Received: 12 October 2020 | Revised: 8 September 2021 | Accepted: 15 September 2021

DOI: 10.1002/col.22740

RESEARCH ARTICLE

COLOR RESEARCH AND APPLICATION WILEY

Lightness perception of structured surfaces

Felicitas V. Muth[1] | **Michael Heilig**[2] | **Dorothea Marquardt**[2] | **Linda Mittelberg**[2] | **Albrecht Sebald**[1] | **Wilfried Kunde**[1]

[1]Department of Psychology, Julius-Maximilians-University of Wuerzburg, Wuerzburg, Germany

[2]Das Kunststoff-Zentrum (SKZ), Friedrich-Bergius-Ring, Wuerzburg, Germany

Correspondence
Felicitas V. Muth, Department of Psychology, Julius-Maximilians-University of Wuerzburg, Roentgenring 11, 97070 Wuerzburg, Germany.
Email: felicitas.muth@uni-wuerzburg.de

Funding information
Allianz Industrie Forschung, Grant/Award Number: 20094 N

Abstract

Visual perception of surfaces is of utmost importance in everyday life. Therefore, it comes naturally, that different surface structures evoke different visual impressions in the viewer even if the material underlying these surface structures is the same. This topic is especially virulent for manufacturing processes in which more than one stakeholder is involved, but where the final product needs to meet certain criteria. A common practice to address such slight but perceivable differences in the visual appearance of structured surfaces is that trained evaluators assess the samples and assign a pass or fail. However, this process is both time consuming and cost intensive. Thus, we conducted two studies to analyze the relationship between physical surface structure parameters and participants visual assessment of the samples. With the first experiment, we aimed at uncovering a relationship between physical roughness parameters and visual lightness perception while the second experiment was designed to test participants' discrimination sensitivity across the range of stimuli. Perceived lightness and the measured surface roughness were nonlinearly related to the surface structure. Additionally, we found a linear relationship between the engraving parameter and physical brightness. Surface structure was an ideal predictor for perceived lightness and participants discriminated equally well across the entire range of surface structures.

KEYWORDS

appearance, color perception, maximum likelihood difference scaling, psychophysics, surface structure

1 | INTRODUCTION

The interplay of optical attributes of materials, namely color, gloss, translucency, and structure, determines how (visually unimpaired) humans perceive color.[1,2] Vision is typically the first, and in many cases most important, sensory channel when encountering new products. Even though substantial research has tried to address the question of how we perceive visual stimuli, the actual visual impression remains something scarcely understood as it is unique to the individual.[3,4] Among the most important factors shaping the overall appearance of an object is its surface. Not only is the structure itself a perceivable feature, rather small variations of surface properties affect other visual features such as brightness or color as well. In addition, material characteristics such as texture and gloss interact to evoke certain impressions in the viewer since visual stimuli are perceived as one conjoint object rather than a compilation of independent characteristics.[5] Consequently, two items of identical material characteristics, for example, material composition or pigmentation, can evoke completely different appearance impressions in the viewer if they differ in their surface structure.[6] Such variations in the visual impression are most likely caused by the context of the visual stimulus, that is, how it is presented and what other stimuli are presented alongside.[7] Commonly, humans perceive the highest luminance of an ambiguous illumination as white to anchor their estimation of the object's overall lightness.[8,9] Hence, perceived lightness is dependent on luminance, contrast and depth perception. This poses a challenge for polymer producing industries, since production-related variations in a material's surface also cause barely predictable variations in perceived color. Consequently, it is difficult to produce two materials that are not only objectively the same but also evoke the impression of sameness in the viewer.[10]

The state-of-the-art to address this issue is the assessment of material probes by specifically trained evaluators. Despite tolerance margins, the visual assessment often leads to the rejection of fully functional but visually not acceptable manufacturing parts. Consequently, the manufacturing part has to be disposed of and the production process has to restart. It often happens that there are multiple iterations before the final product actually meets the client's visual demands. Such reconciliation loops are time and money consuming and oftentimes superfluous.

Thus far, some research has been conducted on the interaction of lightness and glossiness, for example, Beck[11] found that glossy surfaces appear darker than matte surfaces and have a higher color constancy when illumination changes. In another study, perceived lightness was increased and the saturation of colored stimuli decreased with increasing specular roughness of the surface.[12] Additionally, there is a negative correlation between perceived gloss and specular roughness.[12,13] According to Toscani et al.,[14] humans profit from using the brightest parts of matte surfaces and rather dark parts of glossy surfaces when judging an object's albedo. They also indicate that humans ignore specular reflections on glossy objects, which might help explain why glossy surfaces are perceived as darker than matte surfaces. In contrast to this, to the best of our knowledge, little systematic research has been conducted to reveal how physical surface properties might affect lightness and color perception. For example, Luo et al.[15] used standardized measures to investigate how surface texture of fabrics influences instrumental color and found that irrespective of the structure the normalized reflectance curves were identical. Xiao and Brainard[16] asked participants to match colors to three-dimensional objects that varied in color and gloss. Results showed that the visual systems seem to compensate for some of the surface structure to extract the underlying body rather than simply translating cone excitation into a color-impression. However, such studies either do not take into account the human observer and their perception or use computer-generated stimuli, which do not necessarily mirror real-life interactions with such stimuli. Therefore, we aim at contributing to a more profound understanding of lightness perception of structured surfaces. To this end, we conducted two experiments with human observers to scrutinize the influence of surface structure on the visual perception of achromatic, opaque surfaces.

As color perception is multidimensional, we decided to concentrate on achromatic stimuli because the human visual system is more sensitive to variations in brightness variations of achromatic stimuli compared to brightness variations of colored objects.[17]

2 | EXPERIMENT 1

The first experiment aimed at systematically identifying differences in perceived lightness resulting from the variation of the surface's structure. Specifically, we intended to identify the point of subjective equality (PSE) as well as the just noticeable difference (JND) for a set of 19 achromatic polymer samples with varying surface structures, but identical pigmentation. In this study, we varied the surfaces of these samples by CO_2 laser processing. The laser was programmed to copy master stimuli of a test stimulus with varying structure onto the polymer samples, by changing the surface of the samples only. This is what we refer to as engraving sparsity—the more structure the less sparsity. That is, the CO_2 laser transferred areas with systematically varying shades of gray onto the plastic sheets by converting the respective gray shading into a specific structural pattern. By doing so, the visually perceived lightness of the sample actually changed as a function of various surface features (for more details refer to the Section 2.1.2). As the specific physical surface features that caused these lightness changes are yet unknown, we analyzed which physical surface features predicted the optometric and perceived lightness changes.

2.1 | Methods

2.1.1 | Participants

Thirty-two participants between 19 and 66 years (5 male, 1 left handed, mean age = 30.2, *SD* = 11.8) recruited from the university's participant pool (SONA) took part in the experiment. All participants had normal vision as measured with the Ishihara test for color blindness.[18]

Prior to the experiment, participants signed an informed consent form and they received either monetary compensation or partial course credit for their voluntary participation. All participants were naïve to the purpose of the study and were debriefed afterwards. The experiment was approved by the Ethics Committee of the psychological department of the Julius-Maximilians-University Würzburg (GZ 2019-05).

2.1.2 | Apparatus and stimuli

Viewing booth setup

The experiment took place in a dark laboratory with black walls, floor and ceiling to avoid extraneous light and reflections on the sample from disturbing the setup. During the experiment, participants were seated in front of a viewing booth (LED Color Viewing Light M[19]) at a distance of 35 cm. This distance granted participants a full view of the interior of the viewing booth without having to move while shielding potentially distracting input outside the box from their visual field. In order to ensure a constant viewing setup, the head was stabilized with a headrest that was equipped with a visual shield, which the experimenter lowered while changing the samples to avoid reflections. This way, we could guarantee controlled lighting (D65) on the samples at an angle of 45° and a viewing angle of 0°. Two buttons placed in front of the viewing booth were used to make judgments. Figure 1 shows the schematic setup of the experiment.

Polymer samples

Nineteen achromatic and opaque plates of the polymer acrylonitrile butadiene styrene (ABS) were used as visual stimuli in this study. We chose achromatic ABS plates, because the human visual system is most sensitive to changes in brightness.[17] They measured 100 × 100 × 3 mm and were gray due to the natural color of ABS itself as well as the additional coloring. The initially glossy surfaces of the ABS samples were structured by using a CO_2 laser (Speedy 100 C45[20]). The laser has a maximum output of 45 W and operates with a laser wavelength of 10.6 μm. As a result, the main effects of engraving and sublimation of the material dominate at this wavelength, whereby, as with metals, edge throwing due to thermal expansion or material bursting can occur and was observed.[21]

To prepare the ABS sample plates for laser application, they were cleaned and discharged with isopropanol and ionized compressed air. The JobControl software[22] was used to generate the structure by laser. Areas of systematically varying gray shading were printed into the laser software. Individual structures had the size of approximately 180 μm by 100 μm and were placed together by the laser parameters so that superordinate structures were created. These structures can also partially overlap. The lower the engraving sparsity, the more individual structures were lined up so that more area of the sample was processed (see Figure 2). The "engraving sparsity" as one of the laser process parameters was set between 30% and 90%. Based on graphical transparency (see Figure 3) (=engraving sparsity) values, corresponding sample structures were generated by the laser software. These structures were applied with 9 W laser power, 0.56 m/s speed, 250 dpi print resolution and 1000 PPI by laser-induced ablation and engraving with edge throw-up. Finally, the plates were cleaned with isopropanol in an ultrasonic bath to remove material residues and dried with ionized compressed air. Figure 2 shows surface pictures of three of the samples under a VHX-5000 microscope[23] illustrating the surface structure.

The range of these laser structures was set to ensure that all participants would classify the most extreme categories correctly. This was achieved by visual validation through experts who determined the engraving sparsity of the laser patches to vary between 30% and 90% in intervals of 5%. The mean value of 60% engraving sparsity was set as standard. A preliminary pilot study indicated the range between 50% and 70% engraving sparsity to be critical for determining the PSE and JND. Therefore, we varied engraving sparsity in intervals of 2% in this range.

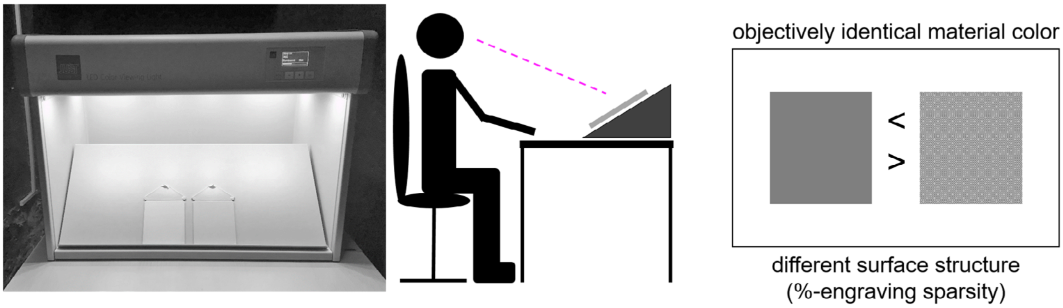


FIGURE 1 Left: Photo of the color viewing cabinet with two samples of identical material color with different surface structures (the left sample is the standard). Right: Sketch of the setup during the experiment

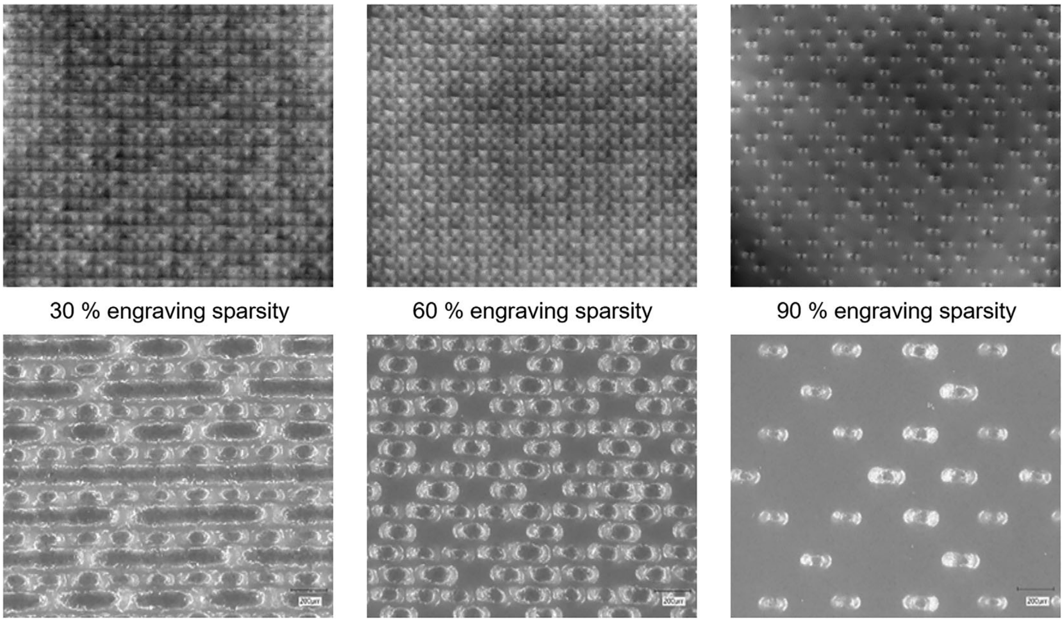


FIGURE 2 Top: Surface pictures of three of the polymer samples with 30%, 60%, and 90% engraving sparsity (from left) recorded with TRACEiT.[24] Bottom: Surface pictures of the same three samples under a VHX-5000 microscope (increased 200-fold)[23]

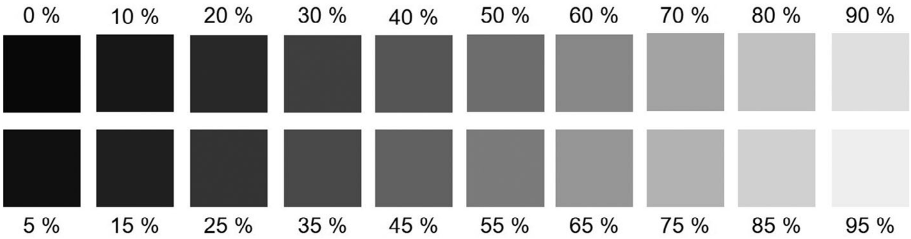


FIGURE 3 Graphical transparencies serving as input for the laser to generate structured surfaces (engraving sparsity)

The structured sample plates were characterized by microscopy and TRACEiT[24] with three white light optics for topography measurement. As there is no one single physical measure, which best reflects visual roughness, we used two common industrial standards to specify the surface roughness of the ABS plates.[25] From a multitude of parameters average roughness S_a and surface skewness S_{sk} of the standard ISO25178-2:2012-09[26] were chosen. S_a is the absolute mean of height that represents an overall measure of the texture comprising the surface, while S_{sk} describes the skewness, that is, refers to the surface's distribution symmetry of heights. S_a is defined as the arithmetic mean of the absolute ordinate values within the definition range A (1) and S_{sk} is defined as the quotient of the mean third power of the ordinate values and the third power of S_q within the definition range A (2). S_q is the mean square value of the ordinate values within the definition area A (3).

$$S_a = \frac{1}{A}\iint_A |z(x,y)|dxdy \tag{1}$$

$$S_{sk} = \frac{1}{S_q^3}\left(\frac{1}{A}\iint_A z^3(x,y)dxdy\right) \tag{2}$$

$$S_q = \sqrt{\frac{1}{A}\iint_A z^2(x,y)dxdy} \tag{3}$$

In order to compare the data and to obtain roughness parameters, the raw data was preprocessed and analyzed with the Software “Mountains Map 7.4” from Digital Surf. All measured profiles, called extracted profiles, with an area of 5 mm × 5 mm were leveled by fifth degree polynomic method to remove the nominal form of the planar work pieces. Using a low-pass filter λs to remove micro roughness was skipped, since this is also the case in most German automotive industry standards such as the VDA 2006:2003-7.[27] The primary profile obtained after leveling was filtered using a Gaussian L-filter with a cut-off wavelength of 0.25 mm. The main reason for using this filter is to separate long-scale components from short-scale components, in other words to separate waviness of the profile from roughness, and calculate roughness parameters according to the specification.[26] Finally, the roughness parameters S_a and S_{sk} were calculated according to Equations (1) and (2).

Spectral analysis of the samples, measured with BYK Additives & Instruments,[28] Spectro2guide 45°/0° under D65 illumination with the 10° observer over a measurement area with 8 mm diameter, provided us with the objective changes in brightness (L^*). For a reliable average throughout the surface L^* was calculated as the average of measurements at three different positions on the surface.

2.1.3 | Procedure

To assess the PSE and the JND, we used the method of constant stimuli.[29] Participants always saw two samples simultaneously. One of these (60% engraving sparsity) served as standard and did not change neither throughout the study nor between subjects. The second sample was the comparison stimulus that changed every trial. On each trial, participants were to judge which of the two samples was lighter in color. The allocation of the two buttons as well as the presentation side of the standard sample was counterbalanced across participants to control for side preferences and answer tendencies. Additionally, we incentivized conscientious participation by offering additional monetary compensation for correct classifications of perceived differences in lightness.* The experiment consisted of five blocks of 38 trials separated by short breaks, that is, each sample served twice as comparison probe in every block. Thus, participants rated each stimulus as comparison probe in 10 trials across the entire experiment. Raw data and supplemental material is available on the Open Science Framework, https://osf.io/xqm8r/.

2.1.4 | Data analysis

To assess the PSE and the JND of the polymer samples, we conducted a binary logistic regression analysis with the comparison stimulus as predictor. Therefore, we computed the proportion of trials in which the comparison plate was judged to be brighter than the standard stimulus as a function of the surface structure in %-engraving sparsity. PSEs were determined by identifying the comparison stimulus at which the psychometric function, that is, the logistic regression, yielded a likelihood of 50% judgments that the comparison stimulus was brighter than the standard. The two possible outcomes were coded as 0 (=comparison stimulus is brighter than standard) and 1 (=standard is brighter than comparison stimulus). We also report the JNDs, which were determined by identifying the %-engraving sparsity values of the comparison stimuli corresponding to the 25% and 75% thresholds and then dividing the difference between these values by two. These values indicate a change in surface structure that was reliably detected by the participants. Additionally, we conducted individual logistic regression analyses for each participant. Trials in which participants answered "too quickly," that is, 2.2 *SD* faster than on average trials, were discarded. We also discarded trials that the experimenter marked as erroneous during the experiment. These included trials in which participants made comments about pressing the wrong button or in single cases where the vision shield did not function properly. All analyses were conducted with IBM SPSS.[30]

For further assessment of the relation between participants' perceived lightness and physical roughness measures, we calculated Pearson correlations (4) between the industrial roughness measures S_a, S_{sk}, L^* and surface structure in %-engraving sparsity. This allows for the detection of linear trends and relationships between two variables. Per convention, a correlation of $r > .5$ is considered a strong relationship.

$$r = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\Sigma x^2 - \left(\sum x\right)^2\right]\left[n\Sigma y^2 - \left(\sum y\right)^2\right]}} \tag{4}$$

2.2 | Results

2.2.1 | Regression analysis

Trials with impossibly fast reaction times (0.5%—30 trials across all participants) as well as error trials marked by the experimenter (0.1%—6 trials across all participants) were excluded from the analyses. Results of the binary logistic regression indicated that there was a significant association between the structure of the surface and the perceived lightness of the ABS plate, $\chi^2(1) = 2683.20$, $P < .001$. The model coefficients can be retrieved from Table 1.Through logistic regression, we determined the PSE, this reflects the %-engraving sparsity value at which participants chose either stimuli to be brighter equally often. This is achieved by determining the x-value for the logistic regression when $y = 0.5$. Despite being statistically different the PSE at 56.7% engraving sparsity proved to be close to the actual point of equality at 60%, $t(31) = -3.71$, $P = .001$, $d_z = 0.66$. The JND, which is defined as the value at 25% and 75% discrimination strength, differed 9.1%-engraving sparsity from the PSE.† Within this range, participants could not reliably make out differences in lightness between the two stimuli. Figure 4 shows the individual regression slopes as well as the average model of the logistic regression across participants. Steeper slopes indicate better discriminatory capabilities.

TABLE 1 Model coefficients for the prediction of perceived brightness of the comparison stimulus

		95% CI for odds ratio		
	***b* (*SE*)**	**Lower**	**Odds ratio**	**Upper**
Included				
Constant	6.848^{*} (0.191)		942.0	
%-engraving sparsity	-0.121^{*} (0.003)	0.881	0.886	0.892

Note: $R^2_L = .68$, $R^2_{CS} = .36$, $R^2_N = .48$. Model $\chi^2(1) = 2683.20$, $P < .001$. $^{*}P < .001$.

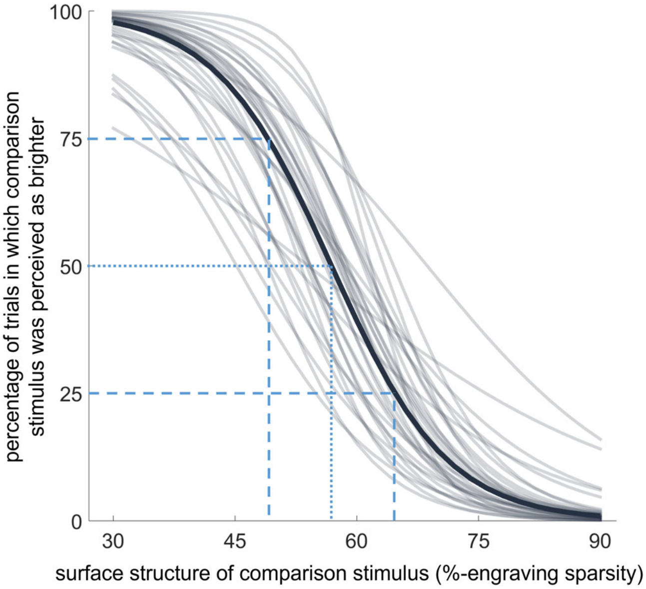


FIGURE 4 Individual regression slopes (light gray) for the relationship between surface structure and the perceived brightness of the comparison stimulus. The bold black line depicts the average logistic regression

2.2.2 | Correlation with physical roughness

Both the physical roughness parameters (S_a and S_{sk}) as well as brightness (L^*) were correlated with the samples' surface structure in %-engraving sparsity, using the Pearson correlation coefficient. Table 2 shows the Pearson correlations between the physical parameters and the surface structure. Additionally, Figure 5 plots all three parameters against the sample's surface structure in %-engraving sparsity. Only L^* as measure of the sample's brightness showed an almost perfect linear trend while the two other physical roughness parameters appeared to have an exponential relationship with surface structure.

TABLE 2 Pearson correlations between physical roughness parameters S_a and S_{sk} and brightness with the manipulated surface structure in %-engraving sparsity

	S_a	S_{sk}	L^*
Surface structure	$-.741^{**}$	$.895^{**}$	$-.990^{**}$
S_a		$-.921^{**}$	$.790^{**}$
S_{sk}			$-.923^{**}$

$^{**}P < .01$ (two-sided).

2.3 | Discussion

We investigated how and when differences in perceived lightness arise due to the variation of the surface structure. Experiment 1 served the purpose of determining the PSE and JND for the given gray ABS samples by using the method of constant stimuli. Participants simultaneously saw two identically colored stimuli with differing surface structures in a viewing booth and had to decide which one they perceived as being brighter. With increasing structure of the surface (in terms of engraving sparsity), the perceived lightness of the stimulus increases. This increase follows a psychophysical function. The JNDs at 65.8% and 47.6% engraving sparsity indicated that participants did not perceive lightness changes below a surface structure change of 10%-engraving sparsity. At 56.7% engraving sparsity, the PSE just slightly deviated from the objective PSE at 60% engraving sparsity. This observation probably reflects a response bias to judge the standard stimulus to be brighter.

As there is an almost perfect linear relationship between surface structure and the physically measured brightness, further analyses revealed that changes in perceived lightness could also be predicted by regressing perceived lightness on ΔL^* as measured with spectral analyses. These analyses also showed that changes in brightness could reliably be detected below a ΔE_{ab}^* of 0.8, which is usually said to have very mild influence on color perception.[31] This sensitivity mainly arose due to only minimal changes in Δa^* and Δb^* but rather substantial changes in ΔL^*. For achromatic colors changes in ΔE_{ab}^* as small as 0.3 can lead to altered color perception.[32] Consequently, the intended lightness change of the ABS plate was indeed achieved by varying the structure of the surface. A lower engraving sparsity lead to more surface structure, which lead to an increase in undirected gloss leading to increased measured and

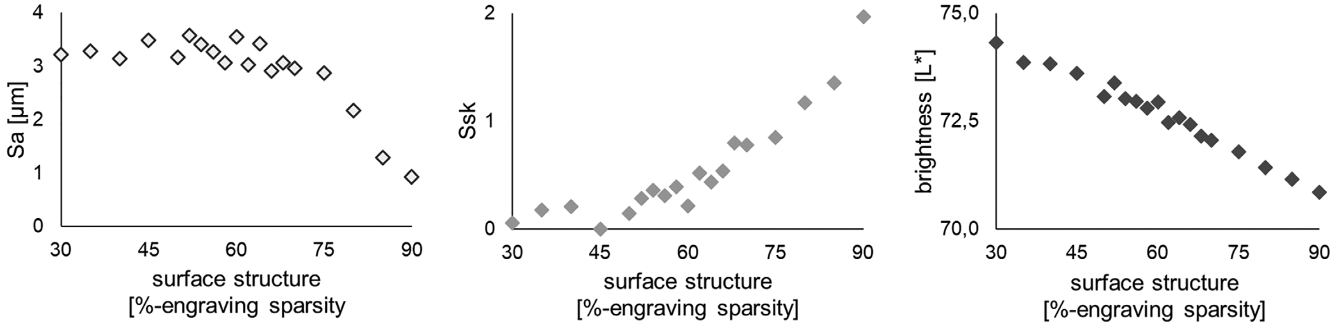


FIGURE 5 Scatterplots for S_a (left), S_{sk} (middle) and brightness (right) plotted against the stimulus' surface structure in %-engraving sparsity

perceived brightness. Physical measures of the surfaces showed that changes of the laser-manipulated engraving sparsity and thus objective brightness correlated reasonably systematic with the S_{sk} measure, whereas correlations with S_a were lower.

However, these results raise the question whether participants discriminate equally well across the whole spectrum of stimuli. In order to test this, we conducted a second experiment with the same ABS plates. Herewith, we intended to find out whether the empirically determined JND of 10%-engraving sparsity generalizes across the entire range of transparencies.

3 | EXPERIMENT 2

With the second experiment we intended to test whether participants discriminate equally well across the entire range of presented surface structures. To address this question, we used maximum-likelihood difference scaling (MLDS) for estimating the function that relates a physical parameter (in this case, surface structure) with its perceptual.[33] This method scales supra-threshold differences on a dimension free difference scale. To this end, participants are presented to pairs of stimuli and are asked to judge which pair presents the larger within-pair difference. Given the pattern of responses, it is possible to estimate the underlying maximum-likelihood perceptual scale. We applied this method to the perception of plastic plates that varied solely in their surface structure.

3.1 | Methods

3.1.1 | Participants

We recruited a new sample of 16 participants between 20 and 52 years (5 male, 1 left handed, mean age = 27.3, *SD* = 7.9) over the university's participant pool (SONA). All participants had normal vision as measured with the Ishihara test for color blindness.[18] Prior to the experiment, participants signed an informed consent form and received monetary compensation for their voluntary participation upon completion of the study. All participants were naïve to the purpose of the study and were debriefed afterwards. The experiment was approved by the Ethics Committee of the psychological department of the Julius-Maximilians-University Würzburg (GZ 2019-05).

3.1.2 | Apparatus and stimuli

In this experiment, we used the same viewing booth and polymer sample plates as in Experiment 1. The setup of the viewing booth was almost identical to the setup in the first study. However, this time four samples were presented simultaneously, thus the four samples were placed in a holdfast, which was then placed before the participants inside the viewing booth (see Figure 6). Additionally, the MLDS method calls for presentation of supra-threshold differences, thus, the sample size was reduced to seven polymer plates. As the results of Experiment 1 suggested that structure differences below 10% engraving sparsity are not reliably detected, only the samples with surface structure variations between 30% and 90% engraving sparsity were presented in steps of 10%.

3.1.3 | Procedure

The experiment was conducted in a dark room and participants saw two pairs of stimuli simultaneously in the viewing booth on every trial. Based on these two non-overlapping pairs (AB and CD) which always followed the logic A < B < C < D participants decided on each trial for which pair the within-pair difference, that is, AB or CD, was larger. To control for positional effects, the selection of the stimuli as well as the presentation

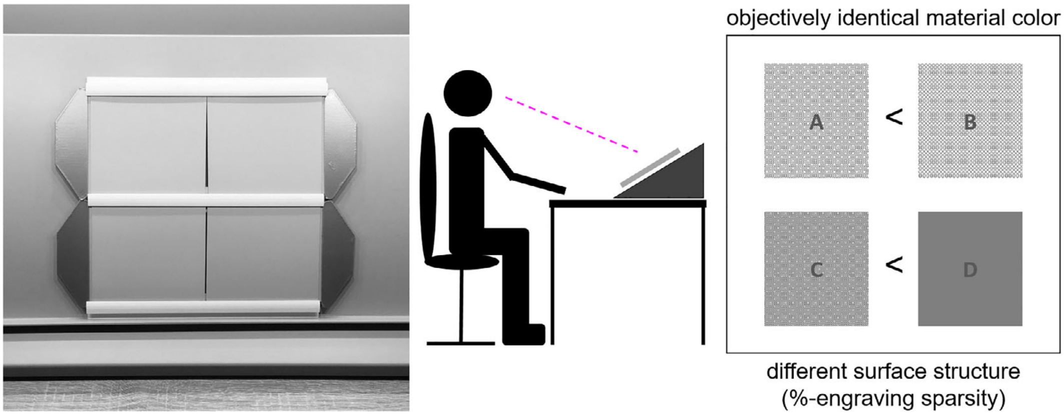


FIGURE 6 Left: Photo of two sample pairs within the color viewing cabinet (top: 30% and 40% engraving sparsity, bottom: 80% and 90% engraving sparsity). Right: Sketch of the setup during the experiment

location (top or bottom) were randomly selected taking into account the aforementioned criterion of A < B < C < D. The experiment consisted of four blocks during which each possible combination of pairs was presented once resulting in 35 trials per block. In-between blocks participants could take short breaks. Upon completion of the experiment, participants were asked for criteria which they used to make the judgments. Raw data and supplemental material is available on the Open Science Framework, https://osf.io/xqm8r/.

3.1.4 | Data analysis

The aim of this experiment was to assess the change in lightness perception of the polymer samples as a function of changes in the surface structure while the color itself was held constant. Therefore, we used the MLDS package for R[33] to estimate both the individuals perceptual scales for each participant and the mean across participants. It is a method employed to estimate perceptual scales. The model uses a stochastic model of how participants decide which of the displayed supra-threshold differences is greater. This results in scale values which capture the participant's judgment of the perceptual difference between two stimuli (for more detailed information see Boschman[34]; Maloney and Yang[35]).

Trials in which participants answered “too quickly,” that is, 2.2 *SD* faster than on average trials, were discarded. We also discarded trials that the experimenter marked as erroneous during the experiment. These included trials in which participants made comments about pressing the wrong button or in single cases where the vision shield did not function properly. The data of two participants had to be excluded from the analysis as they showed little to no consistency and discriminatory sensitivity in their answers and we thus have to assume that these participants did not take the task seriously. All analyses were conducted using R.[36]

3.2 | Results

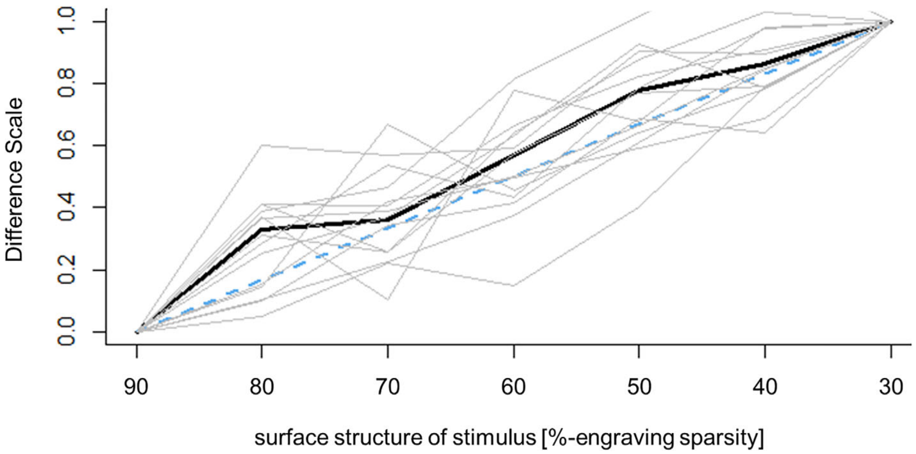
Trials with impossibly fast reaction times (0.4%—9 trials across all participants) were excluded from the analyses. Figure 7 plots the estimated perceptual scales for each individual subject (gray lines) as well as the mean across subjects (bold line). The data suggests that participants discriminated equally well across the entire engraving sparsity range.

In response to the question, which criteria were used to judge the within-pair differences of the polymer samples, most participants answered with visible differences in surface structure (33%) and reflected light (33%). The observations that surface structure was named as criterion to judge stimulus similarity suggests that engraving sparsity came with visible changes of surface structure on top of changes on brightness. Another quarter named surface patterns as criterion to base their decisions on. Moreover, a substantial number of participants ($n = 5$) also reported that they decided intuitively for one or the other pair without paying attention to specific characteristics.

3.3 | Discussion

We examined whether participants' changes in lightness perception are stable across the entire range of stimuli used in the presented study. Experiment 2 served the purpose of unveiling potential areas of higher or lower sensitivity for perceived lightness. Participants simultaneously saw two pairs of identically colored stimuli differing only in surface structure. These pairs were presented in a viewing booth and participants were to decide for which of the pairs the

FIGURE 7 Individual estimated perceptual scales as a function of stimulus transparency using maximum-likelihood-difference-scaling (MLDS) shown in gray. The mean across subjects is also plotted (black line)



within-difference was greater. There was an almost linear trend in participants' ability to discriminate between the sample pairs. This observation indicates that participants' difference perception was fairly constant across the entire range of ABS samples. Thus, it can be assumed that the JND determined in Experiment 1 generalizes across other surface structures of the same material and color.

When asked for decision criteria, participants predominantly reported surface structure, the stimuli's surface pattern as well as the reflected light to judge the differences. The presentation of the samples within the viewing booth, however, was possibly not ideal, as the light might not have spread equally across both sample pairs. Such differences in illumination could have influenced participants' judgments. To minimize this effect, we randomly presented the pairs at either the top or the bottom so each pair was judged at both positions.

4 | GENERAL DISCUSSION

With the two experiments, we aimed at contributing to the still very sparse understanding of perceived changes of lightness of surfaces as a function of surface structure. Specifically, we systematically manipulated the surface structure of achromatic ABS plates in %-engraving sparsity to study its influence on the perception of lightness of the respective sample plates. Overall, our surface variation resulted in the predicted changes of lightness perception. That is, increasing surface structure reliably resulted in more perceived lightness of the respective stimuli. Structure changes below 10% engraving sparsity were not reliably detected while supra-threshold changes were equally well discriminated across the entire range of stimuli. Among the various spatial parameters of surface, S_{sk} turned out to be a reasonable predictor for both objective and subjective lightness changes. In other words, by measuring S_{sk} it was possible to predict reasonably well, which perceptual lightness changes occur when S_{sk} changes. The correlation of skewness of the luminance histogram with the perceived lightness was shown before by Motoyoshi et al.,[37] however solely for synthetic images of stucco-like surfaces or, as shown by Landy,[38] for manipulated images of peaches but not yet for real surface samples. In a follow-up study, Sharan et al.[39] found the reflectance histograms considering interreflections and surface gloss to be a good predictor for human lightness perception. It is important to see that this correlation also applies for real surfaces. It may be useful, especially for cases in which optometric analyses of samples is not possible. However, it should also be noted that the perceived lightness changes in this study were almost perfectly, and in this case linearly, correlated with the optometric L^* measurements. So for precise quality assessment of materials regarding brightness, the L^* measure as used here, is certainly recommendable. Moreover, we can say that lightness changes caused by surface structure of more than $\Delta L^* = 0.3$ are discriminated by observers reliably, and have to be taken into account, if lightness equivalence is a strong quality criterion. Variations below that range are not that reliably discriminated. Experiment 2 showed that the discrimination accuracy is very similar across the entire L^* range studied here. Additionally, future studies will probably profit from correlating measured and perceived lightness with image statistics or BRDF (eg, Ashikmin et al.[40]) as this should contribute to a better understanding of how surface structures influence the visual perception of lightness.

These results bring about a multitude of possible implications for the plastic industry. First and foremost, studying physical roughness parameters and their influence on changes in visual perception could possibly render visual assessment by specially trained evaluators superfluous in the future. Our results suggest that human observers do not reliably detect lightness changes caused by structure variations below $\Delta L^* = 0.3$. Additionally, of the multitude of physical roughness parameters, in our study skewness had the largest influence on perceived color differences. Thus, manufacturers should put special emphasis in keeping the skewness constant across different charges. As the presented findings can help to reduce unnecessary excess production in the future, additional research needs to assess the generalizability of the reported results to additional surface structures.

Second, knowledge about changes in lightness perception due to variations in the surface structure can not only inform perception of achromatic plastics but also give first insights, how color perception is affected by surface changes. However, it is to be expected that surface structure variations have different influence on the perceived color depending on the inherent color of the plastic. The changes in green materials can probably be larger than those in blue materials for example.[41] Consequently, to extend the study's findings, future studies should include colored materials. It will be interesting to see how the addition of color contributes to the perception of lightness.

Additionally, in this study, we solely focused on visual perception of the given plastic plates. However, product perception is a multifaceted oeuvre and no single domain suffices to represent it entirely. Thus, future research should be extended to the interaction between haptic and visual product characteristics.

ACKNOWLEDGMENTS

The project 20094 N by the research association “Foerdergemeinschaft fuer das SKZ” was funded through the Arbeitsgemeinschaft industrieller Forschung e. V. (AiF) within the scope of the program for the promotion of cooperative industrial research and development by the Federal Ministry for Economic Affairs and Energy due to a decree by the German Bundestag. We would like to thank for the financial support. Open Access funding enabled and organized by Projekt DEAL.

ENDNOTES

[*] As participants did not receive feedback after individual decisions but rather at the end of the experiment, rewarding participants for certain classifications could not have led to learning certain strategies. Moreover, it is common to incentivize participants to motivate them to do their best. Thus, after careful consideration we decided for additional monetary compensation between 1-3 Euros. This procedure was approved by the university's ethics committee.

[†] The JND reported here stems from the fitted model. Across participants the average range of % engraving sparsity within which participants could not reliably discriminate was 17.32% ($SD = 6.79$).

DATA AVAILABILITY STATEMENT

All raw data and supplemental material is available on the Open Science Framework, https://osf.io/xqm8r/.

ORCID

Felicitas V. Muth https://orcid.org/0000-0001-9706-8578

REFERENCES

[1] Lozano RD. A new approach to appearance characterization. *Color Res Appl*. 2006;31(3):164-167.

[2] Hunt Robert W. G., Pointer Michael R. RGB Colorimetry. *Measuring Color*. 4th ed. Chichester, UK: John Wiley & Sons; 2011:231-240.

[3] Schifferstein HNJ. The perceived importance of sensory modalities in product usage: a study of self-reports. *Acta Psychol (Amst)*. 2006;121(1):41-64. https://doi.org/10.1016/j.actpsy.2005.06.004

[4] Kuehni RG. Color difference formulas: an unsatisfactory state of affairs. *Color Res Appl*. 2008;33(4):324-326.

[5] Ho Y-X, Landy MS, Maloney LT. Conjoint measurement of gloss and surface texture. *Psychol Sci*. 2008;19(2):196-204.

[6] Bastian M, Hochrein T. *Einfärben von Kunststoffen: Produktanforderungen-Verfahrenstechnik-Prüfmethodik*. Carl Hanser Verlag GmbH Co KG; 2nd ed. München: Carl Hanser Verlag GmbH Co KG; 2018.

[7] Gilchrist AL. Perceived lightness depends on perceived spatial arrangement. *Science*. 1977;195(4274):185-187.

[8] Gilchrist A, Kossyfidis C, Bonato F, et al. An anchoring theory of lightness perception. *Psychol Rev*. 1999;106(4):795.

[9] Anderson BL. The perceptual representation of transparency, lightness, and gloss. In: Wagemans J, ed. *Handbook of Perceptual Organization*. Oxford, UK: Oxford University Press; 2015:466-483.

[10] Leiby FA. Visual color matching of plastic materials. In: Charvat RA, ed. *Coloring of Plastics: Fundamentals*. Hoboke, NJ: John Wiley & Sons; 2005.

[11] Beck J. The effect of gloss on perceived lightness. *Am J Psychol*. 1964;77(1):54-63.

[12] Isherwood ZJ, Huynh-Thu Q, Arnison M, et al. Surface properties and the perception of color. *J Vis*. 2021;21(2):7. https://doi.org/10.1167/jov.21.2.7

[13] Honson V, Huynh-Thu Q, Arnison M, Monaghan D, Isherwood ZJ, Kim J. Effects of shape, roughness and gloss on the perceived reflectance of colored surfaces. *Front Psychol*. 2020;11:485. https://doi.org/10.3389/fpsyg.2020.00485

[14] Toscani M, Valsecchi M, Gegenfurtner KR. Lightness perception for matte and glossy complex shapes. *Vision Res*. 2017;131:82-95.

[15] Luo L, Tsang KM, Shen H-L, Shao S-J, Xin JH. An investigation of how the texture surface of a fabric influences its instrumental color. *Color Res Appl*. 2015;40(5):472-482.

[16] Xiao B, Brainard DH. Surface gloss and color perception of 3D objects. *Vis Neurosci*. 2008;25(3):371-385. https://doi.org/10.1017/S0952523808080267

[17] Hunt RWG. Light and dark adaptation and the perception of color. *JOSA*. 1952;42(3):190-199.

[18] Stilling J, Hertel E, Velhagen K. *Tafeln zur Prüfung des Farbensinnes/Farbensehens*. 35th ed. Georg Thieme Verlag: Stuttgart, NY; 2016.

[19] JUST Norm Licht GmbH. LED Color Viewing Light 2.0; 2020. https://www.just-normlicht.com/en/articlelist.html?id=15&name=LED-Color-Viewing-Light-2.0. Accessed April 20, 2020.
[20] Trotec Laser GmbH. Speedy 100 C45. Marchtrenk, Austria.
[21] Lake M. *Oberflächentechnik in der Kunststoffverarbeitung: vorbehandeln, beschichten, bedrucken, funktionalisieren, prüfen*. München: Carl Hanser Verlag GmbH Co KG; 2016.
[22] JobControl(R) [computer program]. https://www.troteclaser.com/en-au/laser-machines/laser-software/jobcontrol/. Accessed August 12, 2020.
[23] Keyence Germany GmbH. VHX-5000 microscope.
[24] Innowep GmbH. TRACEiT(R).
[25] Bergmann Tiest WM, Kappers AM. Haptic and visual perception of roughness. *Acta Psychol (Amst)*. 2007;124(2):177-189. https://doi.org/10.1016/j.actpsy.2006.03.002
[26] ISO 25178-2:2012-09. Geometrical product specifications (GPS)—Surface texture: Areal—Part 2: Terms, definitions and surface texture.
[27] VDA 2006:2003-7. Geometrical Product Specifications—Surface Texture—Rules and Procedures for the Assessment of Surface Texture.
[28] BYK-Gardner GmbH. spectro2guide.
[29] Gescheider GA. *Psychophysics: The Fundamentals*. 3rd ed. Lawrence Erlbaum Associates: Mahwah, NJ; 1997.
[30] IBM SPSS Statistics for Windows [computer program]. Version 25.0. Armonk, NY: IBM Corp; 2017
[31] Klein GA, Meyrath T. *Industrial Color Physics*. New York: Springer; 2010.
[32] Kirchner EJ, Ravi J. Setting tolerances on color and texture for automotive coatings. *Color Res Appl*. 2014;39(1):88-98.
[33] Knoblauch K, Maloney LT. MLDS: maximum likelihood difference scaling in R. *J Stat Softw*. 2008;25(2):1-26.
[34] Boschman MC. DifScal: a tool for analyzing difference ratings on an ordinal category scale. *Behav Res Methods Instrum Comput*. 2001;33(1):10-20.
[35] Maloney LT, Yang JN. Maximum likelihood difference scaling. *J Vis*. 2003;3(8):5.
[36] *R: A Language and Environment for Statistical Computing [computer program]*. Vienna, Austria: R Foundation for Statistical Computing; 2019. https://www.R-project.org/. Accessed July 1, 2020
[37] Motoyoshi I, Nishida S, Sharan L, Adelson EH. Image statistics and the perception of surface qualities. *Nature*. 2007; 447(7141):206-209. https://doi.org/10.1038/nature05724
[38] Landy MS. A gloss on surface properties. *Nature*. 2007; 447(7141):158-159.
[39] Sharan L, Li Y, Motoyoshi I, Nishida S, Adelson EH. Image statistics for surface reflectance perception. *J Opt Soc Am A*. 2008;25(4):846-865. https://doi.org/10.1364/JOSAA.25.000846
[40] Ashikmin M, Premože S, Shirley P. A microfacet-based BRDF generator. In: Brown JR, Akeley K, eds. *Computer Graphics: SIGGRAPH 2000 Conference Proceedings, July 23–28, 2000: Electronic Art and Animation Catalog. Bog + 1 cd-rom i lomme*. New York: ACM Press; 2000:65-74.
[41] Witzel C, Gegenfurtner KR. Categorical sensitivity to color differences. *J Vis*. 2013;13(7):1. https://doi.org/10.1167/13.7.1

AUTHOR BIOGRAPHIES

Felicitas V. Muth completed her studies in psychology at the Julius-Maximilians-University in Würzburg in 2018. Since then she has been a researcher and PhD candidate at the chair for experimental psychology at the University of Würzburg, Germany.

Michael Heilig completed studies in functional materials at the Julius-Maximilians-University in Würzburg in 2014. Since 2014, he has been working as a research assistant at the Plastics Centre SKZ, Germany, in the area of joining and surface technology.

Dorothea Marquardt received a PhD in chemistry from the Heinrich-Heine-Universität of Düsseldorf, Germany, in 2012. She worked as a research associate in the spectroscopy research group at the Plastic Centre SKZ.

Linda Mittelberg received a MS in physics from the Westfälische Wilhelms-Universität of Münster, Germany, in 2012 and a PhD in physics in 2016. She is currently leading the spectroscopy research group at Plastics Centre SKZ in Würzburg, Germany.

Albrecht Sebald received a degree in psychology from the University of Würzburg, Germany, in 1983, and a PhD in psychology from the University of Würzburg, Germany, in 1986. He is a researcher at Julius-Maximilians-University Würzburg.

Wilfried Kunde received a degree in psychology from the University of Hamburg, Germany, in 1994, and a PhD in psychology from the University of Würzburg, in 1998. He is currently a full Professor of Experimental Psychology at the University of Würzburg.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Muth FV, Heilig M, Marquardt D, Mittelberg L, Sebald A, Kunde W. Lightness perception of structured surfaces. *Color Res Appl*. 2022;47(2):377-387. doi:10.1002/col.22740