# Perceived glossiness and lightness under real-world illumination

**Maria Olkkonen**

Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

**David H. Brainard**

Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

Color, lightness, and glossiness are perceptual attributes associated with object reflectance. For these perceptual representations to be useful, they must correlate with physical reflectance properties of objects and not be overly affected by changes in illumination or viewing context. We employed a matching paradigm to investigate the perception of lightness and glossiness under geometric changes in illumination. Stimuli were computer simulations of spheres presented on a high-dynamic-range display. Observers adjusted the diffuse and specular reflectance components of a test sphere so that its appearance matched that of a reference sphere simulated under a different light field. Diffuse component matches were close to veridical across geometric changes in light field. In contrast, specular component matches were affected by geometric changes in light field. We tested several independence principles and found (i) that the effect of changing light field geometry on the diffuse component matches was independent of the reference sphere specular component; (ii) that the effect of changing light field geometry on the specular component matches was independent of the reference sphere diffuse component; and (iii) that diffuse and specular components of the match depended only slightly on the roughness of the specular component. Finally, we found that equating simple statistics (i.e., standard deviation, skewness, and kurtosis) computed from the luminance histograms of the spheres did not predict the matches: these statistics differed substantially between spheres that matched in appearance across geometric changes in the light field.

Keywords: 3D surface and shape perception, lightness/brightness perception, perception of surface gloss, high-dynamic-range display, real-world illumination, environment map

Citation: Olkkonen, M., & Brainard, D. H. (2010). Perceived glossiness and lightness under real-world illumination. *Journal of Vision, 10*(9):5, 1–19, http://www.journalofvision.org/content/10/9/5, doi:10.1167/10.9.5.

## Introduction

Color, lightness, and glossiness are perceptual attributes associated with object reflectance. For these perceptual representations to be useful, they must correlate with physical reflectance properties of objects and not be overly affected by changes in illumination or the viewing context provided by surrounding objects. Because the light reflected from objects to the eyes confounds reflectance and illumination, generating stable perceptual correlates of object reflectance is not trivial. Despite this difficulty, however, judgments of object color and lightness are known to exhibit considerable constancy across changes of illumination, at least for flat objects with matte reflectance (e.g., Brainard, 2004; Shevell & Kingdom, 2008; Smithson, 2005). Less is known about the constancy of other perceptual correlates of surface reflectance, such as glossiness, although this issue has recently begun to receive experimental attention (e.g., Doerschner, Boyaci, & Maloney, 2010; Fleming, Dror, & Adelson, 2003; Nishida & Shinya, 1998).

Object surface reflectance is characterized by the bidirectional reflectance distribution function (BRDF), which defines the amount and direction of light reflected from a surface as a function of the angles of the incoming and reflected light, relative to the surface normal. The BRDFs of natural surfaces can be quite complex (Alldrin, Zickler, & Kreigman, 2008; Dana, van Ginneken, Nayar, & Koenderink, 1999; Oren & Nayar, 1994). Nonetheless, simple parametric models of the BRDF are widely used in computer graphics and capture salient features of the variation in object surface reflectance (e.g., Yu, Debevec, Malik, & Hawkins, 1999). In the work reported here, we employed the isotropic Ward (1992) reflectance model. This model characterizes reflectance as the sum of a diffuse component and a specular component. Intuitively, the diffuse component describes how much light the object reflects in a nondirectional fashion while the specular component describes the strength and spread of mirror-like reflection. Objects with purely diffuse reflection typically appear as matte, whereas objects with a strong specular reflection component are likely to be perceived as glossy.

The majority of studies on the perception of object color and lightness have considered the appearance of matte flat objects. For three-dimensional objects with both diffuse and specular reflectance components, there is considerably

doi: 10.1167/10.9.5 Received April 6, 2010; published September 2, 2010 ISSN 1534-7362 

more richness in the relation between object reflectance, illumination, and the light reflected to the eye. In particular, because the specular component is directional, the geometric structure of the light field impinging on the object can have a large effect on the reflected light. In addition to overall variation in the intensity and spectrum of the illumination, the geometric structure varies from scene to scene (Debevec, 1998). This fact in turn raises the question of how well the visual system stabilizes object appearance across variation in the geometry of the light field and how this stabilization depends on object BRDF. These are the broad questions we address in this paper.

Two previous studies investigated the perception of surface glossiness under changes in illumination geometry (Doerschner et al., 2010; Fleming et al., 2003). In a seminal paper, Fleming et al. (2003) used a matching task to show that observers' perception of object specular reflectance components showed considerable stability across geometric changes in light field, as long as the light fields contained rich geometric structure. In their experiments, observers manipulated both the strength and spread (roughness) of the specular component, and they reported that these two aspects of the matches were approximately independent of each other. Doerschner et al. (2010) used a forced-choice procedure to study how the specular component of matches at the point of subjective equality (PSE) varied across changes in the geometric structure of the light field. Although their data are qualitatively similar to those of Fleming et al., they emphasized the deviations from veridical matching as opposed to the relative stability of the matches. In addition, they examined whether the PSEs were consistent across changes between three light fields and found that they were. Our work considers the same basic questions as these previous studies but extends the measurements in several ways.

First, we focus explicitly on identifying and testing principles that allow measurements of a small number of stimulus configurations to predict what will happen for a large number of combinations. In particular, we allowed observers to adjust both the diffuse component of reflectance and the strength of the specular component and asked whether changing the light field acts independently on these two components of the matches. We made this comparison for two levels of specular component roughness. We view the identification and testing of such principles as a high priority as we seek to understand object surface perception in the case of three-dimensional objects with realistic BRDFs viewed under geometrically complex lighting; without such principles it will be difficult to manage the fact that the number of experimental conditions increases geometrically with the number of stimulus parameters.

Second, our data are collected using a custom high-dynamic-range display, so that we did not need to truncate (tone map) the luminance of the most intense regions of the specular highlights.

# Methods

The purpose of this experiment was to measure how changing the geometrical structure of the light field affects the perceived glossiness and lightness of a three-dimensional object, across variations in both diffuse and specular components of the object's reflectance. To this end, we employed an asymmetric matching procedure, in which the observer adjusted the reflectance of a test object seen under one light field so that its glossiness and lightness matched that of a reference object seen under a second light field. In our experiments, observers could adjust both the diffuse reflectance and the strength of the specular reflectance independently during each match. The roughness parameter of the specular component was held constant during each match. As a control condition, we also measured matches when the objects were seen under the same light field. As part of the experiment, we explored the effect of the complexity of the surrounding scene on the matches.

## Observers

Three naive observers participated in the experiment. Observers VIL and BNW ran all conditions, while observer MVI ran two out of three conditions. Visual acuity and stereo vision were assessed with the Keystone VS-II vision screener. Color vision was assessed with the Ishihara color plates. Uncorrected (VIL, MVI) or corrected (BNW) visual acuity was at least 20/20 for all observers. All observers had normal stereo acuity and normal color vision.

## Stimuli

The test and reference objects were grayscale spheres; examples are shown in Figure 1. We used the isotropic Ward model to render spheres with different surface reflectance properties (Ward, 1992). In the model, surface reflectance is defined by three parameters: $\rho_d$ controls the diffuse component ("albedo"), $\rho_s$ controls the strength of the specular component ("glossiness"), and $\alpha$ controls the spread of the specular component ("roughness"). All scenes were rendered with the RenderToolbox package for Matlab (http://rendertoolbox.org). RenderToolbox acts as an interface to the RADIANCE rendering software (Ward, 1994), allowing scenes to be rendered independently at different monochromatic wavelengths. This ensures that the spectral interaction between illuminants and surfaces is simulated in a physically correct manner. This feature was not critical here because the stimuli were grayscale, but it remained convenient to use this software.

To provide spectral input to the RenderToolbox routines, the objects were assumed to have reflectance that

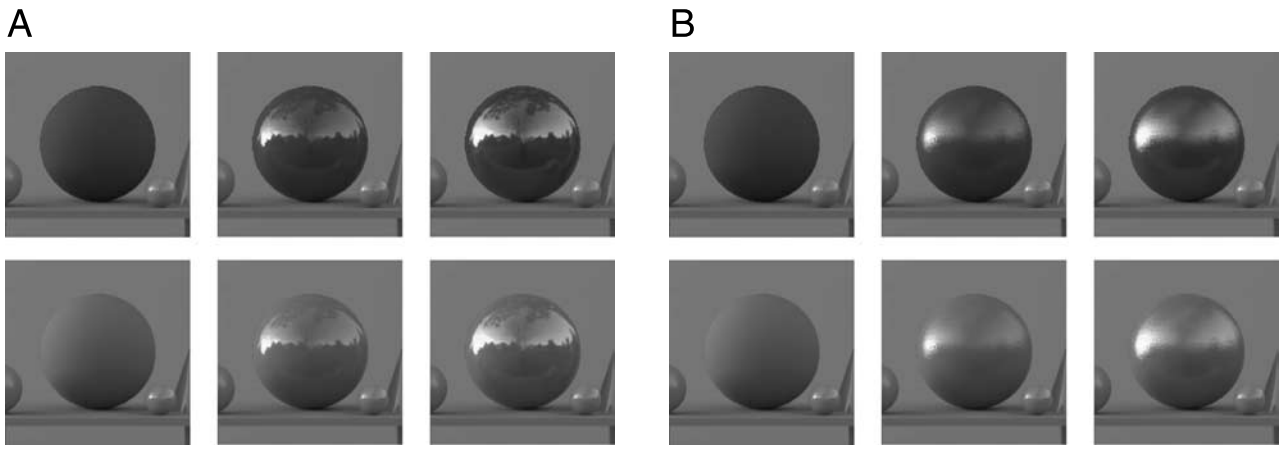


Figure 1. Sample stimuli rendered under the Campus light field. (A) Stimuli with the roughness parameter $\alpha$ set to 0.001. The two levels of diffuse reflectance are shown in rows, and the three specularities in columns. (B) Stimuli with the same parameters for diffuse and specular reflectances as in (A), but with $\alpha$ set to 0.1. The images shown here were created as described in Methods section, after which they were scaled, tone-mapped, and gamma-corrected so that they looked reasonable when viewed on a conventional display or in print. Images shown in subsequent figures were created similarly.

did not vary with wavelength. We used two levels of diffuse reflectance ($\rho_d$ = [0.15 0.35]) and three levels of specularity ($\rho_s$ = [0 0.06 0.12]) for the reference stimuli. One set of spheres was rendered with a smooth specular component ($\alpha = 0.001$), while another was rendered with a rough specular component ($\alpha = 0.1$).

Stimuli were rendered under four real-world light fields measured by Debevec (1998). These are illustrated in Figure 2, which also shows log luminance histograms for the image pixels corresponding to the sphere. We chose light fields measured both indoors and outdoors. The original light field measurements were reported in color, with separate values for nominal red, green, and blue (RGB) planes. To render the scenes spectrally within the RenderToolbox environment, the RGB light field images had to be converted to images that represented image intensity as a function of wavelength. This was done as follows. First, we converted the RGB values of the images at each pixel to *XYZ* values on the assumption that the RGB values represented linearized RGB primary intensities with respect to the sRGB standard (International Electrotechnical Commission, 1999). A three-dimensional linear model for surface reflectance, computed in our laboratory from measurements of 462 Munsell papers (Newhall, Nickerson, & Judd, 1943; Nickerson, 1957; Nickerson & Wilson, 1950), was then used to convert the *XYZ* values of each pixel to spectra. This was done by forming a 3 by 3 transformation matrix between *XYZ* values and linear model weights, using standard linear model methods (e.g., Brainard, 1995). Spectra were produced by multiplying the basis functions by the obtained weights. The particular choice of linear model was not of deep theoretical significance and was motivated primarily by convenience; we currently know little about the spectral variation of real-world light fields. The rendering procedure resulted in an hyperspectral image with 31 planes, with each plane corresponding to one wavelength band. Sample wavelengths were between 400 and 700 nm at 10-nm steps. For the current experiment, only the 500-nm band resulting from this process was used. This single image plane was replicated three times to produce identical red, green, and blue image planes for display. Although the conversion to spectral light fields and subsequent hyperspectral rendering was not necessary for the current experiment, we implemented it in preparation for planned future experiments where the spectral properties of the stimuli will be manipulated.

For the observer to perform the matching task, we needed to rapidly render scenes containing spheres of different reflectances. Because re-rendering the entire scene using Radiance was too slow, we generated scenes containing spheres of different reflectances by adjusting and combining three prerendered basis images (Griffin, 1999; Xiao & Brainard, 2008). Each basis image contained a sphere, and across the three basis images the sphere varied in reflectance. One basis image contained a matte sphere, one a smooth glossy sphere and one a rough glossy sphere. By taking the difference between the glossy and matte basis images, we could extract a difference image that represented the effect of adding a specular component to the sphere's reflectance. To generate stimuli with different levels of specularity, we then combined the matte sphere basis image with either the smooth or the rough difference image and varied the weight on the difference image to simulate the effect of varying the strength of the specular component. To simulate different levels of diffuse reflectance, we extracted the image pixels corresponding to the sphere in the matte basis image, scaled these, and reinserted them. This was done prior to adding in the specular component. We verified that images rendered in this manner provided a good pixel-by-pixel approximation to directly rendered scenes. For

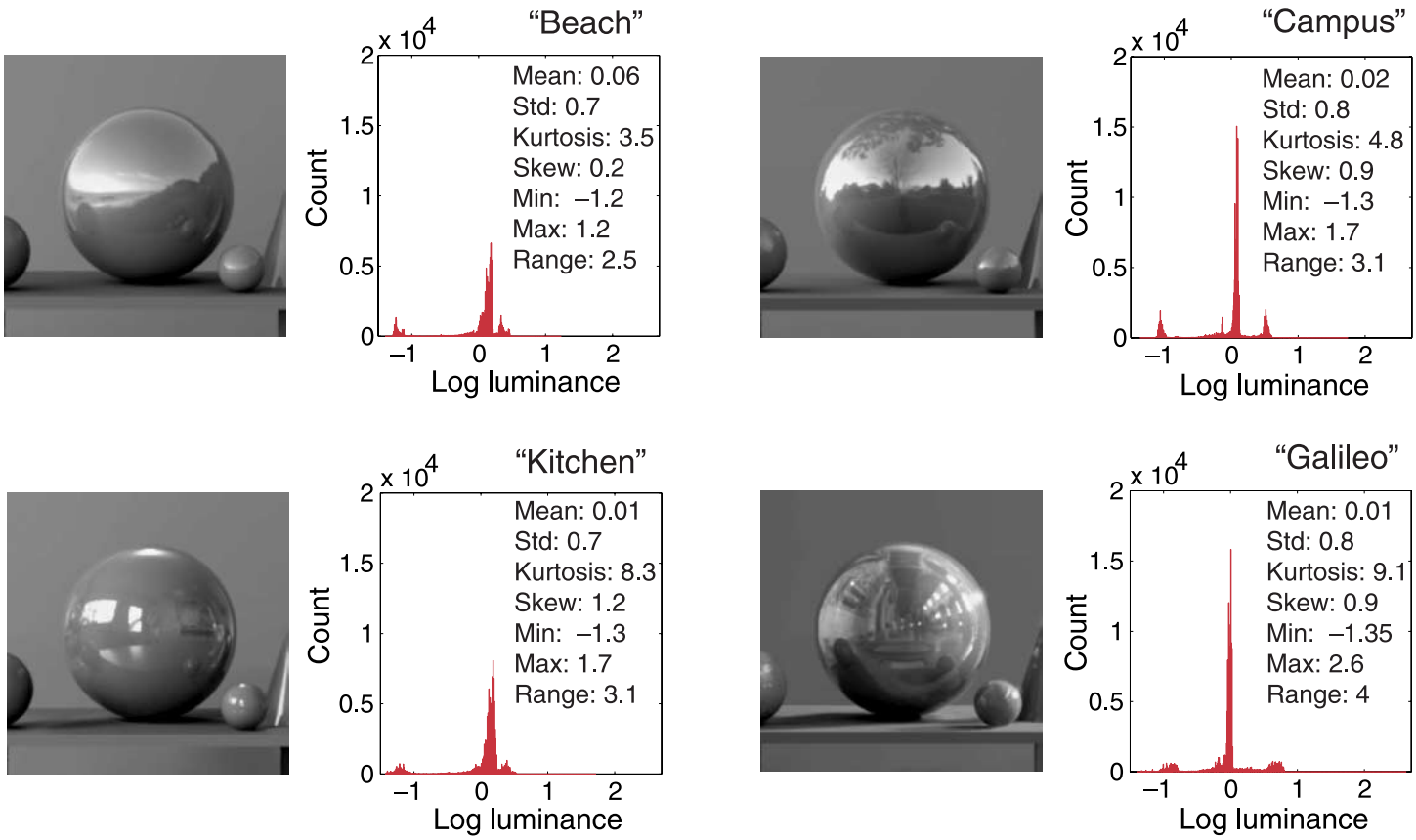


Figure 2. The left-hand images in each panel show a cutout of a scene rendered under the four different light fields used in this study. The spheres have Ward parameters $\rho_d = 0.5$, $\rho_s = 0.2$, and $\alpha = 0.001$. The histograms on the right from the images show the distribution of log luminances across the sphere. The insets list basic log luminance statistics of the spheres.

the three light fields and two sphere specularities we checked, the mean pixelwise difference between generated and directly rendered images of the spheres was at most 6% of the mean pixel value in the rendered image.

Two scene contexts were used. The spheres could either be embedded in the scene in which they were rendered (complex context), or against a simple checkerboard background (simple context). For each light field, the checkerboard background was created so that it had the same mean luminance and contrast as the corresponding complex background. A key difference between the two contexts is that the complex context was rendered using the same light field as its sphere and thus carried information about the geometry of the light field, while the simple context only varied with light field in terms of its mean luminance and contrast, and thus did not provide information about lighting geometry. Figure 3 shows a stimulus pair embedded in simple contexts in the upper row and the same pair embedded in complex contexts in the lower row.

The light fields were scaled so that the mean luminance reflected from the matte sphere basis image was the same for each. A common scale factor was then applied to all of the rendered images so that as a set they were mapped into the luminance range of the display.

## Apparatus

The stimuli were displayed on a custom high-dynamic-range (HDR) display, designed along the lines reported by Seetzen et al. (2004). The HDR display consisted of an LCD screen (ViewSonic 19″) where the commercial backlight was replaced by a projector (Panasonic DLP PT-D7600U) that illuminated the LCD screen with a projected image. The largest difference between our HDR display and a conventional transmissive LCD display is thus that the light pattern of the projector can be modulated in concert with spatial modulation provided by the LCD panel. This makes the LCD screen into a spatio-chromatic filter for the projector image and provides an overall dynamic range that is the product of that provided by the projector and the LCD in isolation. Both display devices were driven at a pixel resolution of 1280 by 1024 and at a refresh rate of 60 Hz through a dual-port video card (NVIDIA GeForce GT 120). The host computer was an Apple Macintosh G5.

The displays were arranged so that the LCD panel was enclosed in a box that prevented stray light within the experimental room from reaching the front of the panel and reflecting back to the observer. Visible surfaces within this box were lined with light-absorbing black cloth. The

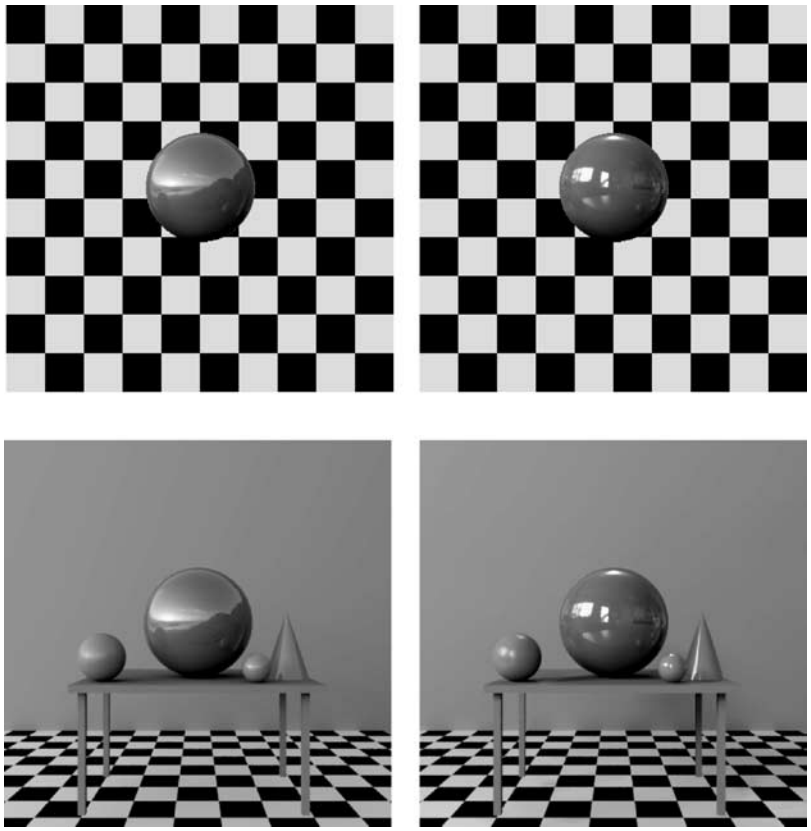

Figure 3. Experimental scenes. Upper row shows an example of a stimulus pair rendered under the (left) Beach and (right) Kitchen light fields and presented against the checkerboard background. Lower row shows the same sphere pair embedded in the corresponding complex scene.

observer viewed the LCD panel monocularly from a distance of 73 cm through a circular aperture (6.1 cm in diameter) at the end of the enclosing box. The observer's head was stabilized with a chin rest, which could be adjusted so that the eye was centered in the circular aperture. Interposed between the observer and the LCD panel was a black reduction screen (44 cm from the viewing aperture, square aperture of dimensions 16 × 16 cm; 20.6 × 20.6 degrees of visual angle), which prevented stray light from the projection system from reaching the observer's eye.

To display calibrated high-resolution images on the HDR display, it is necessary both to align the projector image with the LCD panel and to map desired stimulus values to appropriate RGB inputs for the video card. These tasks were accomplished using custom software, following the general methods outlined by Seetzen et al. (2004). To control the chromaticity and luminance of the overall display system, we used a spectroradiometer (PhotoResearch, PR-650) to characterize the properties of the projector and the LCD panel separately. This was done in situ, with the radiometer placed at the observer's eye position. First, we characterized the projector, which we used as a grayscale device so that its RGB input values were always set with R = B = G. We set the RGB input values of the LCD panel to their maximum level (corresponding to maximum transmission through the panel) and measured the relation between the R = G = B input values to the projector and luminance output for a series of input values. We then splined these to produce a full gamma curve for the projector. Second, we set all projector pixels to their maximum input values (full light output) and measured separately the gamma curves of the R, G, and B channels of the LCD panel, as well as the transmitted spectrum for each channel. For any desired display luminance and chromaticity, the characterization data were used to compute an R = G = B value for the projector and R, G, and B values for the LCD panel that produced the desired output.

The maximum luminance of the display was 423.5 cd/m$^2$. The minimum luminance was below the measurement range of our radiometer, but at least a factor of 10,000 below the maximum luminance. There were some deviations between desired and displayed chromaticities, due primarily to shifts in the chromaticity of the nominally neutral projected light as a function of luminance. These shifts were not corrected for by the display control software but were not readily apparent in the displayed images. Through analysis of the calibration data, we estimate that mean chromaticity of the scenes (i.e., the

neutral point of the display) was $x = 0.306$ and $y = 0.353$ across all input RGB values and changed gradually as a function of luminance from $x = 0.307$ and $y = 0.338$ to $x = 0.302$ and $y = 0.375$ as luminance varied between 0.025 and 423 cd/m$^2$, with additional shifts at very low luminances.

### Procedure

On each trial of the experiment, the observer viewed a pair of side by side images, one containing the reference sphere and one containing the test sphere (see Figure 2). Each image subtended 9.8 × 9.8 degrees on the display, with the test and reference spheres subtending 1.85 degrees. The two images were separated horizontally by 0.4 degrees; the distance between the centers of the reference and test spheres was 10 degrees. The observer adjusted both the diffuse and specular parameters of the test sphere so that it appeared to be made out of the same material (i.e., have the same reflectance properties) as the reference sphere. The exact instructions provided to the observers are provided in the supplemental material available at http://color.psych.upenn.edu/supplements/glossiness_lightness. The reference image was always presented on the left side of the display, and the test image on the right side. Adjustments were made by pushing one of four buttons on a button box. One pair of buttons increased/decreased the specular component of the test sphere, while the other two increased/decreased the diffuse component. Observers could cycle between four different step sizes by pressing a fifth button on the controller. Observers were required to adjust the test at least once at one of the two smallest step sizes to be able to move to the next trial. When the observer was satisfied with the match, he or she pushed a final button to accept it and move to the next trial.

Observers set matches in blocks of 20 trials each. A block was defined by a choice of two light fields and 10 reference stimuli. In symmetric matching blocks, observers set symmetric matches for each of the symmetric reference stimuli/light field pairs. In asymmetric matching blocks, observers set asymmetric matches for each reference stimulus, with each of the two light fields serving as the reference. Different background conditions (simple/complex) and adjustment tasks (symmetric/asymmetric) were run in separate blocks. Within each block, the 20 possible trial types were presented in random order. Observers were allowed to rest between trials and between blocks. Each block type was repeated three times over the course of the experiment, with 2–4 (MVI) or 4–7 (VIL, BWN) blocks per session. Across the experiments, blocks were ordered so that all blocks for one pair of light fields were completed before the observer moved on to the blocks for the next light field pair. The order of light field pairs was the same for all observers: Kitchen/Beach, Kitchen/Campus, and Kitchen/Galileo.

## Results

Observers' symmetric specular matches were close to veridical. Such matches from two observers are shown in Figure 4A for the Kitchen/Kitchen and Galileo/Galileo comparisons. Data for both simple (top row of panel) and complex (bottom row of panel) contexts are shown. In all cases, the data fall close to the unity line; the slopes of fitted regression lines ranged between 0.97 and 0.99.

Asymmetric matches were not always veridical. That is, changing the geometric structure of the light field had an effect on the appearance of the spheres. Figure 4B shows asymmetric specular matches for the same observers for the Kitchen/Galileo comparison. Matches for tests in the Galileo context against references in the Kitchen context are plotted with red symbols. Matches for tests in the Kitchen context against references in the Galileo context are plotted with gray symbols. The latter data have been flipped such that the match values (Kitchen) are shown on the $x$-axis and the reference (Galileo) values on the $y$-axis. This allows direct comparison between the two cases. The broad effect in the data is that the points fall well below the unity line. The slopes of the regression lines shown in Figure 4B for the simple context were 0.51 and 0.40, and for the complex context, 0.55 and 0.45 for VIL and BNW, respectively.

In an experiment similar to ours, Doerschner et al. (2010) reported differences in asymmetric specular matching depending on which of two light fields served as the reference. This effect would show up in Figure 4B as a divergence of the red and gray plotted points and is not readily apparent in the data shown. We have, however, observed such effects in preliminary experiments that employed a larger parameter range than those reported here.

Symmetric matches for the diffuse component were also close to veridical. In addition, the change in the geometric structure of the light field did not have a large effect on the diffuse component of the asymmetric matches. Figures 5A and 5B show the symmetric and asymmetric diffuse component matches, respectively, for the same observers and conditions as Figure 4 for the specular component matches. For the symmetric matches, the slopes of the regression lines fit to the data in Figure 5A ranged between 0.97 and 0.99. The slopes for the asymmetric matches (Figure 5B) were also close to 1 (range 0.92–1.05). The fact that the diffuse component asymmetric matches are close to veridical may arise because we scaled all of the light fields to reflect the same luminance from a purely diffuse sphere. We did this because our interest here is in the effect of changing the geometric structure of the light field, not in overall effects of changing light field luminance.

The type of data shown in Figures 4 and 5 may be summarized by the slope of the fitted regression line. Figure 6 provides the slopes of the fitted regression lines

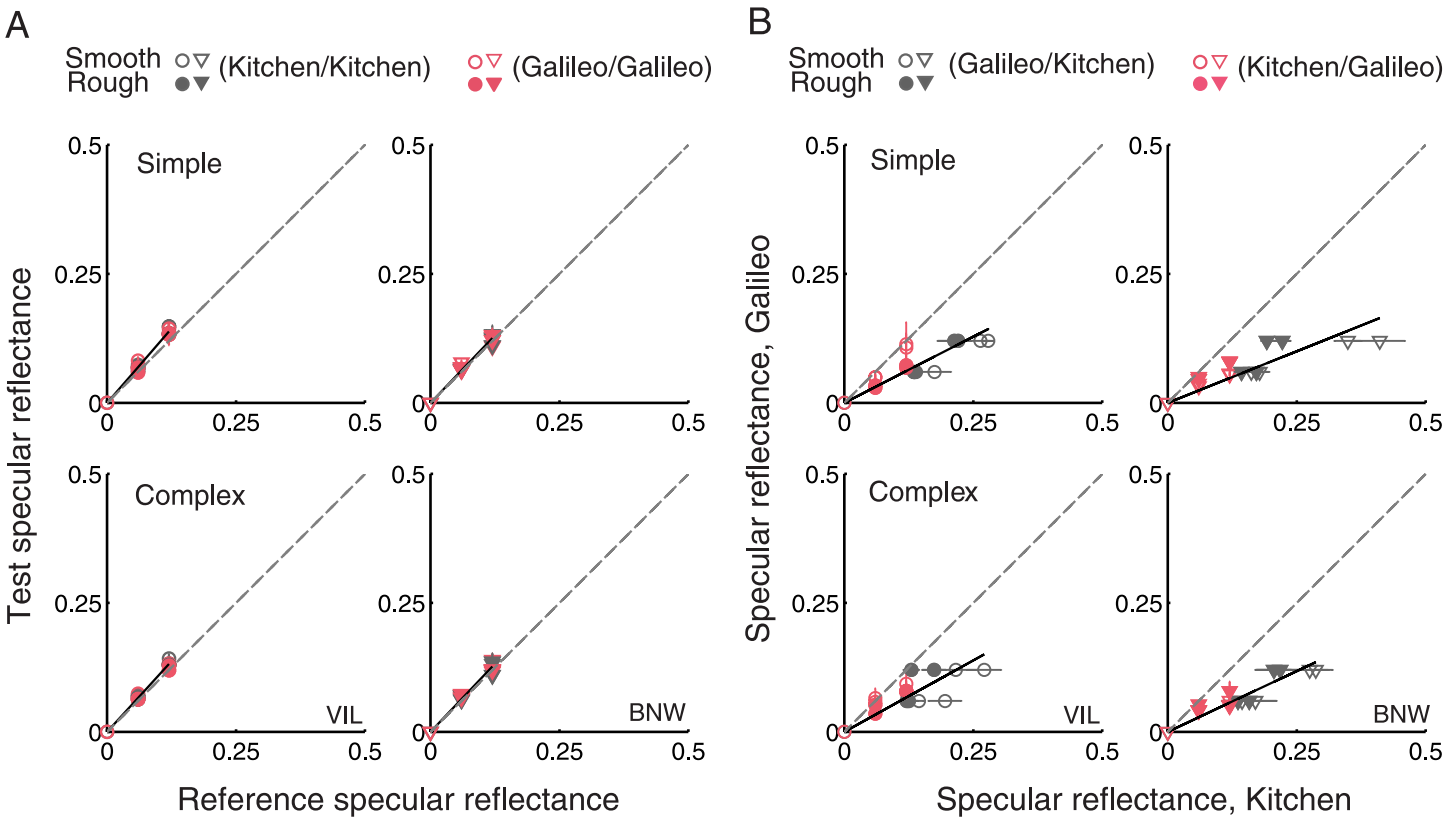


Figure 4. Specular component matches for the Kitchen/Galileo comparison. (A) Symmetric specular matches for observers (left) VIL and (right) BNW for the (top row) simple and (bottom row) complex conditions. Specular values for the Kitchen tests are plotted against specular values for the Kitchen references in gray, and specular values for the Galileo tests are plotted against specular values for the Galileo references in red. Settings for smooth surfaces are shown with open symbols and for rough surfaces with filled symbols. The unity line is shown with a gray dashed line. Error bars show the standard error of the mean over three matches. Black solid lines are regression lines fitted through the data points. Here one line was fit to the mean data shown in each panel and the lines were constrained to include the origin. Data shown include matches for reference spheres with different levels of diffuse reflectance. To avoid excessive clutter in the figure, these differences are not indicated explicitly. (B) Asymmetric specular matches for the same two observers for the (top row) simple and (bottom row) complex contexts. The specular value for the sphere seen under the Kitchen illuminant is always plotted on the *x*-axis, whether that sphere served as the reference (red) or the test (gray). Similarly, the specular value for the sphere seen under the Galileo illuminant is always plotted on the *y*-axis. Error bars show one standard error of the mean and are plotted vertically for the red symbols and horizontally for the gray symbols.

for all three light field comparisons. Here lines were fit separately to the data from each session, with the reported slopes representing the mean of the three fits and the error bars representing the standard error of this mean. The left panels show the slopes for the specular matches. A number of features of the data emerge from these plots. First, the slopes for all of the symmetric conditions were close to one for all observers (range 0.97–1.02). This confirms that our observers reliably performed the underlying matching task. Second, for each light field comparison the slopes for asymmetric specular matches deviated from one, at least for some observers. This deviation was most apparent for the Kitchen/Beach (Figure 6A) and Kitchen/Galileo comparisons (Figure 6C), where it was shown by all observers. It was also clearly present for observer MVI in the Kitchen/Campus comparison (Figure 6B); for this comparison, there was not much effect for observers VIL and BNW. The sign of the deviation is a matter of convention (depending on which light field is defined as the reference) but was consistent across observers in all cases. Third, there was little overall effect of changing from simple to complex background. Although for some conditions (Figure 6A, all observers) the deviations of the slope from unity were larger for the simple background, there was essentially no effect in other cases. If there is a systematic effect of background, further experiments would be needed to document it persuasively. Fourth, there were some individual differences in the strength of the effects. For observer VIL in Figure 6A, the slopes for the asymmetric matches were between 2.5 and 3, meaning that she had to adjust the specularity of the test sphere under the Beach illuminant to over two times that of the sphere under the Kitchen illuminant. For the two other observers, the effect was considerably smaller.

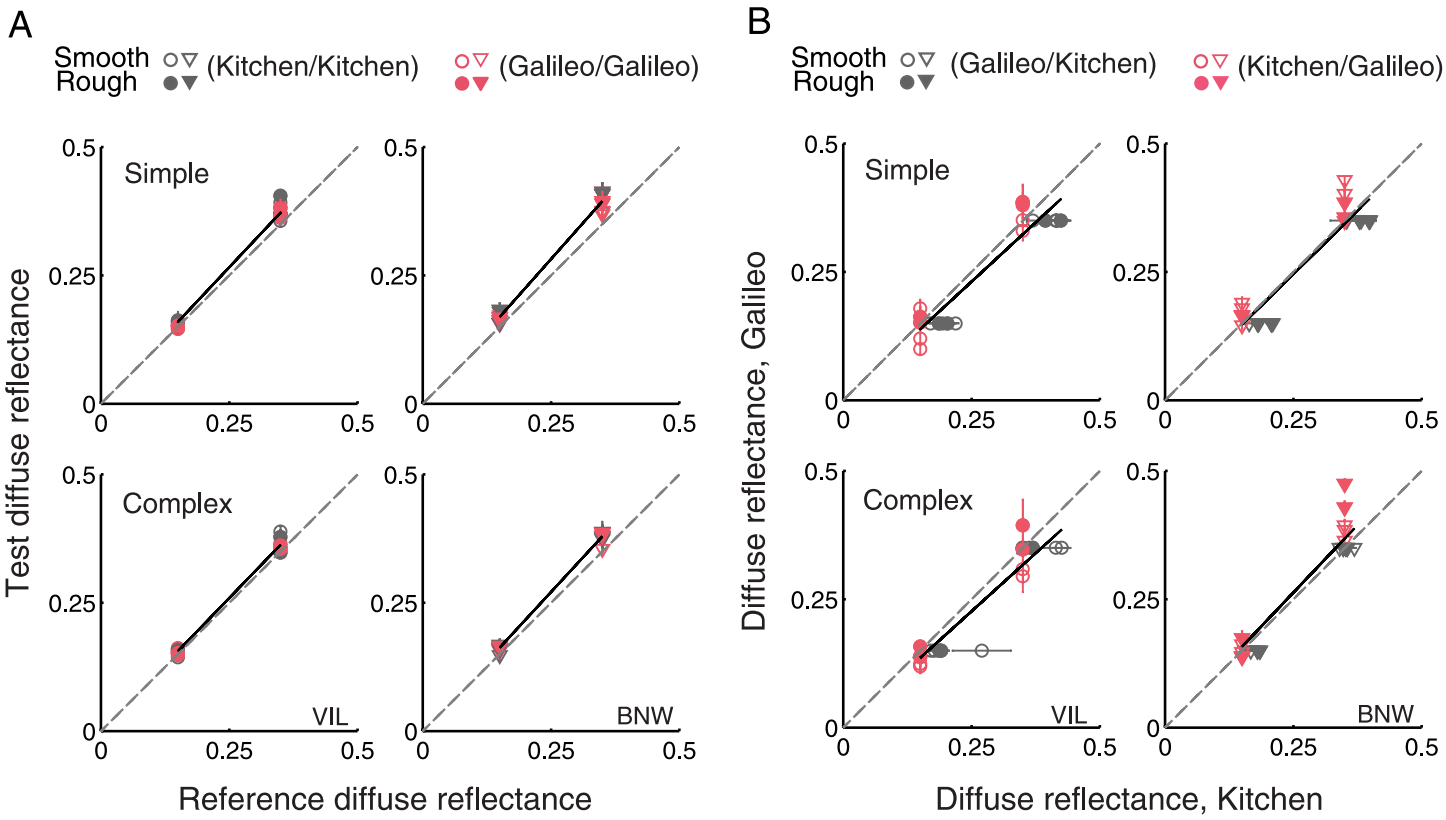


Figure 5. Diffuse component matches for the Kitchen/Galileo comparison. (A) Symmetric diffuse component matches for observers (left) VIL and (right) BNW for the (top row) simple and (bottom row) complex contexts. (B) Asymmetric diffuse component matches for observers (left) VIL and (right) BNW for the (top row) simple and (bottom row) complex contexts. Plotting conventions as in Figure 4.

The right panels in Figure 6 show the corresponding slopes for the diffuse matches. Here the matches were always close to veridical (slope range 0.89–1.28, means 0.97, 1.05, 0.97 for Figures 6A–6C).

## Component independence

We analyzed the data to ask whether the effects of light field geometry could be characterized independently for specular and diffuse components, and whether these effects were modulated by surface roughness. To the extent that such independence principles hold, the experimental enterprise of characterizing the effect of light field is greatly simplified, since it is not then necessary to measure the effect separately for all possible combinations of the reflectance components (see, e.g., Brainard & Wandell, 1992).

Diffuse component matches were independent of surface specular reflectance. This is illustrated in Figure 7, which shows the difference between the diffuse component matches for the low and medium specularities in the top panels and the difference between the diffuse component matches for the high and medium specularities in the bottom panels. Figure 7A shows this for the Kitchen/Beach comparison. Differences were scattered around zero for all observers and reference stimuli, which indicates that there were no systematic effects of stimulus specular reflectance on diffuse component matches. Figure 7B shows the differences for the Kitchen/Campus comparison, and Figure 7C for the Kitchen/Galileo comparison. For these two comparisons, differences were also scattered around zero. Bonferroni corrected *t*-tests confirmed that none of the differences were significantly different from zero (the six uncorrected *p*-values ranged from 0.05 to 0.93).

Specular component matches were also independent of diffuse reflectance. Figures 8A–8C show the differences between specular component matches for the two levels of diffuse reflectance for the Kitchen/Beach, Kitchen/Campus, and Kitchen/Galileo comparisons, respectively. Except for two data points from observer VIL in Figure 8A, the differences scatter around zero. Bonferroni corrected *t*-tests confirmed that none of the differences shown in Figure 8 were significantly different from zero (3 tests; uncorrected *p*-values of 0.02, 0.36, and 0.78 for Figures 8A–8C, respectively). We attribute the significant uncorrected *p*-value for Figure 8A to the outliers in the data.

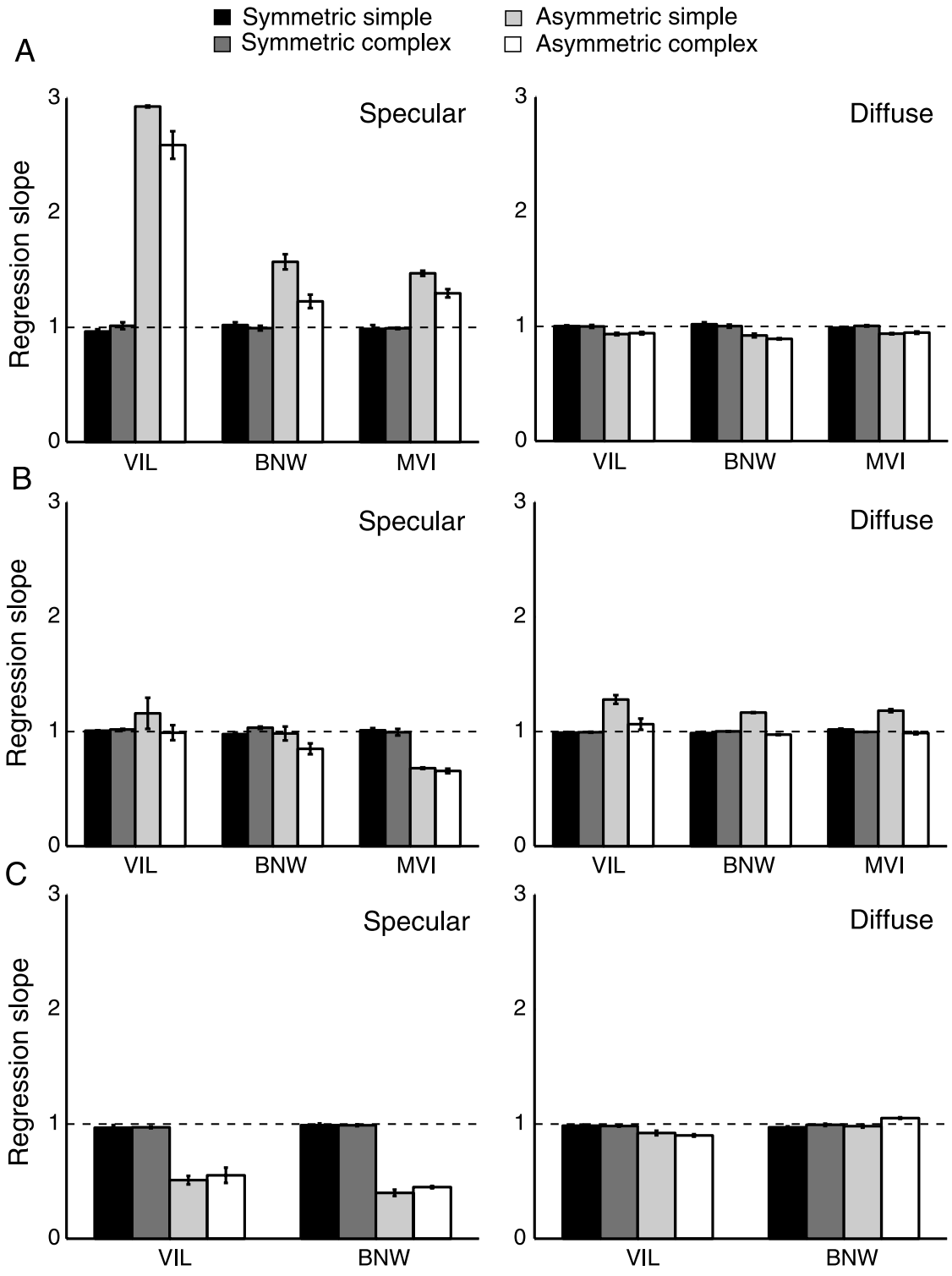


Figure 6. (A) Regression slopes for the Kitchen/Beach comparison. The left panel shows the slopes for the specular component matches; the right panel shows the slopes for the diffuse component matches. Each set of bars in each panel is for one observer. The bars in each group are from left to right: simple symmetric, complex symmetric, simple asymmetric, complex asymmetric. Slopes plotted were obtained by fitting the data for each of the three replications separately and taking the mean; error bars show the standard error of this mean. The horizontal black lines indicate veridical matches. (B, C) The slopes for the Kitchen/Campus comparison and for the Kitchen/Galileo comparison, respectively.

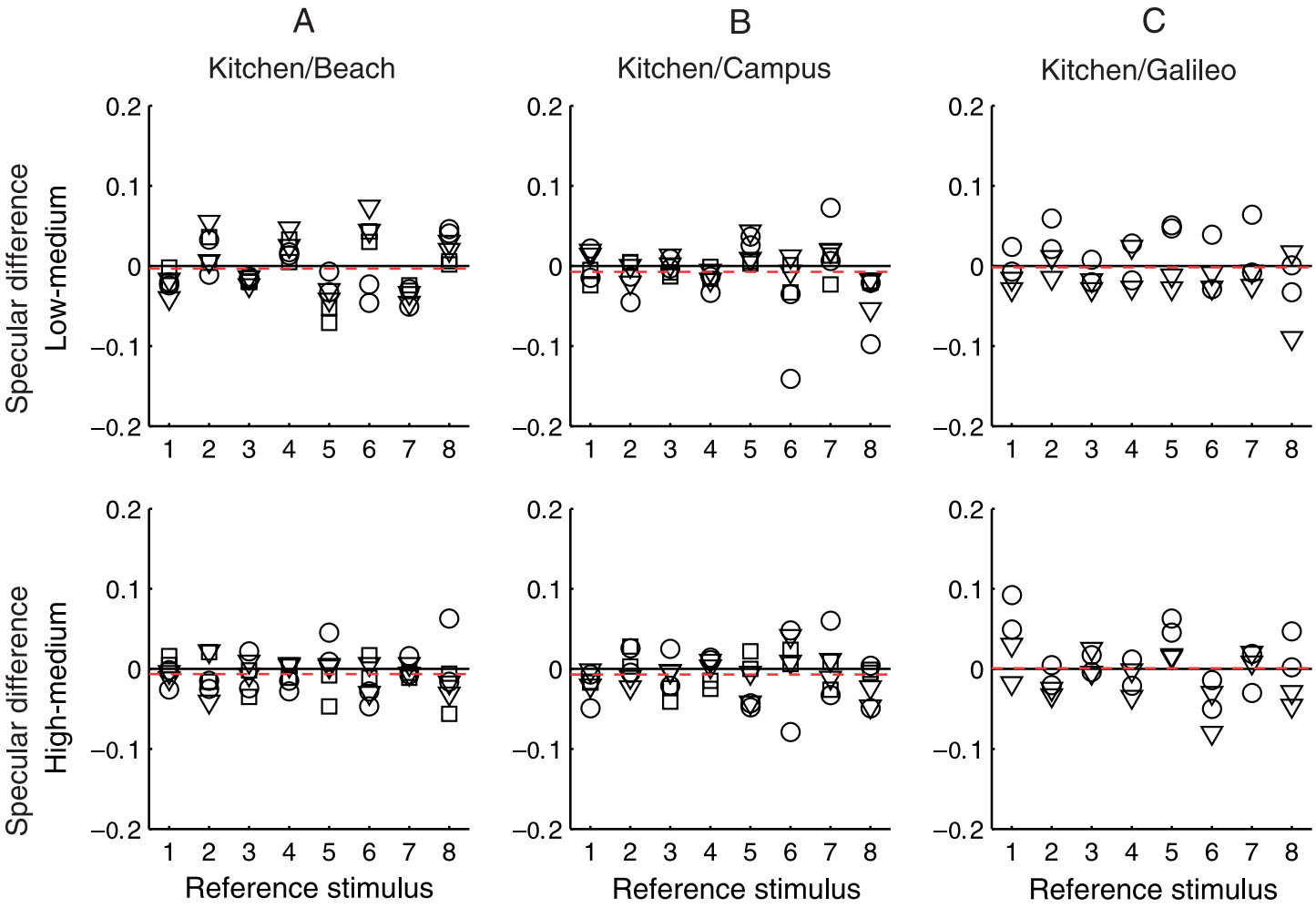


Figure 7. Effect of specular reflectance on diffuse component matches. (A) The top panel shows the difference between diffuse component matches for the low and medium reference specularities for the Kitchen/Beach comparison. The differences are plotted on a nominal *x*-axis with 8 levels (2 levels of diffuse reflectance × 2 levels of roughness × 2 light fields). The bottom panel shows the differences between diffuse component matches for the high and medium reference specularities for the same comparison. Symbols indicate different observers as follows: circle = VIL, triangle = BNW, square = MVI. The black solid line indicates zero difference. The red dashed line indicates the mean difference across reference stimuli. (B, C) Differences in diffuse component matches for the Kitchen/Campus and Kitchen/Galileo comparisons, respectively. Details are as in (A).

Finally, diffuse component matches for smooth surfaces were similar to the diffuse matches for rough surfaces. Specular component matches, however, were somewhat different for smooth and rough surfaces. This is illustrated in Figure 9, which shows the differences between specular matches for smooth and rough surfaces (green symbols) and the differences between the diffuse component matches for smooth and rough surfaces (gray symbols). The differences for the diffuse component matches were scattered around zero, whereas the differences for the specular component matches were on average below zero, especially for the Kitchen/Galileo comparison (Figure 9C). After Bonferroni correction, only the effect for the specular matches shown in Figure 9C was significantly different from zero (6 tests; uncorrected $p = 0.004$ for specular comparison in Figure 9C; the other five uncorrected $p$-values varied between 0.01 and 0.52).

# Discussion

## Summary

The primary goal of the experiments reported here was to measure the effects of geometric changes in the light field on perceived lightness and glossiness, and to test independence principles that can simplify the problem of characterizing these effects. In our hands, changing the light field has a substantial effect on perceived glossiness as assessed with our matching paradigm: for many of our conditions the asymmetric matches for the specular component differed from veridical. We also found that several independence principles hold to good approximation. The effect of changing light field geometry on specular component matches is independent of the diffuse

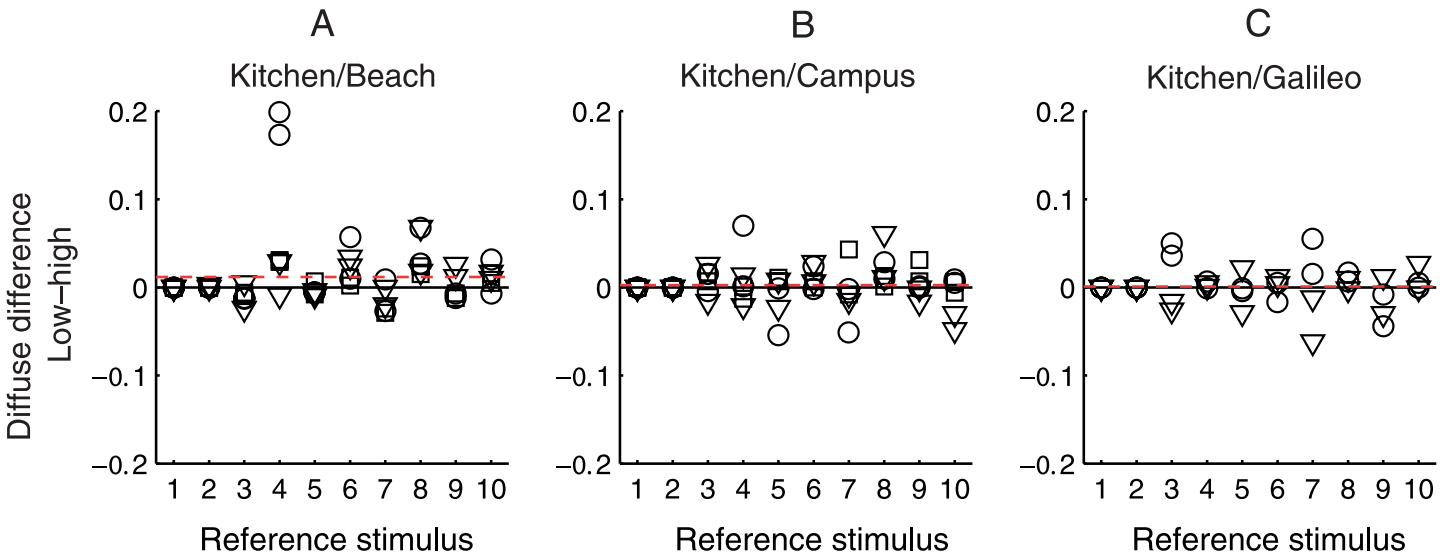


Figure 8. Effect of diffuse reflectance on specular component matches. (A) Differences between specular component matches for the high and low reference diffuse reflectances for the Kitchen/Beach comparison. The differences are plotted on a nominal *x*-axis with 10 levels (3 levels of specular reflectance × 2 levels of roughness × 2 light fields taken as reference; but since the diffuse sphere is the same for both levels of roughness, there are a total of only 10 levels). Symbols indicate different observers as follows: circle = VIL, triangle = BNW, square = MVI. The black solid line indicates zero difference. The red dashed line indicates the mean difference across reference stimuli. (B, C) Specular component matches for the Kitchen/Campus and Kitchen/Galileo comparisons, respectively. Details are as in (A).

component of surface reflectance; the effect of changing light field geometry on diffuse component matches is independent of the specular component of surface reflectance; and the effect of changing light field geometry on both reflectance components is approximately independent of the roughness of the specular reflectance component; for one light field pair, there was a small but statistically significant effect of roughness on the specular matches. Below, we expand on these conclusions and relate them to the prior literature.

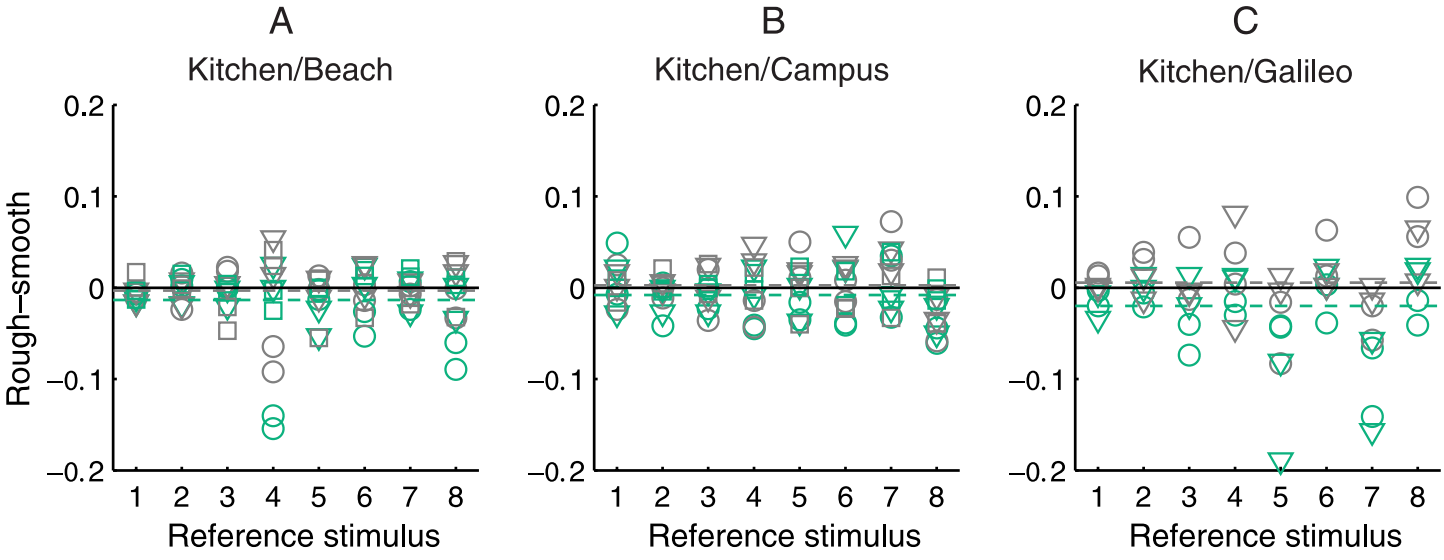


Figure 9. Effect of surface roughness on diffuse and specular matches. (A) Gray symbols show the differences between diffuse component matches for rough and smooth surfaces for the Kitchen/Beach comparison. The differences are plotted on a nominal *x*-axis with 8 levels (2 levels of diffuse reflectance × 2 levels of nonzero specular reflectance × 2 light fields). Green symbols show the same differences for specular matches. Black solid line indicates zero difference. The gray and green dashed lines indicate mean differences over reference stimuli for diffuse and specular matches, respectively. Symbols indicate different observers as follows: circles = VIL, triangles = BNW, squares = MVI. (B, C) The same for the Kitchen/Campus and Kitchen/Galileo comparisons, respectively.

## Prior measurements of effect of light field

Several studies have investigated glossiness perception for 3D surfaces under a single illuminant (Beck & Prazdny, 1981; Berzhanskaya, Swaminathan, Beck, & Mingolla, 2005; Blake & Bülthoff, 1990; Pellacini, Ferwerda, & Greenberg, 2000; Wendt, Faul, & Mausfeld, 2008). The constancy of glossiness perception under varying illumination, however, has only recently begun to receive attention (Doerschner et al., 2010; Fleming et al., 2003; Obein, Knoblauch, & Viénot, 2004; Pont & te Pas, 2006; te Pas & Pont, 2005).

Fleming et al. (2003) used asymmetric matching to measure perceived glossiness for spheres simulated under either real-world light fields or geometrically simple illuminants. Our methods are similar to theirs. Fleming et al.'s broad conclusion was that observers' matches were quite veridical when the light fields were geometrically complex, although this veridicality broke down for geometrically simple illuminants. Their data, however, do show some light field comparisons where there is a definite effect of light field (see, for example, the top middle panel of their Figure 12). We see the difference between their conclusion and ours as primarily one of emphasis. As we will discuss below, the question of how to evaluate whether a deviation from veridical matching in this type of experiment is large or small is subtle. Doerschner et al. (2010) also studied the effect of light field changes on perceived glossiness. They abandoned asymmetric matching in favor of forced-choice comparisons of perceived glossiness across light fields and found deviations from gloss constancy for almost all light field comparisons. In this regard, our data for specular component matches agree well with those of Doerschner et al.

Fleming et al. (2003) demonstrated that the context in which a sphere was viewed did not have an appreciable effect on perceived glossiness. Along similar lines, Sharan, Li, Motoyoshi, Nishida, and Adelson (2008) pointed out that for bumpy surfaces simulated under directional light sources, background was not necessary for estimating albedo. These observations indicate that for stimuli rendered under complex lighting and/or with complex surface structure, background has little influence on appearance. Our results largely agree with these earlier observations: matches for the complex and for the simple contexts were similar for two out of three light field pairs. We did note a small effect toward veridical matching for the Kitchen/Beach comparison, but our data set is not extensive enough to generalize from this one case. If there is a systematic effect of background, it is subtle.

Pont and te Pas (2006) and te Pas and Pont (2005) studied the perception of material properties under geometrically varying illumination. They used a discrimination paradigm and asked how well observers could discriminate between changes of illumination and changes of object reflectance. They did this both for computer-rendered stimuli (Pont & te Pas, 2006) and for photographs of real objects (te Pas & Pont, 2005). Their conclusion was that the rendered stimuli did not support discrimination, while above-chance performance was possible with the photographed stimuli. Although their data are not directly comparable with ours, their results do sound a cautionary note about the generality of data obtained using computer-rendered stimuli.

Obein et al. (2004) used a difference scaling method to derive the relation between perceived gloss and the specular reflectance component of black papers. Their observers viewed physical samples presented in a light booth under directional illumination, and they varied the direction of the light source in relation to the samples. Obein et al. found a monotonic but nonlinear relationship between physical surface specularity and perceived gloss, and the shape of this relation was very similar for the two illuminant directions employed. Because the scales for each light direction were derived from difference scaling data collected only within that condition, however, their data are silent with respect to any shifts in the absolute magnitude of perceived gloss across the two light directions. That is, the "gloss constancy" they report is a constancy of relative perceived glossiness within single light field conditions, and their data do not make predictions about whether asymmetric matches across a light field change would be veridical.

## Luminance histogram statistics

Several papers have suggested that simple statistics (e.g., mean, skewness) extracted from the luminance histograms of objects may provide the key cues used by the visual system to determine object lightness and glossiness (Motoyoshi, Nishida, Sharan, & Adelson, 2007; Nishida & Shinya, 1998; Sharan et al., 2008). This broad conclusion has been criticized, however, on the grounds that the stimulus conditions employed did not sufficiently dissociate variation in histogram statistics from variation in object material properties (Anderson & Kim, 2009). Some of our light field manipulations had a large effect on the luminance histogram of the rendered spheres, even when their reflectance was held constant (see Figure 2). It is thus of interest to ask whether our observers' matches are predicted by simple luminance histogram statistics.

The panels in the top row of Figure 10 show the mean, standard deviation, skewness, and kurtosis of reference spheres under the Kitchen light field ($x$-axis) against the same statistics for the same spheres under the Galileo light field ($y$-axis). As noted above, there was little effect of light field geometry on the histogram means because of the way we normalized light field intensities. For the other three statistics, the change in light field had a large effect. The bottom panels show the same statistics, but rather than computing them from the same physical spheres

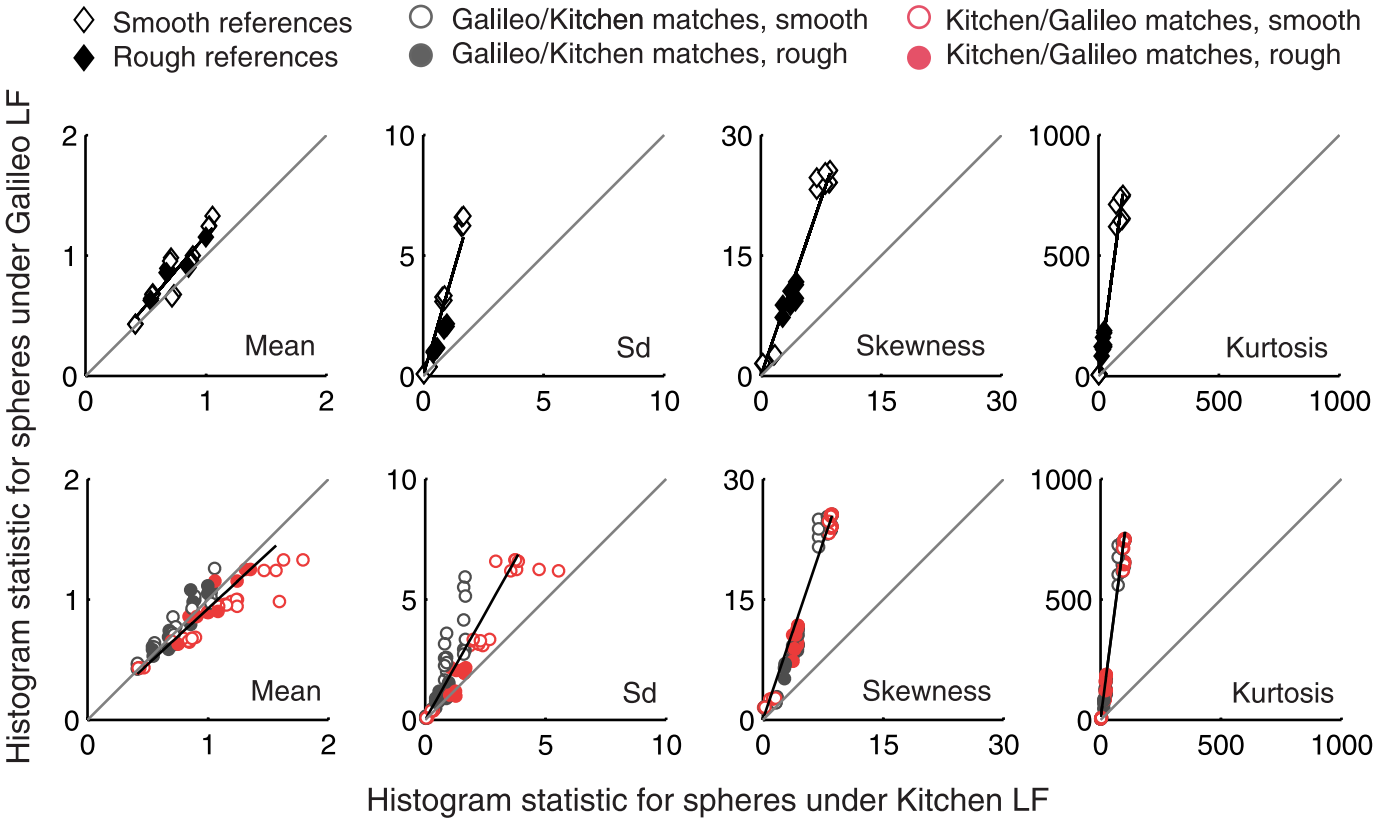


Figure 10. Luminance histogram statistics for spheres under the Galileo light field are shown against statistics for spheres under the Kitchen light field. The top row shows histogram statistics (from left to right: mean, standard deviation, skewness, and kurtosis) across the stimulus set under the Kitchen light field on the *x*-axis against the same statistics for the stimulus set under the Galileo light field on the *y*-axis. Open symbols stand for smooth spheres, and closed symbols stand for rough spheres. The bottom row shows statistics of perceptually matched test and reference stimuli across the change from the Kitchen to the Galileo light field. The plotting convention is the same as in the data figures above: gray symbols represent the cases where stimuli under Kitchen were matched to stimuli under Galileo, and the red symbols represent the reverse case. Data for all observers are presented, with matches for each observer plotted as separate points. See supplemental materials for similar figures for the other light field comparisons.

across the light field change, we computed them from the asymmetric matches. Each point represents the statistic value for a pair of spheres judged to match across the illuminant change. For an observer who matched spheres based on one of these statistics, the data in the bottom panel for that statistic would fall along the positive diagonal of the plot. This pattern is approximated only for the histogram mean, and this is not diagnostic since very little variation in the mean was produced by the light field change. For the other three statistics, the data deviate strongly from the positive diagonal. Specifically, these data falsify the hypothesis that specular component matches are predicted solely by luminance histogram skewness (second panel from the right).

Note that for an observer who judged the same physical sphere to have the same lightness and glossiness across the illuminant change, the data in the bottom four panels would fall along the same lines as the data in the corresponding top four panels. Although as shown in Results section this is not exactly what occurs, the data tend in this direction relative to the positive diagonal.

The supplemental materials provide additional figures in this same format for all of our light field comparisons, and examination of these figures supports the same conclusions. To present this in summary format, Figure 11 shows the slopes of the regression lines for each histogram statistic and light field comparison. For each statistic other than the mean, the slopes obtained from the asymmetric matches can deviate substantially from unity and are, in general, close to those obtained from the analysis of physically constant spheres. This is particularly true for skewness and kurtosis. Repeating the analysis with log luminance histograms produced essentially the same results for skewness and kurtosis (see Figure 5 in the supplemental materials). It turns out, however, that our light field changes had little effect on the standard deviation of the log luminance histogram and thus are not diagnostic with respect to this statistic.

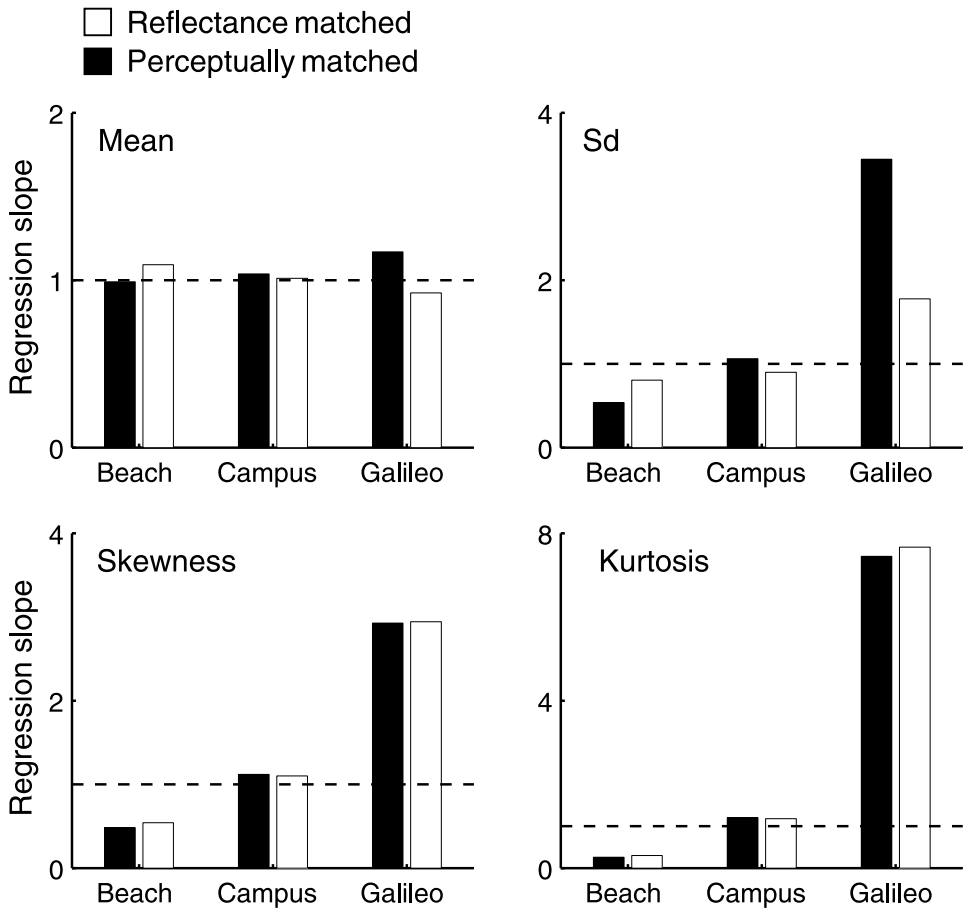


Figure 11. Slopes of the regression lines fitted to luminance histogram statistics for each pair of light fields. Each panel is for one statistic. The three sets of bars in each panel are for each of the three light field comparisons. The names under the bars indicate one of the light fields in a pair; the other one was always the Kitchen light field. Black bars indicate slopes between reference sphere statistics for a pair of light fields; white bars indicate slopes between statistics of perceptually matched spheres across the same pair of light fields. Dashed black lines mark slope of 1. Supplemental materials present the corresponding analysis when histogram statistics are computed from log luminance.

Our conclusion that simple luminance histogram statistics do not fully account for the perception of glossiness does not indicate a contradiction between our data and those of Motoyoshi et al. (2007). Rather, our experiments involved comparisons of appearance across changes of light field, whereas their experiments involved ratings of glossiness under a single light field. It is the effect of light field change that drives our conclusion (see Anderson & Kim, 2009).

The observers in our experiments were able to independently adjust only two surface reflectance parameters. It is conceivable that if observers had had more degrees of freedom for their adjustments, they would have set different perceptual matches and that the matches would have been well predicted by simple luminance histogram statistics. Although we did not investigate this possibility experimentally, we did evaluate it informally as follows. We generated image pairs where the spheres were (i) matched in reflectance, (ii) perceptually matched based on the average data from our observers, and (iii) where the sphere histograms were forced to match by direct remapping of the luminance histogram. Examples for the Kitchen/Beach and the Kitchen/Galileo light field comparisons are shown in Figures 12A and 12B, respectively. The first rows show the reflectance-matched pairs; the middle rows show the perceptually matched pairs; and the bottom rows show the histogram-matched pairs. Because the published images must be tone-mapped, and because the exact gamma correction appropriate for the printed and online versions is uncertain, the images in Figure 12 do not fully represent the type of high-dynamic-range stimuli used in our experiment. Even so, the renderings of the perceptually matched pairs appear more similar than those of the corresponding reflectance-matched and histogram-matched pairs. This is consistent with our conclusion that across light field changes, factors other than simple luminance histogram statistics affect glossiness perception.

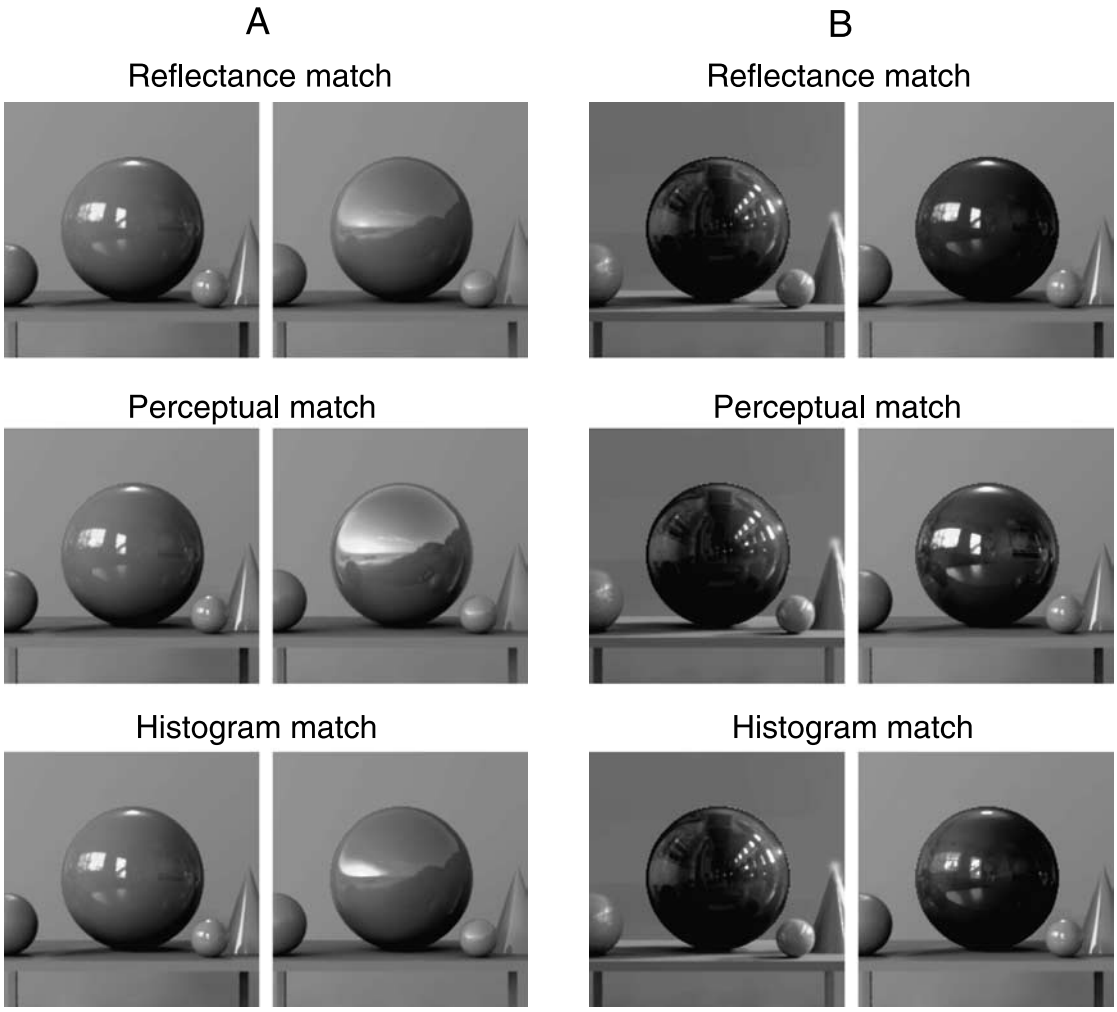


Figure 12. Renderings of reflectance-, perceptually, and histogram-matched spheres. (A) The top row shows a pair of spheres with reflectance parameters $\rho_d$ = 0.35, $\rho_s$ = 0.06, and $\alpha$ = 0.001 simulated under the (left) Kitchen and (right) Beach light fields. The middle row shows (left) the same reference sphere under the Kitchen light field and (right) a sphere under the Beach light field that is a perceptual match to the reference sphere as determined by the average (over observers) of our experimental data. The bottom row again shows the same reference sphere on the left and a sphere on the right whose histogram has been matched to that of the reference sphere. (B) The reference spheres have the reflectance parameters $\rho_d$ = 0.15, $\rho_s$ = 0.06, and $\alpha$ = 0.001 and are simulated under the (left) Galileo and (right) Kitchen light fields. Other details as in (A).

Motoyoshi et al. (2007) acknowledged that the spatial structure of the light reflected from objects plays a role in how surfaces are perceived and demonstrated this by showing pixel-scrambled versions of their stimuli (see also Sharan et al., 2008). Pixel scrambling perfectly preserves the luminance histogram but has a large effect on how a stimulus is perceived. Anderson and Kim (2009) showed a set of image manipulations less extreme than pixel scrambling and across which the predictive value of skewness is low. Specifically, Anderson and Kim (2009) rotated and translated highlights in photographs relative to the underlying object and then applied a pointwise luminance nonlinearity to keep the luminance histogram skew constant. Apparent glossiness was strongly affected by this manipulation. Anderson and Kim (2009) also produced images where the light field was such that an object with diffuse reflectance had a positively skewed histogram but nonetheless appeared matte (see their Figure 9). Our data are consistent with this latter demonstration of Anderson and Kim (2009): we show that when histogram statistics vary due to naturally occurring changes in light field, perceived glossiness assessed by matching is not predicted solely by histogram skew. Note, however, that each of our light field changes simultaneously affected both the luminance histogram and the spatial pattern of the light reflected from the spheres. Because of this covariation, our data are not ideally suited for drawing strong inferences about the precise nature and extent of the interaction between photometric and geometric information in material perception. In the

supplemental materials, we provide image pairs that match perceptually across changes in light field, so that future investigators can determine whether their models are consistent with our experimental data.

## Independence principles

A focus of this paper has been on formulating and testing independence principles that support reduction in the number of conjoint experimental manipulations that need to be explored to develop an empirical foundation for object surface perception. Because of the explosion of stimulus parameter combinations that must be considered when one considers object BRDF, shape, and light field geometry, there are many such principles that could be explored. Other laboratories have considered other aspects of independence.

Pellacini et al. (2000) used multi-dimensional scaling to derive a two-dimensional model of how the specular component of object reflectance is represented perceptually. They described these dimensions as contrast gloss and distinctness-of-image gloss and noted that objects also have perceived lightness. They then examined whether each of these three dimensions was the perceptual correlate of a single easily described stimulus variable, by varying object reflectance within the Ward (1992) BRDF model. Asking this type of question is motivated by the same broad considerations that motivated us to test independence principles: a positive conclusion allows simplification of future empirical studies. Pellacini et al. concluded that distinctness-of-image gloss was affected primarily by the roughness parameter of the Ward model and that lightness was affected primarily by the diffuse reflectance component. Contrast gloss, however, was affected both by the strength of the specular reflectance component and by the diffuse reflectance component. Note that our data do not speak to the principles assessed by Pellacini et al., since they examined the perceptual representation of object reflectance within a single light field context. We, on the other hand, looked at how the effect of light field depended on various stimulus parameters.

In their asymmetric matching experiments across changes of light field, Fleming et al. (2003) included conditions where observers matched either the strength of the specular component or the roughness of the specular component. They note in passing that specular component matches were independent of surface roughness and that roughness matches were independent of surface specularity. Our data are consistent with their first conclusion. Although there was a small effect of roughness on the specular matches for one light field pair, the approximation that specular matches are independent of roughness is quite good. Since we did not ask observers to match surface roughness, our data are silent with respect to their second conclusion.

## Consistency across sets of asymmetric matches

The generality of conclusions drawn from experiments that rely on perceptual comparisons between stimuli viewed in different contexts rests on the assumption that data collected for a set of such pairwise comparisons are self-consistent. One type of consistency is that asymmetric matches should not depend on which object is adjusted. Violations of this type of consistency were reported by Doerschner et al. (2010) and Fleming et al. (2003). In the former case, some of the violations were large. Both Doerschner et al. and Fleming et al. attribute the inconsistencies to a nonspecific response bias. Because of the inconsistencies, Doerschner et al. abandoned asymmetric matching for their main experiments and turned to a forced-choice method in which observers indicated which of two objects appeared most glossy.

The matching data reported here do not show large inconsistencies when the roles of test and reference context are reversed. Our analysis aggregates paired conditions in which each light field served as reference. Because any inconsistencies in our matching data were small and because we measured asymmetric matches in both directions for each pair of light fields, we believe that our data suffer at most minor contamination from any matching response bias. We have, however, observed larger inconsistencies of the sort reported by Doerschner et al. (2010) and Fleming et al. (2003) in preliminary experiments that extend the range of stimulus parameters studied. For those conditions, we found that changing to a forced-choice procedure substantially reduced the inconsistencies, in agreement with Doerschner et al. (2010). We plan to use forced-choice methods for future experiments.

Doerschner et al. (2010) also examined whether pairwise comparisons across light fields satisfied a second important consistency property referred to as transitivity. Suppose we have measurements of the effect of changing light field from A to B and measurements of the effect of changing light field from B to C. Transitivity means that the effect of changing from A to C is predicted by the concatenation of the effect of A to B and B to C. Doerschner et al. concluded that transitivity holds well for glossiness judgments, a reassuring result.

## Limitations

We note here a few limitations of our current study.

We used a high-dynamic-range display so that our stimuli captured the full range of luminances representative of specular stimuli presented under real-world light fields. Although we were able to present a high dynamic range, there were limits on the maximum luminance of the display. This in turn meant that the mean luminance of our stimuli was rather low, around 1.5 cd/m$^2$. Whether

performance is invariant with mean luminance is an aspect that requires future exploration.

We have tested several independence principles, but there remains other important principles to be examined. One important addition to our current paradigm would be to vary roughness in addition to diffuse reflectance and the strength of the specular component. Other aspects of interest are the overall intensity of the light fields, their spectral properties, the spectral properties of the objects' reflectances, and the shape of the objects. In addition, it will be of interest to extend our current measurements to a larger range of surface reflectance parameters so as to probe the limits over which the independence principles hold.

### Constancy

The specular component matches of our observers deviated from those that would be obtained had they veridically matched object reflectance parameters. In this sense, our observers failed to show what might be termed glossiness constancy. Deviations from perfect constancy are also found reliably in the lightness and color literature (for a review, see, e.g., Brainard, 2004). In those literatures, it is common to quantify the degree of constancy using an index, where the index is based on a comparison of the data to two reference points. One reference point is the constancy prediction, obtained by positing that observers match object reflectance. The other reference point is a no constancy prediction, obtained by positing that the visual system makes no adjustment for a change in viewing context. Here the prediction is made on the basis of equating the photometric and colorimetric properties of the test stimuli in the image. That is, a no constancy prediction is made by assuming that observers make their matches by equating image properties rather than object properties. The use of the two reference points provides a natural scale against which to judge deviations from constancy. Although constancy indices provide only a broad strokes summary, they have proved valuable for framing the nature of performance (Arend, Reeves, Schirillo, & Goldstein, 1991; Brainard, Brunt, & Speigle, 1997; Delahunt & Brainard, 2004; Hansen, Walter, & Gegenfurtner, 2007; Kraft & Brainard, 1999; Lucassen & Walraven, 1996; Smithson & Zaidi, 2004).

Here, it also seems desirable to find an index that provides a sense of whether the deviations from constancy are large or small. This has proved difficult, however, because it is not clear what to use as the no constancy reference point. When the light field is changed, there is no matching stimulus that is identical to the reference stimulus. For this reason, it is not clear how to scale the deviations from physical matching and thus to decide whether we should view performance as close to constant or far from it.

This noted, we can use matching based on luminance histogram statistics as a proxy for how a visual system with no constancy would perform. Recall that the positive diagonals in the top panels of Figure 10 and corresponding figures in the supplemental materials predict asymmetric matches for observers who match luminance histogram statistics. These positive diagonals can thus be regarded as a no constancy prediction. The regression lines in the same panels show how the same statistics would be matched if observers made veridical matches. By comparing the top and bottom panels of this figure (see also Figure 11), we see that observer data (bottom panels), expressed in terms of the histogram statistic values at the match, are in general closer to the predictions of constancy than to the predictions of no constancy.

## Acknowledgments

This work was supported by the NIH Grants RO1 EY10016 and P30 EY001583 and by a Young Investigator Grant to MO from the Emil Aaltonen Foundation.

Commercial relationships: none.
Corresponding author: Maria Olkkonen.
Email: mariaol@sas.upenn.edu.
Address: Department of Psychology, University of Pennsylvania, 3401 Walnut St., Philadelphia, PA 19104, USA.

## References

Alldrin, N., Zickler, T., & Kreigman, D. (2008). Photometric stereo with non-parametric and spatially-varying reflectance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, June 23–28* (pp. 1–8). IEEE.

Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of Vision, 9*(11):10, 1–17, http://www.journalofvision.org/content/9/11/10, doi:10.1167/9.11.10. [PubMed] [Article]

Arend, L. E., Reeves, A., Schirillo, J., & Goldstein, R. (1991). Simultaneous color constancy: Paper with diverse Munsell values. *Journal of the Optical Society of America A, 8,* 661–672.

Beck, J., & Prazdny, S. (1981). Highlights and the perception of glossiness. *Perception & Psychophysics, 30,* 407–410.

Berzhanskaya, J., Swaminathan, G., Beck, J., & Mingolla, E. (2005). Remote effects of highlights on gloss perception. *Perception, 34,* 565–575.

Blake, A., & Bülthoff, H. (1990). Does the brain know the physics of specular reflection? *Nature, 343,* 165–168.

Brainard, D. H. (1995). Colorimetry. In M. Bass (Ed.), *OSA handbook of optics* (vol. 1, pp. 26.1–26.54). New York: McGraw-Hill.

Brainard, D. H. (2004). Color constancy. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 948–961). Cambridge, MA: MIT Press.

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image. I. Asymmetric matches. *Journal of the Optical Society of America A, 14,* 2091–2110.

Brainard, D. H., & Wandell, B. A. (1992). Asymmetric color matching: How color appearance depends on the illuminant. *Journal of the Optical Society of America A, 9,* 1433–1448.

Dana, K. J., van Ginneken, B., Nayar, S. K., & Koenderink, J. J. (1999). Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics, 18,* 1–34.

Debevec, P. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH'98: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 189–198). New York: ACM.

Delahunt, P. B., & Brainard, D. H. (2004). Does human color constancy incorporate the statistical regularity of natural daylight? *Journal of Vision, 4*(2):1, 57–81, http://www.journalofvision.org/content/4/2/1, doi:10.1167/4.2.1. [PubMed] [Article]

Doerschner, K., Boyaci, H., & Maloney, L. T. (2010). Estimating the glossiness transfer function induced by illumination change and testing its transitivity. *Journal of Vision, 10*(4):8, 1–9, http://www.journalofvision.org/content/10/4/8, doi:10.1167/10.4.8. [PubMed] [Article]

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision, 3*(5):3, 347–368, http://www.journalofvision.org/content/3/5/3, doi:10.1167/3.5.3. [PubMed] [Article]

Griffin, L. D. (1999). Partitive mixing of images: A tool for investigating pictorial perception. *Journal of the Optical Society of America A, 16,* 2825–2835.

Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision, 7*(4):2, 1–15, http://www.journalofvision.org/content/7/4/2, doi:10.1167/7.4.2. [PubMed] [Article]

International Electrotechnical Commission (1999). International electrotechnical commission standard 61966-2-1.

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 307–312.

Lucassen, M. P., & Walraven, J. (1996). Color constancy under natural and artificial illumination. *Vision Research, 36,* 2699–2711.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature, 447,* 206–209.

Newhall, S., Nickerson, D., & Judd, D. (1943). Final report of the O.S.A. subcommittee on the spacing of Munsell colors. *Journal of the Optical Society of America, 33,* 385–412.

Nickerson, D. (1957). *Spectrophotometric data for a collection of Munsell samples* (Tech. Rep.). Washington, DC: US Department of Agriculture.

Nickerson, D., & Wilson, D. H. (1950). Munsell reference colors now specified for nine illuminants. *Illumination Engineering, 45,* 507–517.

Nishida, S., & Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. *Journal of the Optical Society of America A, 15,* 2951–2965.

Obein, G., Knoblauch, K., & Viénot, F. (2004). Difference scaling of gloss: Nonlinearity, binocularity, and constancy. *Journal of Vision, 4*(9):4, 711–720, http://www.journalofvision.org/content/4/9/4, doi:10.1167/4.9.4. [PubMed] [Article]

Oren, M., & Nayar, S. K. (1994). Generalization of Lambert's reflectance model. In *SIGGRAPH'94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (pp. 239–246). New York: ACM.

Pellacini, F., Ferwerda, J. A., & Greenberg, D. P. (2000). Toward a psychophysically-based light reflection model for image synthesis. In *SIGGRAPH'00: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 55–64). New York: ACM.

Pont, S. C., & te Pas, S. F. (2006). Material-illumination ambiguities and the perception of solid objects. *Perception, 35,* 1331–1350.

Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., & Trentacoste, M., et al. (2004). High dynamic range display systems. In J. Marks (Ed.), *SIGGRAPH'04: Proceedings of the 31st Annual Conference on Computer Graphics and Interactive Techniques* (pp. 760–768). New York: ACM.

Sharan, L., Li, Y., Motoyoshi, I., Nishida, S., & Adelson, E. H. (2008). Image statistics for surface reflectance perception. *Journal of the Optical Society of America A, 25,* 846–865.

Shevell, S. K., & Kingdom, F. A. A. (2008). Color in complex scenes. *Annual Review of Psychology, 59,* 143–166.

Smithson, H. (2005). Sensory, computational and cognitive components of human color constancy. *Philosophical Transactions of the Royal Society B, 360,* 1329–1346.

Smithson, H., & Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and levels of reference. *Journal of Vision, 4*(9):3, 693–710, http://www.journalofvision.org/content/4/9/3, doi:10.1167/4.9.3. [PubMed] [Article]

te Pas, S. F., & Pont, S. C. (2005). A comparison of material and illumination discrimination performance for real rough, real smooth and computer generated smooth spheres. In S. N. Spencer (Ed.), *APGV'05: Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization* (pp. 75–81). New York: ACM.

Ward, G. J. (1992). Measuring and modeling anisotropic reflection. In A. Glassner (Ed.), *SIGGRAPH 92: Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 459–472). New York: ACM, ACM Press.

Ward, G. J. (1994). The radiance lighting simulation and rendering system. In A. Glassner (Ed.), *SIGGRAPH 94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (pp. 459–472). New York: ACM, ACM Press.

Wendt, G., Faul, F., & Mausfeld, R. (2008). Highlight disparity contributes to the authenticity and strength of perceived glossiness. *Journal of Vision, 8*(1):14, 1–10, http://www.journalofvision.org/content/8/1/14, doi:10.1167/8.1.14. [PubMed] [Article]

Xiao, B., & Brainard, D. H. (2008). Surface gloss and color perception of 3D objects. *Visual Neuroscience, 25,* 371–385.

Yu, Y., Debevec, P., Malik, J., & Hawkins, T. (1999). Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *SIGGRAPH'99: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 215–224). New York: ACM Press.