Disparity, motion, and color information improve gloss constancy performance

Gunnar Wendt
Institute of Psychology, Kiel, Germany

Franz Faul
Institute of Psychology, Kiel, Germany

Vebjørn Ekroll
Institute of Psychology, Kiel, Germany

Rainer Mausfeld
Institute of Psychology, Kiel, Germany

S. Nishida and M. Shinya (1998) found that observers have only a limited ability to recover surface-reflectance properties under changes in surface shape. Our aim in the present study was to investigate how the degree of surface-reflectance constancy depends on the availability of information that may help to infer the reflectance and shape properties of surfaces. To this end, we manipulated the availability of (i) motion-induced information (static vs. dynamic presentation), (ii) disparity information (with the levels “monocular,” “surface disparity,” and “surface + highlight disparity”), and (iii) color information (grayscale stimuli vs. hue differences between diffuse and specular reflections). The task of the subjects was to match the perceived lightness and glossiness between two surfaces with different spatial frequency and amplitude by manipulating the diffuse component and the exponent of the Phong lighting model in one of the surfaces. Our results indicate that all three types of information improve the constancy of glossiness matches both in isolation and in combination. The lightness matching data only revealed an influence of motion and color information. Our results indicate, somewhat counterintuitively, that motion information has a detrimental effect on lightness constancy.

Keywords: gloss constancy, gloss perception, specular highlights


Introduction

Perceptual objects, as they appear in the phenomenal world of an observer, are characterized by several attributes, like size, shape, and color. In addition, material qualities like “soft,” “fluid,” “brittle,” or “smooth,” i.e., attributes that constitute the consistency or surface properties of an object, can obviously be inferred by the visual system without difficulty (Adelson, 2001; Mausfeld, in press). One of the most prominent research topics in this area is the perception of gloss—a field of investigation that has already been established in the middle of the 19th century (Dove, 1850; Wendt, 2010; see also Harrison, 1945). The attribute of glossiness is tied to a variety of different perceptual objects or materials, like metals, glass, plastics, or textiles, differing in both qualitative (Bixby, 1928) and quantitative aspects (e.g., Billmeyer & O’Donnell, 1987; Ferwerda, Pellacini, & Greenberg, 2001; Obein, Knoblauch, & Viénot, 2004). However, what exactly determines the degree to which a surface appears glossy?

Physically, gloss is characterized as a certain reflection property of surfaces, described by the bidirectional reflectance distribution function (BRDF, cf. Nicodemus, Richmond, Hsia, Ginsberg, & Limperis, 1977). The reflective behavior of surfaces ranges from ideal diffuse (e.g., chalk or paper that approximately shows an ideal Lambertian reflection) to ideal specular (perfect mirrors). However, most object surfaces reflect the incident light in a manner intermediate between these extreme types of reflection, resulting from a mixture of both (Christie, 1986).

Although such surfaces usually appear more or less glossy, there is also a strong confounding effect of the illumination (Pont & te Pas, 2006; Sève, 1993). Fleming, Dror, and Adelson (2003) investigated how the gloss impression depends on properties of the prevailing illumination. They rendered spheres using different “illumination maps” (Debevec, 1998) to simulate illuminations of varying complexity, ranging from global illumination to a single point-light source (the stimuli were similar to those shown in the top row in Figure 1). They found that subjects could match the gloss parameters of a test sphere in a comparison sphere more accurately if the test sphere was illuminated with a real-world illumination map than when point lights or artificial illuminant maps were used. A limitation of this study is the restriction on spheres as objects, because it is well known that object shape may also influence gloss perception. This can be appreciated, for instance, by comparing the upper and lower rows in Figure 1. Thus, it is well possible that the relatively poor
matching performance found in the conditions with point-light sources is, at least in part, due to the limited information provided by spherical objects in this case.

Intense, localized light sources as the sun or a light bulb lead to isolated highlights in the proximal stimulus. There may be a single highlight as on a sphere or a distributed pattern of light on more complex surfaces (see Figure 1). One can distinguish several aspects of highlights: First, their position and form that depend mainly on the shape of the surface and the relative positions of observer and light source. Second, the blurring of the highlight contour that mainly depends on the material properties described by the BRDF. Third, their (relative) intensity that depends on the intensity of the illumination and the relative degree of diffuse and specular reflections.

Empirical findings indicate that the visual system uses the latter two properties as cues for the glossiness of the surface (Beck & Prazdny, 1981; Forbus, 1977; Hunter, 1975; Wendt, Faul, & Mausfeld, 2008). This seems reasonable, because they depend on the material properties. There are, however, confounding effects that also influence the spatial extent and blur of highlights. A first confounding effect that we shall not deal with in the present paper is implied by the fact that highlights are more or less blurred images of light sources. Thus, the size of highlights and the intensity distribution within them does in general also depend on the size and the intensity distribution inside the light source itself.

In this paper, we focus on a second confounding effect, which stems from the 3D surface geometry: A jagged surface would produce smaller highlights with sharper contours than a smoother surface with lower local curvatures, even if both surfaces have the same BRDF (Figure 2a). This interaction of several influencing factors at the physical stage—which affect the generation of the proximal stimulus—raises the question of whether the visual system is able to reliably determine based on highlights—the reflective properties of a surface independently from its 3D structure. In other words, is perceived glossiness invariant under changes in the 3D geometry of a surface and does the visual system possess a corresponding “gloss constancy” mechanism that takes local curvatures into account in order to estimate the glossiness of a surface?

This question has been addressed by Nishida and Shinya (1998). In a matching experiment, they used computer-generated grayscale stimuli that had complex 3D shapes, varying in spatial frequency and amplitude (similar to the...
stimuli shown in Figure 3). The Phong lighting model (Phong, 1975) was used to simulate the reflection properties of surfaces. The monocularly presented test and match stimuli were displayed side by side on the monitor and rotated in synchrony around their horizontal middle axis. The task of the observers was to match the perceived lightness and glossiness of the two surfaces by manipulating the diffuse component and the exponent of the Phong lighting model in one of the surfaces. Nishida and Shinya found that the reflection parameter values that were chosen for the match surface differed considerably from those of the test unless the test and the match surfaces exhibited similar local curvature. This indicates that the visual system is, in general, unable to provide perfect gloss constancy.

The aim of the present study was to investigate whether gloss constancy performance can be influenced by the availability of different sources of information in the proximal stimulus, and if so, how. In matching experiments similar to the one conducted by Nishida and Shinya (1998), we tested three different kinds of potential glossiness cues for their possible contribution to gloss constancy performance. These sources of information, which are usually present under ecological stimulus and viewing conditions, are (1) disparity information, (2) motion-induced information, and (3) color information.

The possible contribution of disparity information to gloss constancy

Under stereoscopic viewing, two types of disparity must be distinguished in glossy surfaces: Surface disparity (i.e., disparity between corresponding surface points) and highlight disparity.

Surface disparity is assumed to provide the visual system with relevant information for inferring the 3D structure of surfaces (“structure from stereo”). The spatial properties of highlights are influenced both by the reflection properties of the surface and by the local curvature. Assuming that surface disparity helps to determine the latter factor, one would expect that also the estimation of the former factor is facilitated. This has already been pointed out by Nishida and Shinya (1998).

The disparity of highlights generally differs from the disparity of corresponding surface points in the two monocular half-images (cf. Blake, 1985; Gräper, 1922; Kirschmann, 1895). In several studies, it has been found that the presence of highlight disparity in a stereoscopic stimulus leads to an enhancement of the realism (Blake & Bülthoff, 1990; Wendt et al., 2008) as well as the strength of perceived glossiness (Hurlbert, Cummings, & Parker, 1991; Wendt et al., 2008). The availability of highlight disparity may also contribute to a hypothetical gloss constancy mechanism: The light reflected from surface areas that contain highlights can be described as an additive mixture of diffuse and specular components.
In a static monocular grayscale image, or generally, in an image where both components have the same chromaticity, it is impossible for the visual system to reliably infer which portion of the incoming light is produced by the diffuse component and which is produced by the specular component. Due to the confounding of these two factors, it is also hard to determine the spatial extent of highlights, especially if they are highly blurred (i.e., when the specular portion merges smoothly into the diffuse underlay of the surface). However, due to the fact that the areas taken by the highlights will not cover exactly the same areas of the surface in the two monocular half-images, the availability of highlight disparity could help identify the boundaries of the highlights (Figure 2b): If the two eyes receive unequal amounts of light from the same surface point, then this indicates that this particular surface point reflects the incident light specularly to some degree (cf. Oppel, 1854; von Helmholtz, 1867). Some phenomenological observations suggest that the visual system is actually able to build separate representations of the perceptual components “essential surface color” and “gloss layer” (associated with the diffuse and specular components, respectively) when such binocular information is present, whereby the latter tends to appear somehow detached from the former (cf. Bixby, 1928; Hering, 1879; Wendt et al., 2008).

The possible contribution of motion information to gloss constancy

A number of theoretical considerations (Kirschmann, 1895; Preston, 1931; von Helmholtz, 1867) and some empirical observations (Kirschmann, 1892; Zocher & Reinicke, 1925) suggest that motion information has an effect on perceived glossiness. More recently, this hypothesis has been supported by a demonstration (Hartung & Kersten, 2002) and an experimental study (Sakano & Ando, 2008).

With respect to potential underlying mechanisms that may explain a contribution of motion information to gloss constancy, essentially the same holds as for the stereoscopic cues: First, motion information may contribute to a better estimation of the 3D shape of a surface (“structure from motion,” cf. Ullman, 1979), at least in combination with stereoscopic information (Johnston, Cumming, & Landy, 1994; Richards, 1985). Second, also a mechanism aiming to separate the two components “surface color” and “gloss layer” from each other could benefit from motion information. A moving surface will cause successive changes that are, in a sense, the temporal analogues of the spatial differences between the two half-images in the stereoscopic case: The moving (e.g., rotating) surface will successively change its orientation toward the position...
of the light source and/or the observer. Since the amount of reflected light sent to the eye depends on these geometrical factors, the eye will be confronted with different amounts of light from the same surface point, according to different orientations of the surface (Figure 2c). In an extreme case, a certain surface point may send a great amount of specularly reflected light to the eye in one orientation and only diffusely reflected light in another. Given that the visual system is able to relate these different successive intensity signals to each other, a segregation of the two underlying components would be facilitated.

**The possible contribution of color information to gloss constancy**

Highlights that appear on dielectric materials (e.g., plastics) approximately have the color of the illumination (however, for a critique on this simple rule, see Angelopoulou & Poger, 2003), which generally differs from the color of the diffuse component. This information seems to be taken into account by the visual system to provide color constancy (Lee, 1986; Maloney, 2002; Yang & Maloney, 2001).

Color information also seems to play a role in gloss perception. This is suggested by the finding that the visual system does not accept all imaginable combinations of diffuse and specular colors: If a surface exhibits physically improbable color combinations (e.g., red highlights on an otherwise white surface), then the perceived glossiness of the surface diminishes and the overall surface color tends to look heterogeneous, resulting into a “somewhat strange” appearance of the surface (Nishida et al., 2008, p. 339a).

Regarding a possible contribution to gloss constancy performance, such color information may also support a separation of the diffuse and specular reflection components that are additively combined in the stimulus. As already mentioned above, it is impossible for the visual system to make an appropriate estimate of the spatial extension of the highlights on a surface when both reflection components are characterized by the same chromaticity (at least when the surface is presented monocularly and static). The perceived colors of the surface points then vary only along one dimension (i.e., the intensity axis within a color space while the chromaticity stays constant) and there is no reliable cue available in the proximal stimulus that could indicate the boundary of the highlights. If, however, the diffuse and specular components exhibit different chromaticities, then not only intensity differences will occur at the boundary between the highlight and the surface color, but also chromatic transitions (Figure 2d): If one plots the color coordinates of an image showing a glossy surface in a color space, then two clusters result, which would be unambiguously relatable to the two reflection components (Tominaga & Tanaka, 2000). Thus, these transitions in color could provide the visual system with useful information for the estimation of spatial properties of highlights.

In the following sections, we report the results from three matching experiments. In the first experiment, we replicated the experiment of Nishida and Shinya (1998). In the second experiment, we investigated the influence of disparity information and motion information on gloss constancy performance. As in the first experiment, the observers matched the perceived lightness as well as the glossiness of the test and match surfaces by adjusting the diffuse component and the exponent of the Phong lighting model, respectively. In the third matching experiment, we investigated the contribution of color information to gloss constancy performance.

**General methods**

**Stimuli**

The computer-generated stimuli used in the present study were similar to those of Nishida and Shinya (1998). The virtual surfaces had a square base area in the (x, z)-plane and a complexly shaped height profile (y-coordinate), consisting of a combination of a number of sinus gratings with random orientations and phases (see Appendix A for details on stimulus construction). The shape of the stimuli was controlled by two parameters $f$ and $a$, which determined the spatial frequency and the amplitude of the height profile, respectively (see Figure 3). The projection of the square base of the surfaces on the monitor screen had a side length of 6.5 cm. The viewing distance was 50 cm.

The color of each pixel was calculated by means of the Phong lighting model (Phong, 1975) given as

$$
\begin{align*}
\begin{pmatrix}
I_R \\
I_G \\
I_B
\end{pmatrix} &= I_d k_a \begin{pmatrix}A_R \\
A_G \\
A_B \end{pmatrix} + I_p k_d \cos(\theta) \begin{pmatrix}D_R \\
D_G \\
D_B \end{pmatrix} \\
&+ I_p k_s \cos^n(\alpha) \begin{pmatrix}S_R \\
S_G \\
S_B \end{pmatrix}.
\end{align*}
$$

Thus, the resulting color vector $(I_R, I_G, I_B)^T$—where the indices refer to the relative intensities of the three monitor lights—was an additive mixture consisting of three different reflection components, namely the ambient, diffuse, and specular components. The relative amount to which each component contributes to the total intensity is determined by the parameters $k_a$, $k_d$, and $k_s$, respectively, and by the intensity of two different light sources: The intensity of the ambient light $(I_a)$ and the intensity of a point-light source $(I_p)$. The diffuse component is scaled by the cosine of the angle between the surface normal and the direction of the light source $(\theta)$, as defined by Lambert’s law. The specular component is scaled by the cosine to the power of $n$ of the angle between the cardinal direction of the reflected light and the viewing direction $(\alpha)$: The size of
the exponent \( n \) determines the spread of the specularly reflected light ("shininess"). Additionally, each component contains a triple of color weighting factors for the R, G, and B channels of the monitor that were used to assign different hues to the single reflection components.

In recent work (e.g., Fleming et al., 2003), the Phong model was often replaced by alternative, physically more plausible rendering models, for example, the isotropic Ward model (Ward, 1992). For the kind of stimuli used in the present experiments (i.e., complex surfaces that are illuminated by a single point light), the choice of the rendering model has only negligible effects on the resulting image. That is, for a given rendering using the Ward model with specific parameters a virtually indistinguishable image can be produced by using appropriate parameters in the Phong model. The two models can thus essentially be considered as two different parameterizations of the same stimulus set. Our choice to stick with the Phong model was mainly motivated by two considerations: First, the results can then be more easily compared with those of previous studies investigating the influence of highlights on gloss perception that also used this model (Nishida & Shinya, 1998; Wendt et al., 2008). Second, the Phong model has the advantage that the intensity and the blur of the highlights can be manipulated independently, whereas highlight blur and intensity are negatively correlated in the Ward model.

In the first two experiments (Experiments 1 and 2), only grayscale stimuli were used and the weighting factors were therefore set to 1.0 for all components. In the third experiment—where we investigated the contribution of color information to gloss constancy performance—we chose a yellowish hue for the ambient and diffuse components (with the weighting factors \( A_R = D_R = 1.0, A_G = D_G = 1.0, A_B = D_B = 0.0 \)) while the specular color was kept unchanged as white light. When we used such chromatic stimuli we paid heed to equate them in luminance to corresponding grayscale stimuli.

In all experiments, the observer’s task was to match the perceived lightness and glossiness of test and match surfaces as closely as possible by adjusting the values of the diffuse component \( (k_d) \) and the Phong exponent \( n \) in the latter (however, in our experiments we transformed \( n \) into the “Phong index” parameter \( m = n^{0.25} \), which provides an approximately equidistant scale of perceived shininess, cf. Wendt et al., 2008). During the stimulus presentation, these two parameters could be manipulated by the observer using the arrow keys on the keyboard. For the diffuse component \( (k_d) \), values within the interval [0.0, 0.7] in steps of 0.02 could be chosen, and for the Phong index \( m \), values within the interval [1.0, 3.5] in steps of 0.05 could be chosen. Figure 4 illustrates the perceptual effect of manipulating the Phong index. The test surface was presented with 9 different reflection parameter combinations (the cartesian product of \( k_d = \{0.2, 0.3, 0.4\} \) and \( m = \{1.35, 1.8, 2.35\} \)) for each shape condition that was tested. All other parameters in Equation 1 were kept constant with \( I_a = 0.2, k_a = 0.3, I_p = 1.0, k_s = 0.2 \) in Experiments 1 and 2 and \( k_s = 0.45 \) in Experiment 3.

The intensity of the highlights was fixed to reduce the number of variables that must be adjusted by the subjects in order to get more reliable settings. This seems justified because changing the shape of a surface influences only the width and blur of the highlights but not their intensity.

**Apparatus**

All stimuli were presented in front of a black background on a 22-in monitor (Sony Triniton Multiscan 500 PS), driven by an NVIDIA GeForce 7900 GTX graphic card and were rendered using OpenGL. The gamma value of the monitor was 2.53.

**Observers**

Four observers participated in all three experiments reported in this study. All had normal or corrected-to-normal visual acuity. With the exception of author GW, all observers were naive with respect to the purpose of the investigations. None of the observers, except for GW had any specific knowledge about the field of lightness perception or computer graphics.

**Experiment 1**

In Experiment 1, we essentially replicated the matching experiment conducted by Nishida and Shinya (1998). The aim of the experiment was to test to what extent observers
are able to match the reflection properties of two complex-shaped surfaces that differ in shape.

The test and match surfaces were simultaneously presented side by side on the screen with a center-to-center distance of 12 cm. The left stimulus was always the test surface, the right one always the match surface. During the presentation, both surfaces synchronously rotated back and forth around their horizontal middle axes within a 90° range and with a speed of approximately 56°/s. In the initial 0° position, the global surface normal of the stimulus was parallel to the y-axis of the virtual space, and in the 90° end position, the global surface normal was parallel to the z-axis (=observer direction; see Figure 5, left).

Like Nishida and Shinya, we investigated two different kinds of simulated motion: Surface motion and motion of the observer. The first kind of motion was put into practice by the use of a rotating surface and a fixed position of the point-light source (with the coordinates \(x = 0.0, y = 70.71, z = -70.71\)). In the observer motion condition, the surface rotated in exactly the same way, but in this case the light source followed the motion of the surface: In each frame, the point-light source was positioned in the same direction as the global surface normal, with a constant distance of 100 cm to the center of the surface.

The test surface always had fixed global shape parameters, namely a spatial frequency parameter of \(f = 3.0\) and an amplitude of \(a = 0.2\) (see surface no. 4 in Figure 3). The match surface was presented in one of 7 different shape conditions with the shape parameter combinations \((f, a) = (1.5, 0.2), (1.5, 0.4), (3.0, 0.1), (3.0, 0.2), (3.0, 0.4), (6.0, 0.1),\) and \((6.0, 0.2)\).

This resulted in 126 different conditions (7 shape conditions \(\times 3\) values for the diffuse component \(k_d\) \(\times 3\) values for the Phong index \(m\) \(\times 2\) rotation conditions) whereby each condition was presented 4 times. The total of 504 randomly ordered trials was subdivided into 4 blocks that were completed by each observer in 4 separate sessions.

Results

Figure 6 shows the results of Experiment 1, averaged across all four observers. The results for surface motion and observer motion are shown separately. The seven diagrams in each panel correspond to the seven different shape conditions that were used for the match surface. Each diagram contains the nine pairs of reflection parameter values \((k_d, m)\) respectively) of the test stimuli (open circles) and the corresponding pairs of the mean settings (arrowheads). The error ellipses are based on the covariance matrices of the samples.

Our procedure differed from that of Nishida and Shinya (1998) in that the global 3D shapes were varied in the match stimulus instead of the test stimulus, while the shape parameters of the test surface were fixed (to \(f = 3.0\) and \(a = 0.2\)). Thus, one would expect that the data points of the settings in our diagrams will correspond to the effects found by Nishida and Shinya, if the latter are point reflected at the coordinates of the test. This is roughly the case.

A comparison of corresponding diagrams in the top and bottom panels shows that there is virtually no difference between the two types of motion with respect to gloss constancy. The errors in the diffuse settings are, however, systematically larger in the observer motion condition than in the surface motion condition. This may be due to the fact that for each surface point the amount of diffuse light was constant during the rotation of the surface in the observer motion condition (because the light source was following the direction of the global surface normal during the rotation). This, in turn, may suggest that the visual system also takes dynamic luminance changes into account when judging the essential surface color of an object.

In four of the seven shape conditions (namely under shape conditions 1, 3, 5, and 7), the given test parameter values (open circles) and the settings (arrowheads) deviate
Surface motion
Vp: GW+TK+LF+CG
N = 16

Observer motion
Vp: GW+TK+LF+CG
N = 16
considerably. Although the directions and the sizes of the deviations may differ a little under the two motion conditions, there seems to be a common pattern in these errors: A surface that has strong local curvature appears both glossier and darker than a less bumpy surface albeit both have the same reflection characteristics. These results indicate that the visual system is, in general, unable to achieve gloss constancy under the stimulus and viewing conditions used in this experiment.

However, under the three remaining shape conditions (2, 4, and 6), the observers made nearly perfect matches. For shape condition 4, this is trivial, because in this case the match surface had the same global shape parameter values as the test, i.e., the diagram in the middle of each set in Figure 6 shows the data obtained with a symmetric match in this regard. However, also the settings under shape conditions 2 and 6 show no significant deviations from the reflection parameters of the test, and this finding does not agree with the results of Nishida and Shinya (however, it is not clear whether deviations from the results of Nishida and Shinya can in part be attributed to the use of a different monitor gamma in the present study). These are the two shape conditions that differ in both shape parameters $f$ and $a$ from the test values and not just in one. Actually, the shape conditions 2, 4, and 6 are related to each other in a specific way: With regard to local curvatures, each of these shape conditions can be transformed into the others by a simple scaling of the underlying vertices. Therefore, it seems that surfaces whose structures only differ in scaling exhibit similar shading features in the proximal stimulus (when they have equal reflection properties) that are considered as equivalent by the visual system. A possible explanation for this result may be that the shape of the highlights remains approximately invariant under a scaling transformation. It would thus be interesting to explore to what extent the invariance of highlight shape actually holds.

**Experiment 2**

In Experiment 2, we investigated how gloss constancy performance is influenced by the presence or absence of motion and disparity information. The availability of disparity information was varied in three levels: The stimuli were presented either monocularly without disparity, with surface disparity alone, or with both surface and highlight disparities. Motion information was varied in two levels, by using either static or rotating stimuli. The surface rotation used in Experiments 1 and 2 differed from that used in Experiment 1. This change was motivated by a finding from additional experiments not reported in this paper (Wendt, 2010) indicating that there is no difference in gloss constancy performance between a static surface and a surface that rotates around its horizontal middle axis (as shown in Figure 5, left). We assume that this is due to the fact that during rotations of this kind highlights are primarily visible within a relative short sequence of frames. Hence, in order to achieve a glossiness match, the observer has to wait until the orientation of the surface provides an opportunity to judge its glossiness.

For this reason, the surfaces rotated counterclockwise around their vertical middle axes (which were tilted 45° toward the observer direction before, see Figure 5, right) at a speed of ca. 90°/s in the dynamic condition. This ensured that the stimuli contained highlights during the entire revolution of the surfaces. In the static condition, the surfaces were tilted in the same way but presented without motion. In Experiments 2 and 3, the position of the point-light source was fixed at $(x = 0.0, y = 100.0, z = 0.0)$ in the virtual space.

In order to provide binocular cues, a stereoscopic arrangement of the stimuli was necessary. For each stimulus, two monocular half-images were generated, which were displayed side by side on the screen with a center-to-center distance of 12 cm. During each trial, the two pairs of the monocular half-images of the test and match surfaces were presented one below the other on the screen with a vertical center-to-center distance of 12 cm and fused by means of a mirror stereoscope (SA200 Screenscope Pro). In the “monocular condition,” both eyes received identical images of the stimuli. In the “surface disparity condition,” the stimuli were presented with perspective, i.e., the surfaces in the two half-images were rendered taking into account the distance between the surfaces and the observer’s eyes (ca. 50 cm) as well as the interpupillary distance (6 cm). In this condition, no highlight disparity was available, i.e., the highlights appeared on corresponding surface locations. This was realized by feeding identical observer vectors for both eyes into the Phong lighting model (Equation 1) using the mean of the two correctly oriented observer vectors (note that only the specular component in Equation 1 depends on the position of the observer). In the “surface and highlight disparity condition,” the surfaces were presented with perspective and the correct observer vectors for both eyes were used to render the two half-images of the stimuli.

Six different conditions, obtained by combining the three levels of the factor “binocular information” with the two levels of the factor “motion information,” were tested in Experiment 2. During each trial, the given condition
was applied to both the test and match surfaces. The test surface always had the shape condition 4 (Figure 3) while the match surface had the shape condition 1 or 3, respectively—which were the two conditions under which the largest and most systematic errors occurred in Experiment 1. As in Experiment 1, the test surface was presented with one of nine different reflection parameter combinations of the diffuse component $k_d$ and the Phong index $m$ (see General methods section) and also the task and procedure were the same as before. In total, 108 different condition combinations resulted (3 levels of the factor “binocular information” × 2 levels of the factor “motion information” × 2 shape conditions for the match surface × 3 values for the diffuse component of the test × 3 values for the Phong index), which were each repeated four times. The set of 432 trials was randomized and split in four equally sized blocks that were completed by each observer in four different sessions.

Results

Figure 7 shows the results of Experiment 2 pooled over all four observers. The left half shows the results for shape condition 1, the right half those for shape condition 3. The top row in each of the two sets of diagrams represents the data in the same way as in Figure 6: For each combination of the levels of the factors “binocular information” and “motion information,” the mean settings of the observers (arrowheads) are linked to the given reflection parameter values of the test (origin of the arrows) for each of the nine parameter pairs that were investigated. To facilitate the comparison of the results from the two conditions, the level combinations that belong to the statically presented stimuli are shown in the left diagram and those that belong to the dynamic condition are shown in the right-hand one. In both diagrams, the three levels of the factor “binocular information” are indicated by different colors.

The diagrams at the bottom of each panel show the same data in a more condensed form: The data for the two reflection parameters “diffuse component” and “Phong index” are shown separately. Each data point shows the mean difference (averaged across all nine reflection parameters and all observers) between the values of the respective parameter set by the observers and the preset value of the test surface. The colored lines connect the data points obtained under the two motion conditions.
“static” (red line) and “dynamic” (green line), respectively; the three levels of the factor “binocular information” are indicated by the abscissa position.

Our results show that both motion and disparity information have a significant effect on gloss constancy performance in the predicted direction. This holds for both shape conditions tested in this experiment (see the bottom left diagram titled “Phong index” in each of the two sets in Figure 7). The systematic errors in glossiness matches are significantly smaller under dynamic than under static presentation in all levels of the factor “binocular information.” A two-factorial ANOVA revealed a significant main effect of the factor “motion information” with \( p < 3 \times 10^{-9} \) for shape condition 1 and \( p < 6 \times 10^{-8} \) for shape condition 3. The main effect of the factor “binocular information” was also significant under both shape conditions (\( p < 0.00047 \) in shape condition 1 and \( p < 0.00013 \) in shape condition 3). Compared to monocularly presented stimuli, the systematic error of the glossiness matches decreased considerably when the stimuli contained surface disparity. This systematic error was reduced further when in addition to surface disparity also highlight disparity was available in the stimuli. This indicates that the visual system takes both kinds of binocular information into account when it estimates the glossiness of a surface.

The lightness matches (“diffuse component”) that generally show non-significant systematic errors under almost all level combinations of the two factors do not seem to be affected by the presence of binocular information or motion information (but see the results of Experiment 3 for the individual observers shown in Figure 9).

**Experiment 3**

In Experiment 3, we investigated the contribution of color information to gloss constancy performance and its possible interaction with the motion and disparity information already investigated in Experiment 2. To test the effect of “color information,” the gloss constancy performance achieved with achromatic surfaces was compared to the performance achieved with stimuli, in which the hue of the ambient and diffuse reflection components differed from that of the specular component. The ambient and diffuse components were yellow and the specular component achromatic (see General methods section). As in Experiments 1 and 2, the test surface was always presented with the global shape parameter values \( f = 3.0 \) and \( a = 0.2 \) (shape condition 4 in Figure 3; cf. Appendix A). In order to keep the total number of trials within a reasonable limit, we only used shape condition 1 (\( f = 1.5; \ a = 0.2 \)) for the match surface. This restriction seems justified in light of the great similarity of the results with shape conditions 1 and 3 observed in Experiment 2.

The task and the procedure were the same as in Experiment 2. In total, 108 different condition combinations resulted (3 levels of factor “binocular information” \( \times \) 2 levels of factor “motion information” \( \times \) 2 levels of factor “color information” \( \times \) 3 values for the diffuse component of the test \( \times \) 3 values for the Phong index), and each was repeated four times with four different 3D structures for the test and match surfaces. The resulting set of 432 trials was randomized and split into four equally sized blocks that were completed in four different sessions by each of the four observers.

**Results**

Figure 8 shows the results of Experiment 3 pooled across all observers. The diagrams in the first two rows give a detailed view on the results. Each of the nine reflection parameter pairs of the test surface (origins of the arrows) is linked to the corresponding pair of the mean settings (arrowheads). This is shown separately for the levels “achromatic” and “colored” of the factor “color information” (rows) and the levels “static” and “dynamic” of the factor “motion information” (columns). In each diagram, the data under the three levels of factor “disparity information” are indicated by different colors.

A comparison of the matching data under the “achromatic” (first row in Figure 8) and “colored” (second row) condition shows that the size of the error ellipses is greatly reduced in the latter case. This indicates that the availability of color information leads to a considerable improvement in the matching accuracy.

In order to determine the contribution of color information to gloss constancy performance, the data are again presented in a more condensed form. The two diagrams in the bottom row in Figure 8 display the results for the parameters “Phong index” (left diagram) and “diffuse component” (right diagram), separately. Each data point in the diagrams represents the mean deviation between the settings and the value of the corresponding parameter preset for the test surface. The data were pooled across all nine reflection parameter pairs and all observers. The data points for “static presentation” and “dynamic presentation” are connected with red lines and green lines, respectively. The points that belong to the “achromatic” and “colored” condition are connected with dashed and continuous lines, respectively. The levels of the remaining factor “binocular information” are indicated by the abscissa position in the diagrams. Thus, the data points that are connected with dashed lines in Figure 8 were obtained under exactly the same condition as those in Experiment 2 with shape condition 1 for the match surface. As is to be expected, the results under the achromatic conditions in Experiment 3 are very similar to the corresponding results in Experiment 2: The glossiness
Figure 8. Results of Experiment 3, averaged across all observers. The two upper rows show a detailed view of the results (compare Figure 7), subdivided into the different levels of the factor “color information” (rows) and the factor “motion information” (columns). The belongingness of the settings to the three different levels of the factor “binocular information” is indicated by the use of different colors. The two diagrams in the bottom show the same results in a more condensed way (cf. Figure 7). The error bars represent ±2 SEM.
Figure 9. Individual results of the four observers in Experiment 3 (cf. Figure 8). The error bars represent ±2 SEM.
matches benefit significantly from the availability of both motion information and disparity information, while the lightness matches do not seem to be affected by the presence of these information sources.

The main concern of the present experiment was to investigate the role played by color information in gloss constancy performance. Our results show that the availability of color information leads to a significant reduction of the systematic errors in glossiness matches, compared to those stimulus conditions where only grayscale surfaces were used (see the bottom left diagram in Figure 8). This was tested with a three-factorial ANOVA that revealed a significant main effect of the factor “color information” for both the glossiness matches ($p < 0.0056$) and the lightness matches ($p < 4.42 \times 10^{-8}$). With respect to the lightness matches, this trend seems to be reversed, i.e., the systematic errors seem to increase under the color conditions compared to the achromatic conditions. On closer inspection, however, this finding is the result of an inappropriate averaging of the data across all observers that ignored the fact that different observers show different polarities in their lightness settings. An inspection of the individual results of the four observers shown in Figure 9 reveals that the lightness matches benefit most from the availability of color information. A further analysis of the individual data also reveals an effect of the factor “motion information” on the lightness matches: All observers show smaller systematic errors under static than under dynamic stimulus presentation. With regard to the glossiness matches, the diagrams for the single observers show a more heterogeneous pattern of results (left column in Figure 9): Observers LF and CG, for instance, seem to take color information more into account than observers GW and TK and while observer GW seems to make strong use of motion information in order to judge the glossiness of a surface, the other observers seem to use this kind of information to a much lesser degree—indicating that different observers show a different receptiveness to certain combinations of information.

### Discussion

The aim of the present study was to investigate to what extent the visual system makes use of different sources of information in order to approach the goal of gloss constancy, that is, the ability to reliably estimate the reflective behavior of a surface independently from its 3D structure.

In this context, we focused on the role of spatial properties of highlights as a cue for glossiness. As Hunter (1975) already pointed out, one has to distinguish between several gloss categories, and different visual criteria may be used to evaluate glossiness in each category. Due to the illumination conditions and the shapes used, our stimuli showed pronounced highlights but a rather low amount of surface gloss—compared to glossy surfaces that exhibit mirror images of the surrounding (cf. Fleming et al., 2003). Hence, our findings with respect to the ability of the visual system to perform gloss constancy relate only to surfaces that feature more or less pronounced highlights.

The results of a prior study conducted by Nishida and Shinya (1998) with stimuli similar to those used in the present study indicated that gloss constancy is rather weak, at least when the surfaces are presented monocularly, monochromatic and with a certain kind of object rotation. These results were generally confirmed in our Experiment 1.

In Experiments 2 and 3, we tested the contribution of three different cues to gloss constancy performance from which it is already known that they play a role in the perception of glossiness: (1) Disparity information, (2) motion information, and (3) color information. The results of our experiments indicate that the visual system makes use of all three sources of information—both separately and in combination. Especially, the glossiness settings in our matching tasks seem to benefit from the presence of these cues. However, with respect to the glossiness matches our results also suggest that observers take these cues into account in different ways: A cue that has a strong effect for one observer may be disregarded by another. The effect of the different cues on lightness matches was more consistent across observers. The availability of color information considerably improved lightness constancy performance; the availability of motion-induced information, however, seems to diminish it. The presence of disparity information had no effect at all on the lightness matches. Our finding that the lightness constancy performance was better under static than under dynamic stimulus presentation is rather astonishing given our assumption that the presence of motion information facilitates a separation between specular and diffuse reflection components. As both components stand in a complementary relation to each other, this assumption implies that the estimation accuracy of both components benefits equally from the presence of motion information (a few studies, however, indicate that the presence of highlights or mirror images barely affects the perceived color of a surface; cf. Todd, Norman, & Mingolla, 2004; Xiao & Brainard, 2006). Furthermore, this finding seems to contradict our results from Experiment 1 where we found that the accuracy of the lightness settings was better under the surface motion condition (where dynamic changes of the diffuse light occurred) than under the observer motion condition (where the amount of diffuse light was constant for each surface point during the rotation).

Although the present investigation provides evidence indicating that the visual system makes use of several cues in order to perform gloss constancy, it remains an open question in which way these sources of information...
are used. An influence of two potential mechanisms was discussed above (see Introduction section): First, a cue (for instance motion and surface disparity information) may contribute to a better estimation of the 3D shape of a surface. This information could then be taken into account by the visual system to infer spatial properties of highlights, which depend on local surface curvature. Second, a cue (for instance, motion and color information, and highlight disparity) may contribute to a better segregation of the diffuse and specular reflection components, which are confounded at each point of a surface.

The experiments reported in the present paper give no clear indication about the underlying processes that led to an improved gloss constancy performance. On the one hand, it is still unclear whether motion information and surface disparity are actually used by the visual system to obtain a “veridical” estimate of the 3D shape of a surface (Todd & Norman, 2003). Richards (1985) has formally shown that at least the combination of these two types of information would suffice to reconstruct the metric structure of a surface, but it seems unlikely that the visual system actually produces such a thing as a veridical metric representation of the 3D shape of a surface; as Todd (2004, p. 120) put it: “[…] the perceptual representation of 3D shape involves a relatively abstract data structure based on qualitative surface properties […]” (see also Mausfeld, 2002). On the other hand, there is some evidence that high disparity could also be used by the visual system to estimate 3D shape (Blake, 1985; Blake & Bülthoff, 1990; Norman, Todd, & Orban, 2004; Todd, Norman, Koenderink, & Kappers, 1997). Likewise, even color information could influence the perceived shape of a glossy object: If the diffuse and specular components are of different hue, then the intensity peaks can be more easily classified as highlights—otherwise, they could be misinterpreted as highly curved locations on a rather matte surface (Todd et al., 1997; cf. Nefs, 2008). Beyond that, the visual system may employ mechanisms that differ completely from those suggested above.

Recent approaches in material perception rely on the assumption that the visual system uses certain image statistics as a cue for the reflection properties of a surface. For instance, Nishida and Shinya (1998) assumed that the reflection properties of two surfaces will be perceived as equal when their corresponding intensity histograms will be equal. Motoyoshi, Nishida, Sharan, and Adelson (2007) assume instead that the skewness of these intensity histograms provides the relevant information for the visual system. Anderson and Kim (2009) and Kim and Anderson (2010) point to some fundamental problems with this approach and there is at least one aspect in our data, namely, the influence of color information on gloss constancy that cannot be explained by these approaches because they have in common that they postulate—at least in their current incarnation—purely luminance-based mechanisms.

Appendix A

Construction of the stimuli

The following equation was used to generate a surface over a 100 × 100 square grid, indexed by integer values x and z:

\[ y = \sum_{k=0}^{40^2-1} \exp\left[-k^2/(40f)^2\right] \sin\left[(x\cos\theta_k + z\sin\theta_k) \frac{\pi k}{10^3} + p_k\right]. \] (A1)

Thus, \((x, y, z)\) are the vertices of this surface. The surface is the sum of sinus gratings with frequencies in the range from 0 to 2f cycles/sl, where sl is the side length of the square base. At each frequency, the orientation \(\theta_k\) and the phase \(p_k\) of the corresponding sinus grating was randomly chosen from the intervals \([0, \pi]\) and \([0, 2\pi]\), respectively. The exponential weighting function gives sinus gratings with high frequencies less weight. The height of the surface was normalized in such a way that an amplitude factor \(a = 1.0\) corresponded to a stimulus with a height range two times the side length of the square base.

The frequency parameter \(f\) and the amplitude \(a\) jointly determine the global shape of the surfaces and were varied in the experiments (see Figure 3).

The viewing direction was parallel to the z-axis. To display the surface, all vertices of the surface were scaled by the factor 0.12. This way, the stimuli (which were centered 10 cm behind the clipping plane in the virtual space) were brought to a size where the projection of the square base of the surfaces on the monitor screen had a side length of ca. 6.5 cm.

Four adjacent vertices in the mesh formed a quadrilateral facet of the complex-shaped surface. Based on these four vertices, the surface normals were calculated for each pixel within the respective facet, using a pixel shading interpolation method (Akenine-Möller & Haines, 2002). All brightness values were computed based on palette values.

Acknowledgments

This research was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) to Rainer Mausfeld (MA 1025/10-3). We also thank two anonymous reviewers for helpful comments.

Commercial relationships: none.

Corresponding author: Dr. Franz Faul.
Email: ffaul@psychologie.uni-kiel.de.
Address: Olshausenstr. 62, Kiel 24118, Germany.
References


