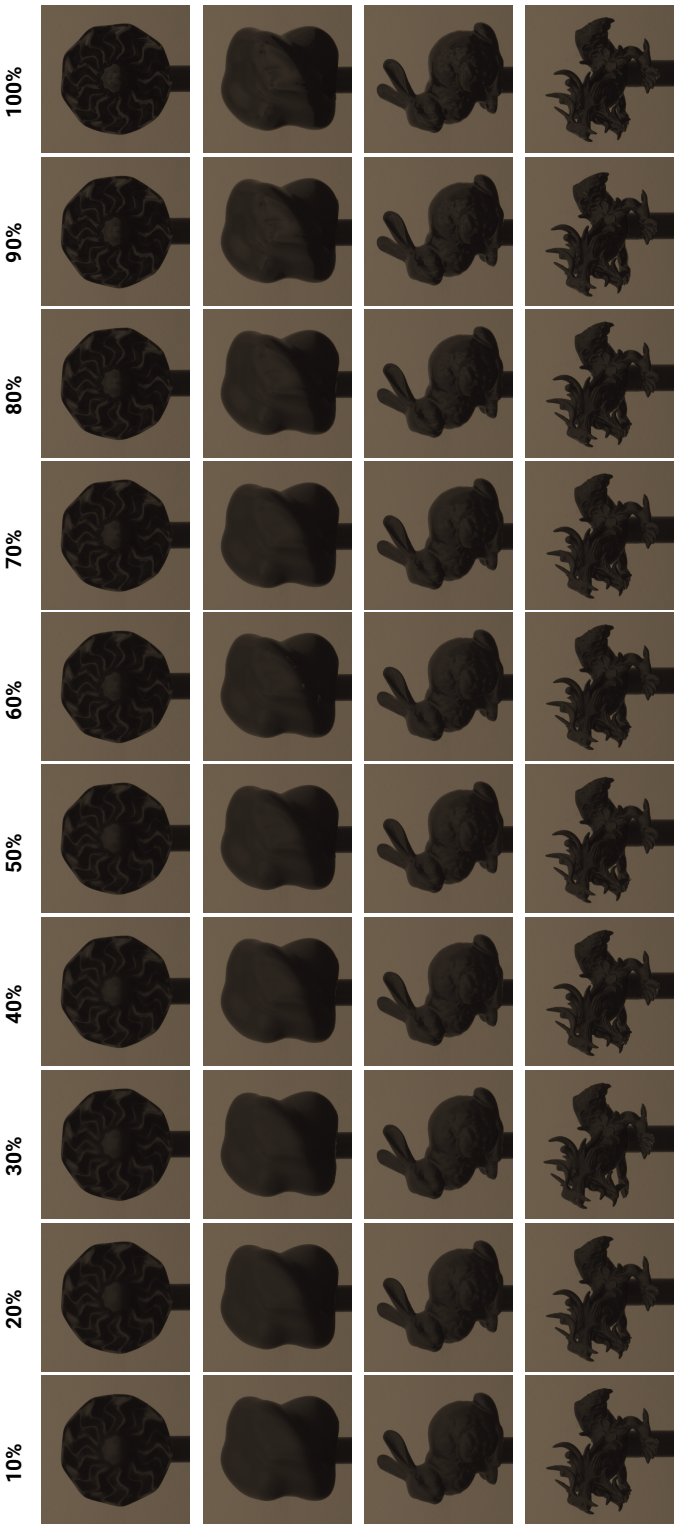

Figure S8.2: All our fabricated samples corresponding to the main experiment photographed under the diffuse illumination.

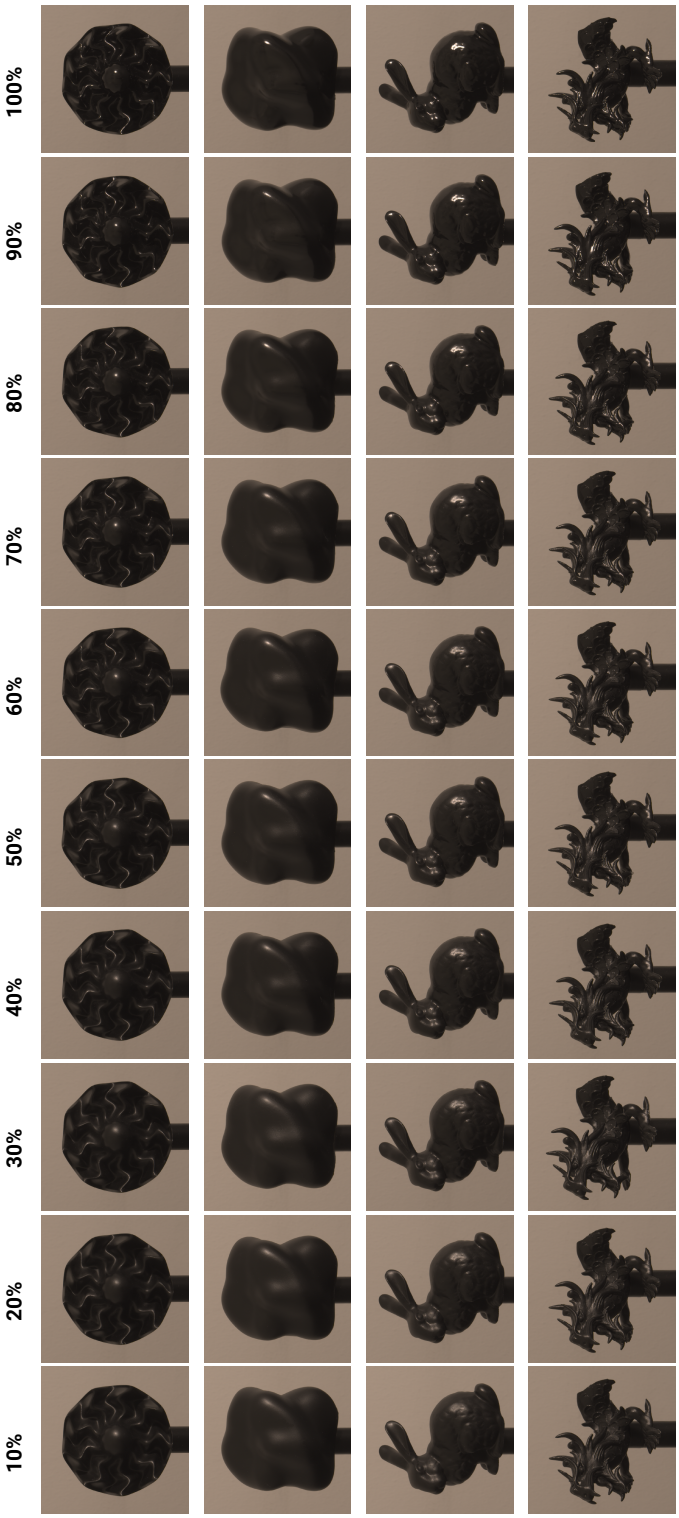

Figure S8.3: All our fabricated samples corresponding to the main experiment photographed under the one spotlight illumination.

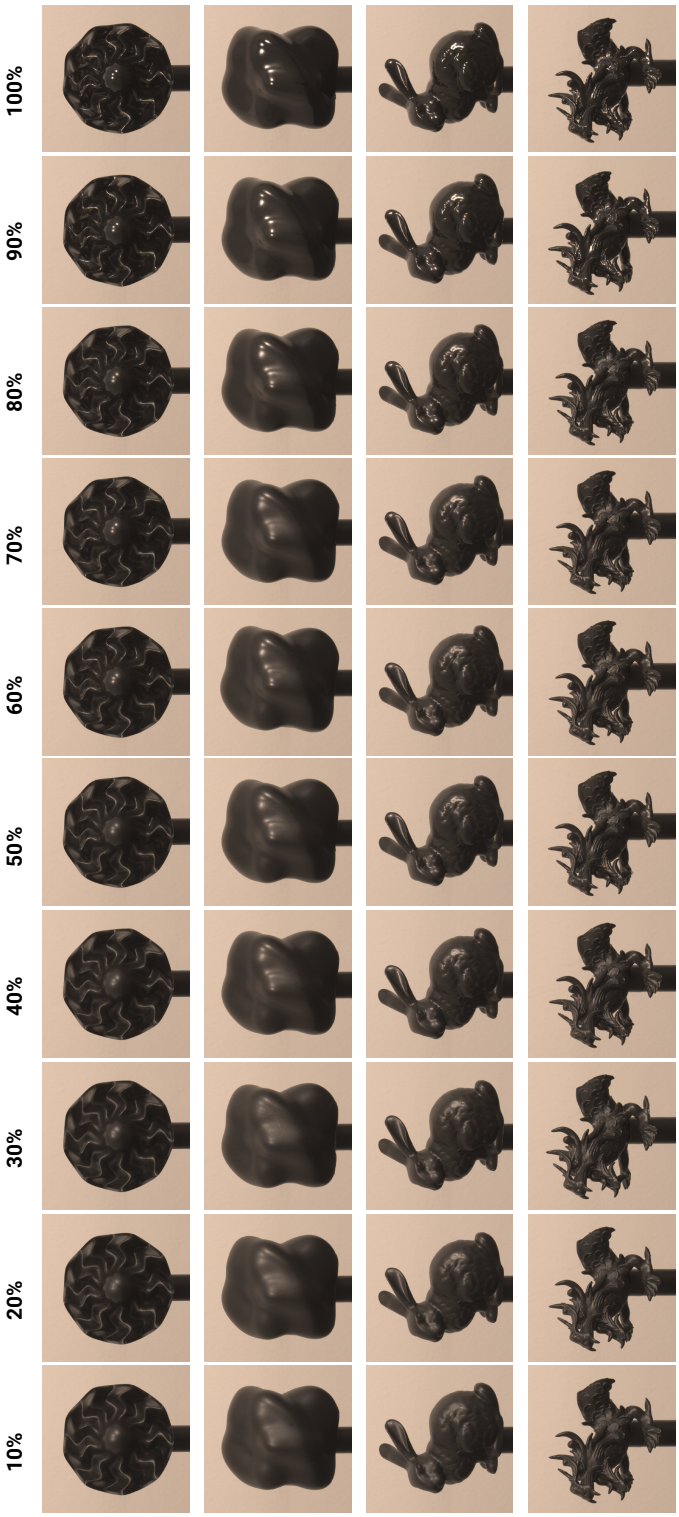

Figure S8.4: All our fabricated samples corresponding to the main experiment photographed under the three spotlights illumination.

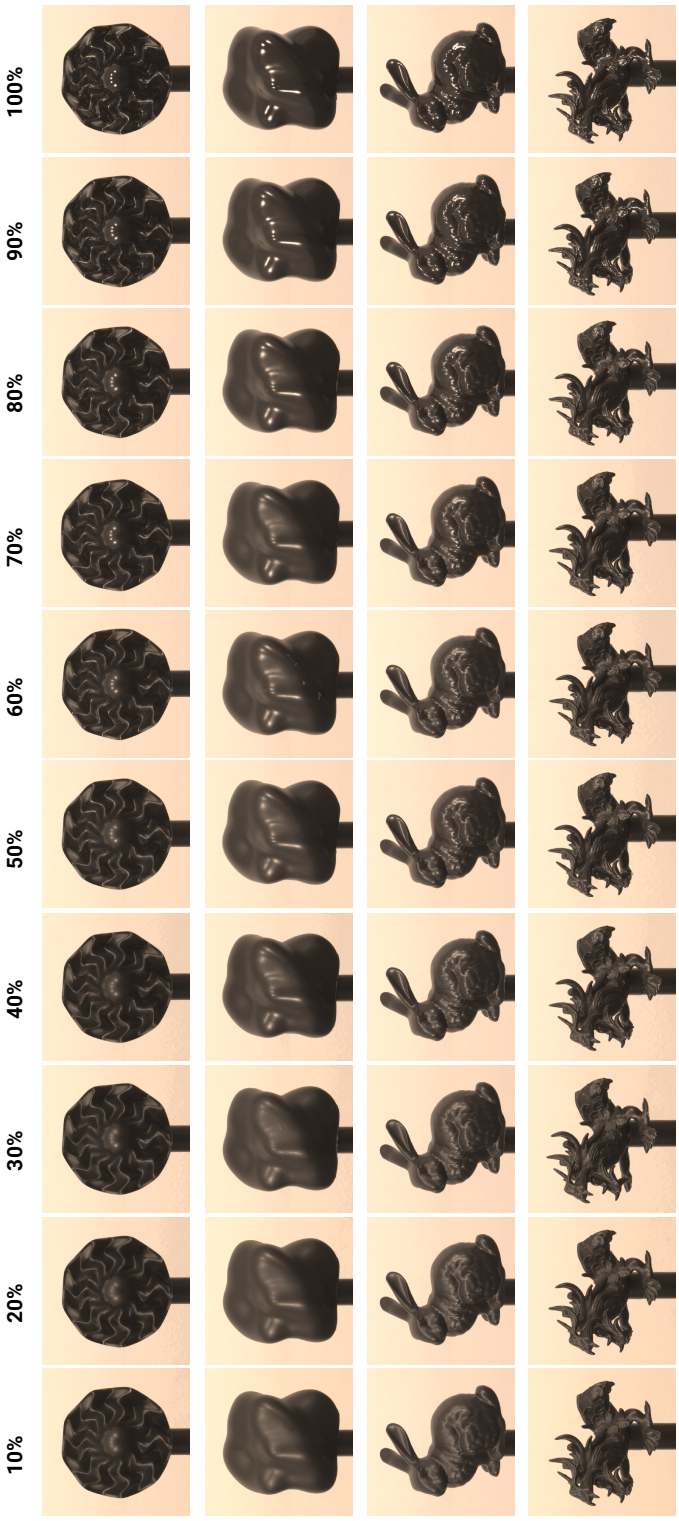

Figure S8.5: All our fabricated samples corresponding to the main experiment photographed under the five spotlights illumination.

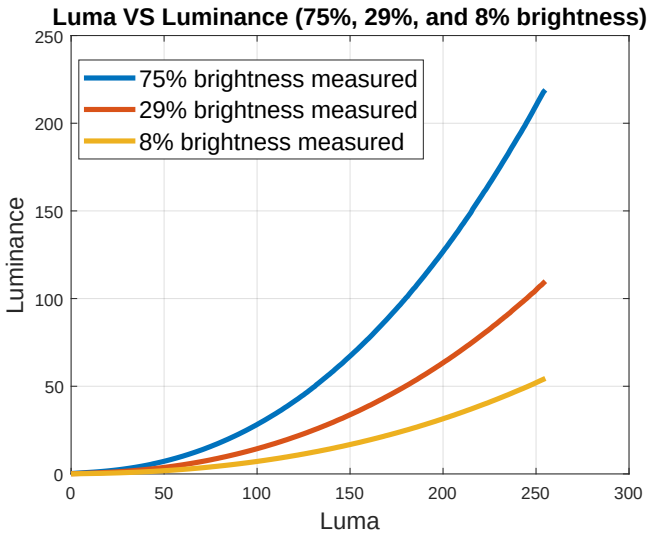

Figure S8.6: The relationship between luma value displayed on monitor and it luminance level in real world.

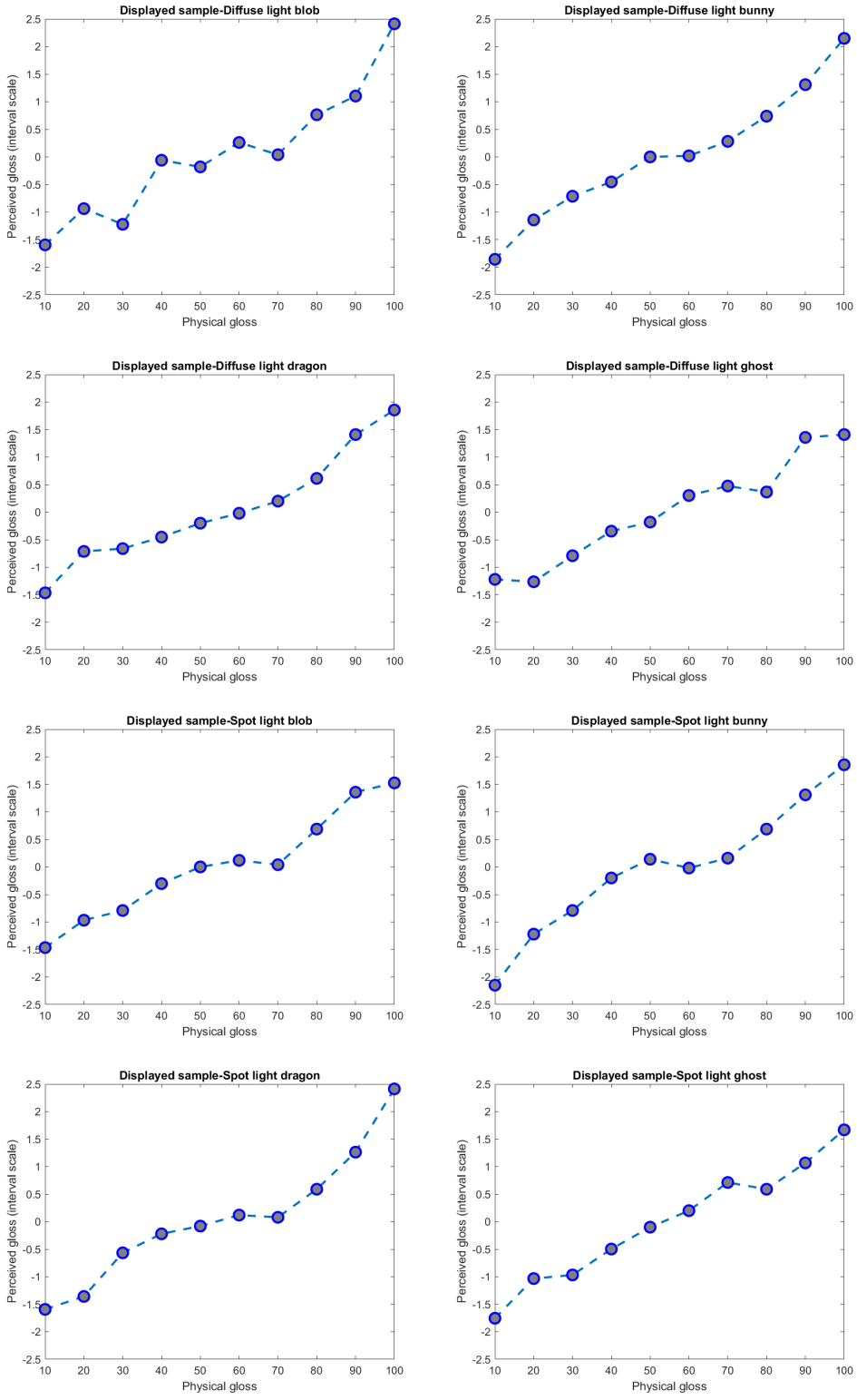

Figure S8.7: Results of the perceptual interval scaling (perceived gloss) obtained from the ordering (preliminary) experiment with Thurstone's Law [Thurstone 1927] for the displayed virtual samples broken up by illumination and geometry.

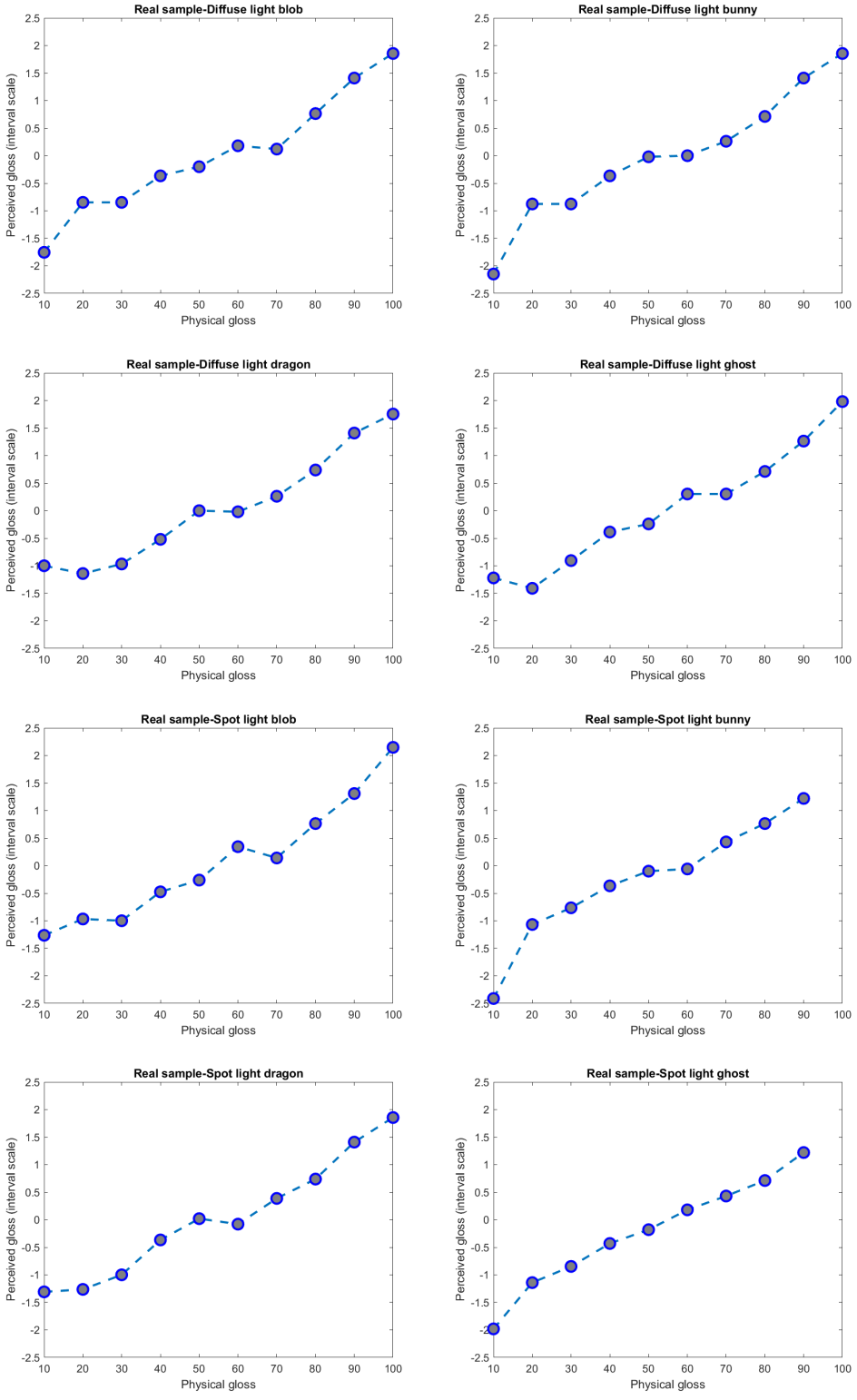

Figure S8.8: Results of the perceptual interval scaling (perceived gloss) obtained from the ordering (preliminary) experiment with Thurstone's Law [Thurstone 1927] for the real-world samples broken up by illumination and geometry. Note that under one spotlight, the perceived gloss values of bunny and ghost with physical gloss 100 are NaN due to zero in numerator.

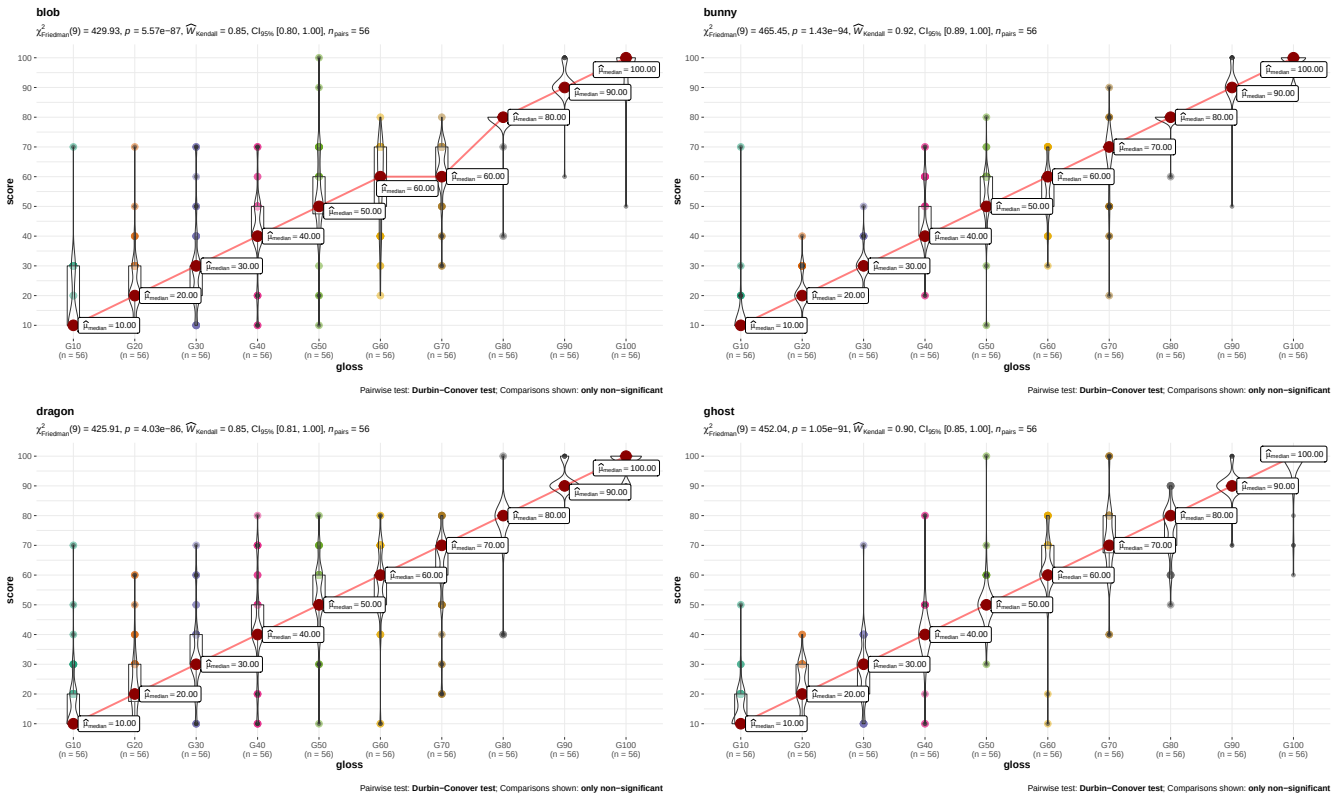

Figure S8.9: Violin plots and medians corresponding to the ordering (preliminary) experiment broken up by geometry. The figure includes results of Friendman's significance test with Durbin-Conover pairwise comparisons. Note that the comparisons marked are only the non-significant ones (i.e., all pairwise comparisons are significant).

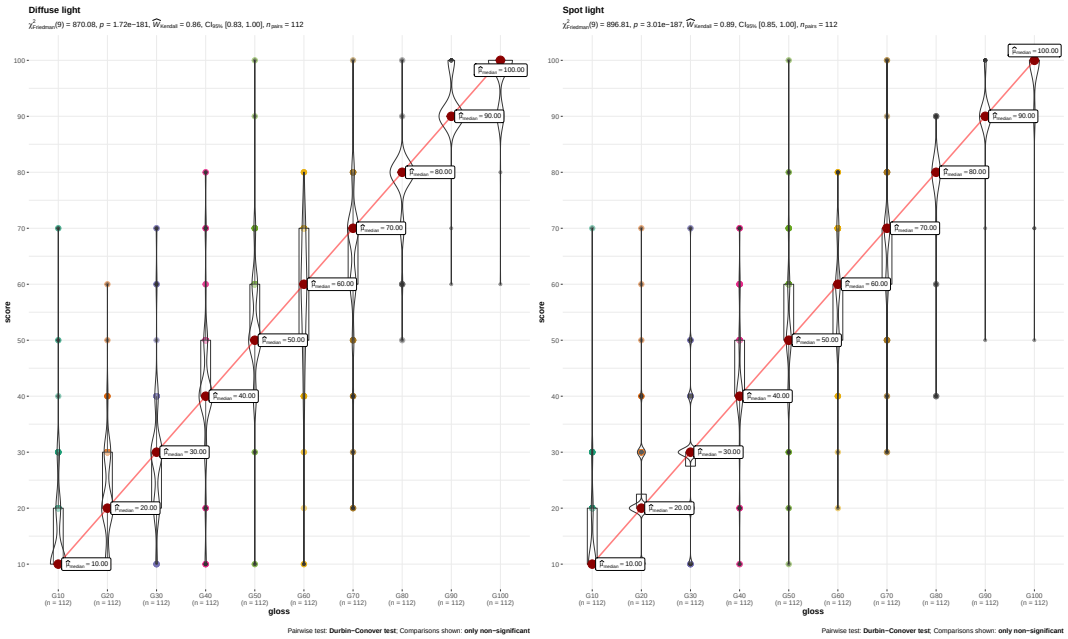

Figure S8.10: Violin plots and medians corresponding to the ordering (preliminary) experiment broken up by illumination. The figure includes results of Friendman's significance test with Durbin-Conover pairwise comparisons. Note that the comparisons marked are only the non-significant ones (i.e., all pairwise comparisons are significant).

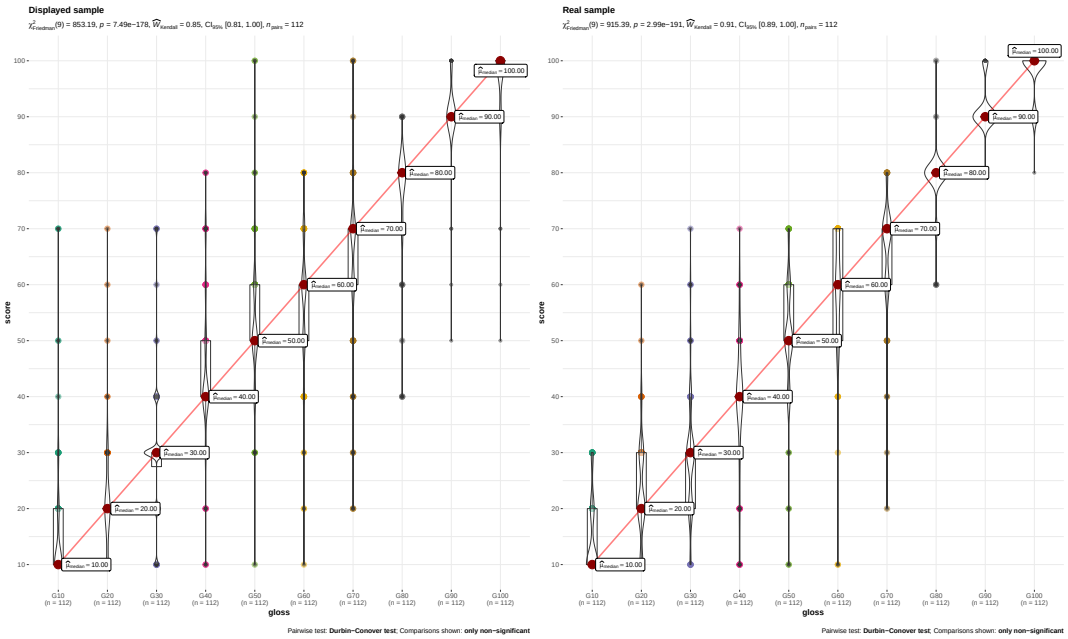

Figure S8.11: Violin plots and medians corresponding to the ordering (preliminary) experiment broken up by visualization condition (real-world sample or displayed sample). The figure includes results of Friendman's significance test with Durbin-Conover pairwise comparisons. Note that the comparisons marked are only the non-significant ones (i.e., all pairwise comparisons are significant).

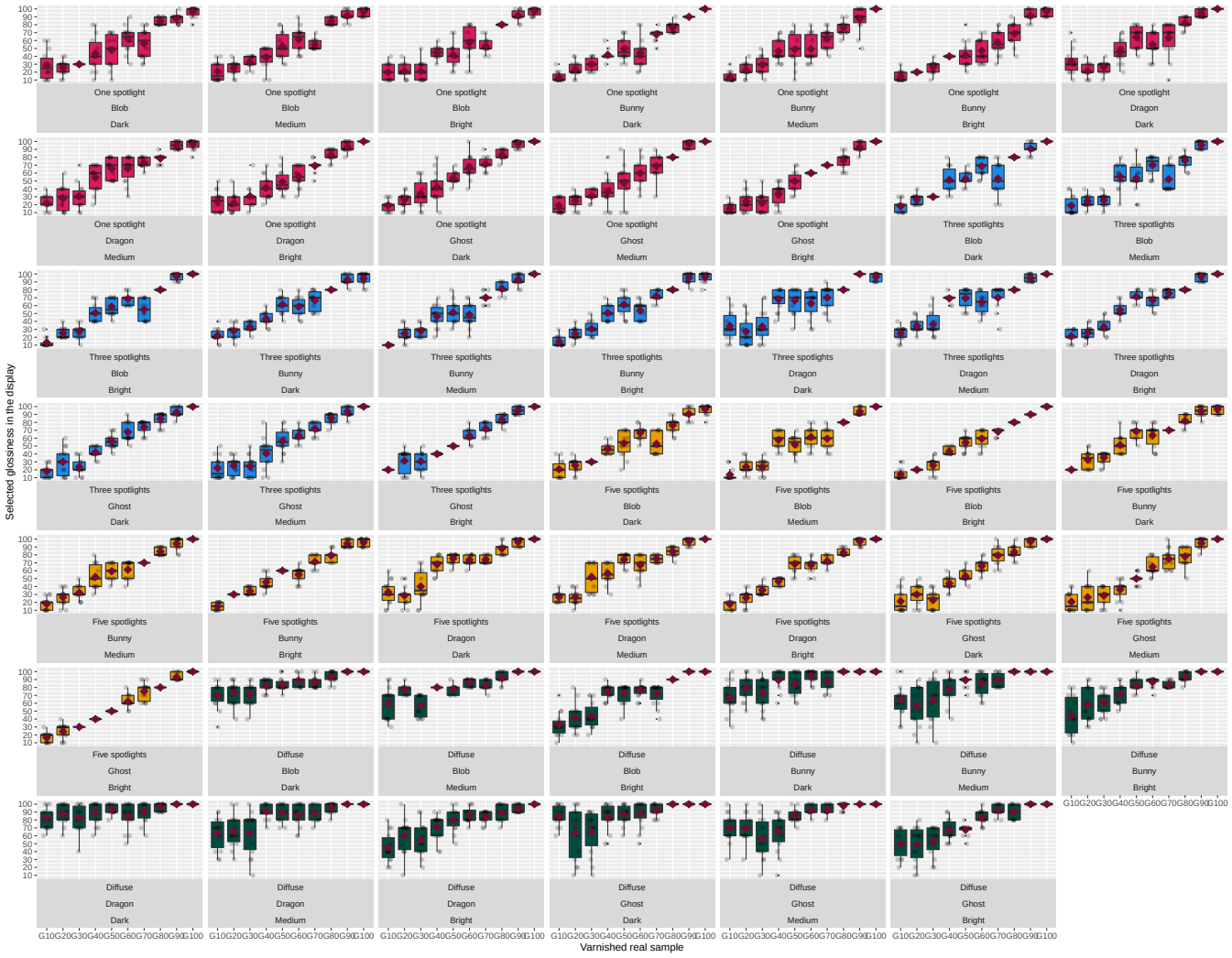

Figure S8.12: Boxplots and individual data points of our collected perceptual data from the main experiment broken up by each combination of conditions (4 illuminations $\times$ 4 geometries $\times$ 3 display brightness). The red diamonds represent the means.

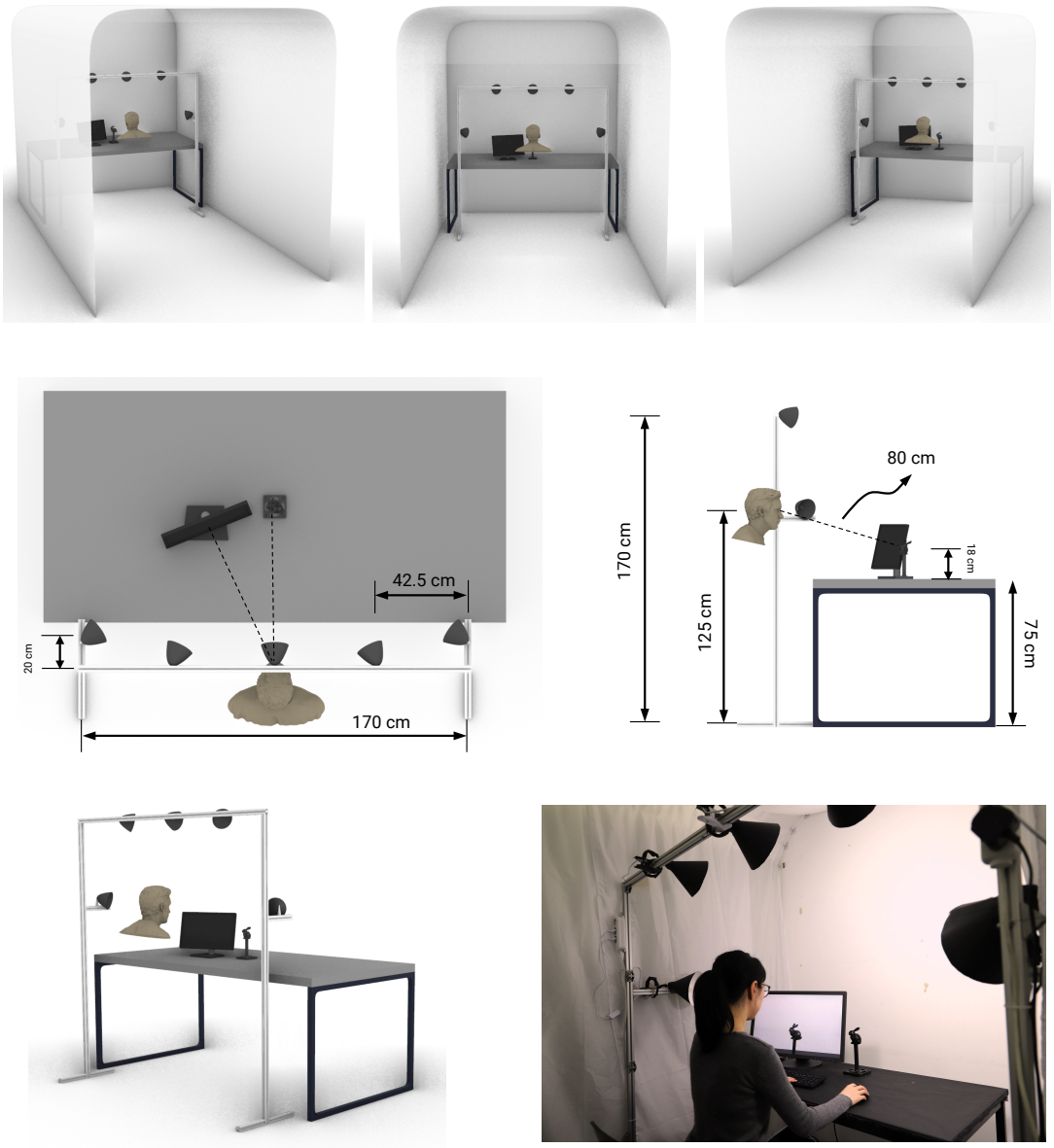

Figure S8.13: Schematics and a picture of our experimental setup. First row: Conceptual visualization of the experimental setup. Second row: The top and side views with information on the physical dimensions. Third row: 3D view on the light source setup and the actual photograph.

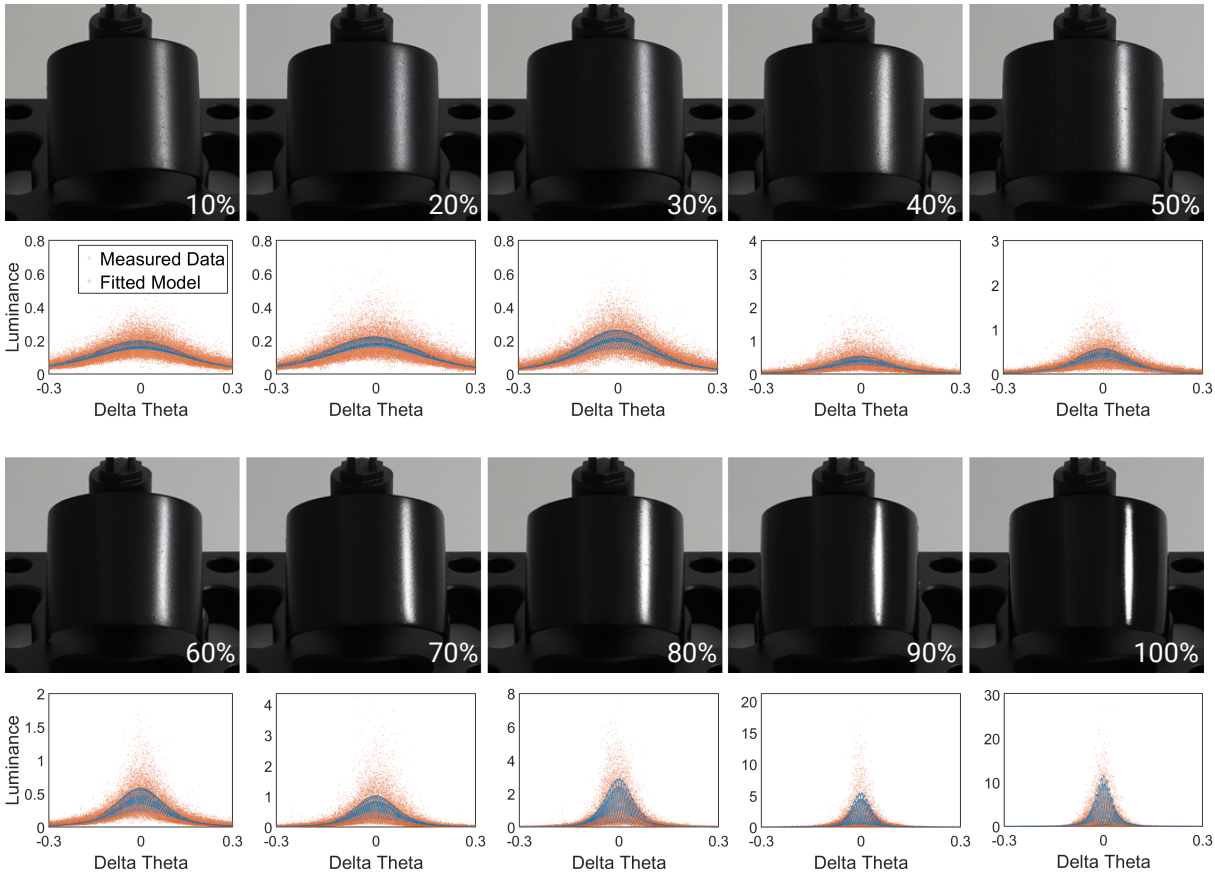

Figure S8.14: Manufactured cylindrical samples for reflectance measurement. We approximate the measured BRDF (orange) with a Cook-Torrance model with GGX distribution (blue).

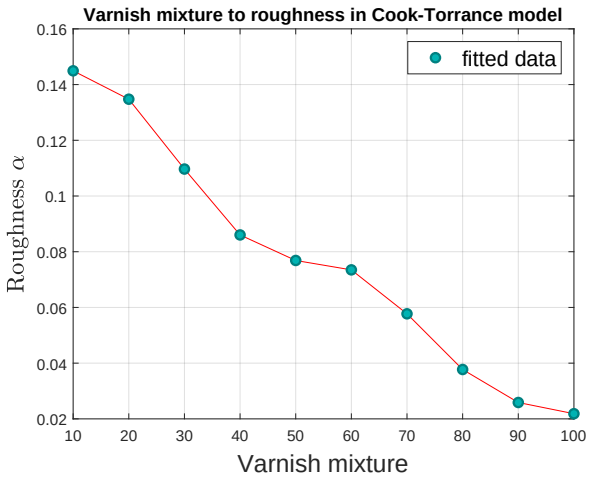

Figure S8.15: Mapping from varnish mixture to measured roughness is realized via a piece-wise linear interpolation of the ground truth values.

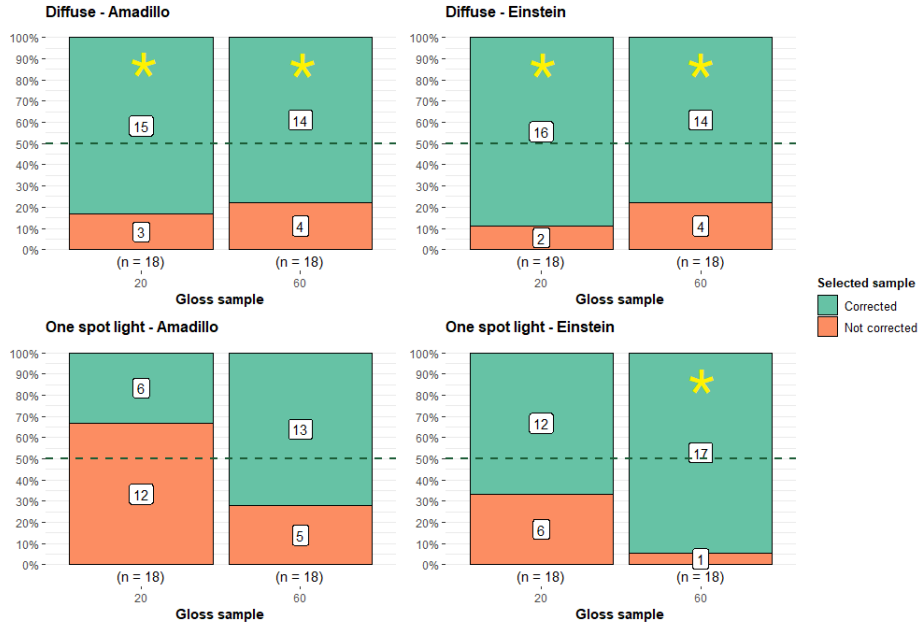

Figure S8.16: Results of the two-alternative forced choice study for evaluating our applications digital product design. Yellow stars mark significant differences (binomial test).

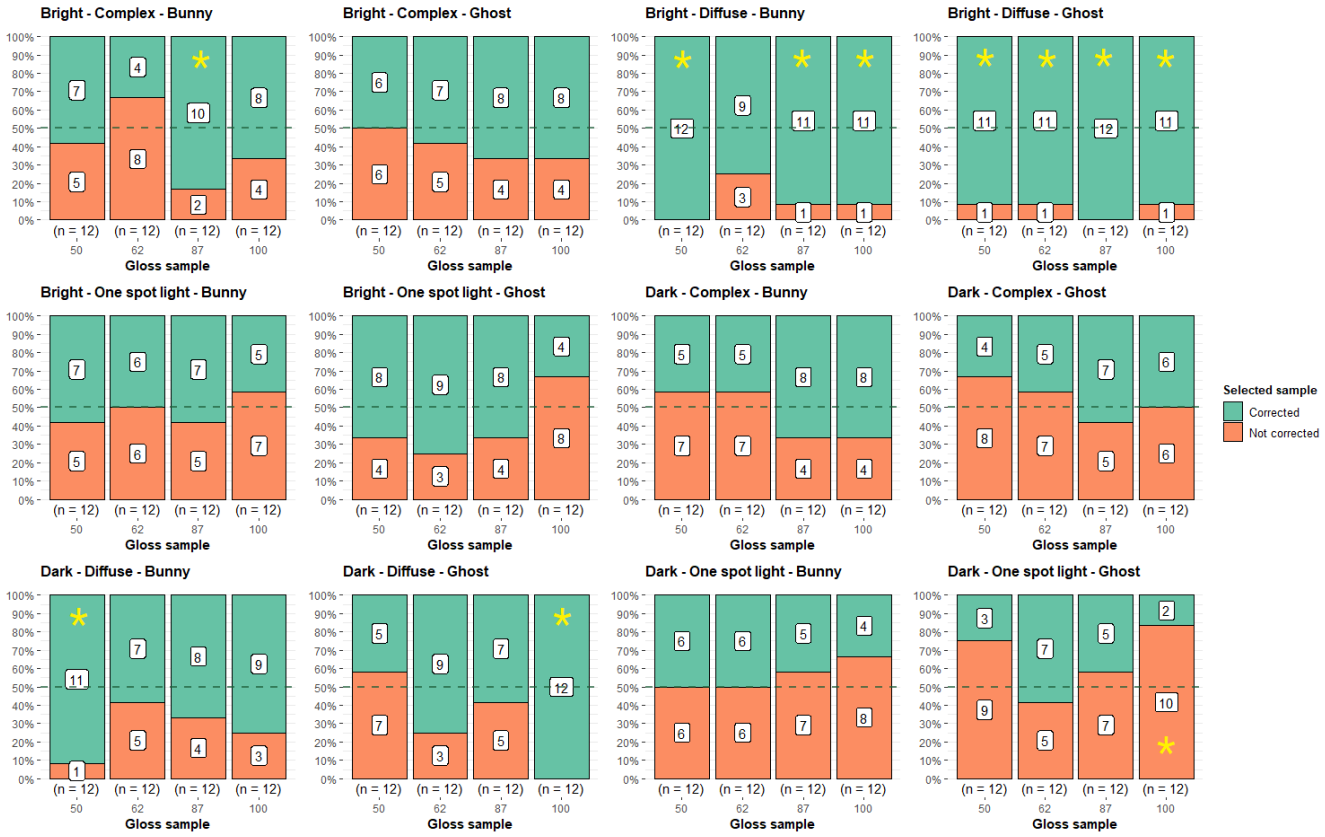

Figure S8.17: Results of the two-alternative forced choice study for evaluating our applications digitizing physical artifacts. Yellow stars mark significant differences (binomial test).

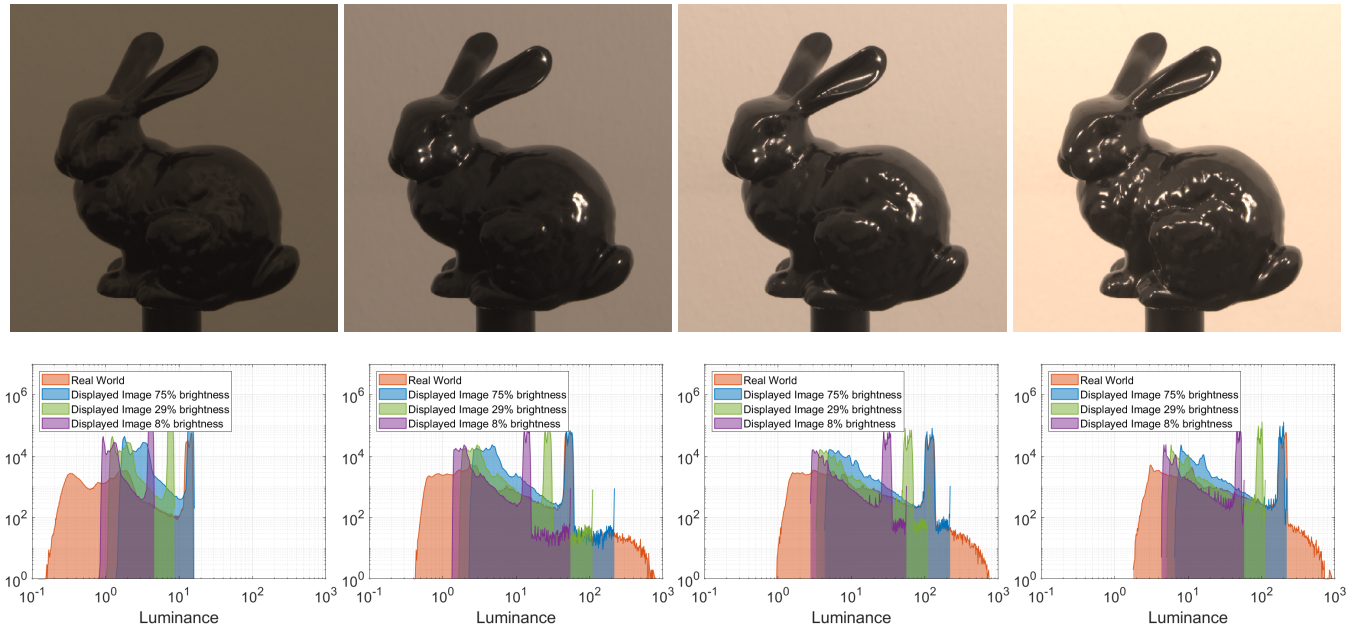

Figure S8.18: Luminance distribution of the real world and its displayed counterpart for our three display brightness levels. From left to right we show the diffuse, one spotlight, three spotlights, and five spotlights illumination scenarios.