Published in final edited form as:
Psychol Sci. 2008 February ; 19(2): 196–204. doi:10.1111/j.1467-9280.2008.02067.x.

CONJOINT MEASUREMENT OF GLOSS AND SURFACE TEXTURE

Yun-Xian Ho[1], **Michael S. Landy**[1,2], and **Laurence T. Maloney**[1,2]
[1]Department of Psychology, New York University, New York, NY
[2]Center for Neural Science, New York University, New York, NY

Abstract

Previous research on visual perception of surface material has focused primarily on smooth, matte surfaces, neglecting surfaces with pronounced three-dimensional (3D) texture or specularity. Furthermore, studies have typically focused on single material properties, with no consideration of possible interactions. In this study, we used a conjoint measurement design to determine how observers represent perceived 3D texture ("bumpiness") and specularity ("glossiness") and modeled how each of these two surface material properties affects perception of the other. Observers made judgments of "bumpiness" and "glossiness" of surfaces that varied in both surface texture and specularity. We found that a simple additive model captures visual perception of texture and specularity and their interactions. We quantify how changes in each surface material property affect judgments of the other. Conjoint measurement is potentially a powerful tool for analyzing surface material perception in realistic environments.

Keywords

material perception; gloss; mesotexture; conjoint measurement

Surfaces can be analyzed at many spatial scales. Koenderink & Van Doorn (1996) delineate three distinct scale ranges in the perception of surface properties: *megascale*, *mesoscale*, and *microscale*. Megascale variations are defined by differences in global shape, e.g., the spherical shape of the orange in Figure 1. Mesoscale refers to the properties of the orange skin, i.e., the irregular and bumpy surface (mesotexture) of the orange which is similar to a lemon, but dramatically different from a smooth rubber ball. Variations at the microscale level include changes in the otherwise invisible surface structure of the orange skin that result in its glossy appearance. To visually identify an orange, we take into account the geometry at all three scales. Certain perceptual tasks such as discriminating an orange from an orange-colored rubber ball require an ability to detect differences in structural geometry at the meso- and microscales, i.e. material properties. Judgments of surface material are made frequently and effortlessly, yet surprisingly little is understood about how the visual system represents materials.

One difficulty in studying material perception is that light interacts with surfaces in a complex way. As a consequence, it may be difficult for a visual system to estimate surface material properties independently of one another and of illumination and viewing geometry. Failures of "material constancy" have been demonstrated for a variety of surfaces viewed under different illumination conditions (Pont & te Pas, 2006). Visual judgments of

Corresponding Author: Yun-Xian Ho, New York University, Dept. of Psychology, 6 Washington Place, 8th Floor, New York, NY 10003, Tel: 1-212-998-7848, Email: yunxian.ho@nyu.edu.

roughness, glossiness, and color are not independent of viewing conditions or other surface properties (e.g., Billmeyer & O'Donnell, 1987; Ferwerda, Pellacini, & Greenberg, 2001; Fleming, Dror, & Adelson, 2003; Ho, Landy, & Maloney, 2006; Ho, Maloney, & Landy, 2007; Hunter & Harold, 1987; Pfund, 1930; Sève, 1993; Zaidi, 2001; although see Obein, Knoblauch, & Viénot, 2004).

Previous work in material perception has considered one perceived surface property at a time; however, most surfaces display several properties (e.g., the gloss and mesotexture of an orange, Figure 1). It would be useful if estimates of gloss were unaffected by surface mesostructure and vice versa. But, cues to one surface property (e.g., size and position of specular highlights) can affect visual judgments of another property (e.g., shape). For example, gloss can affect the perception of global shape by making surfaces appear more curved (Braje & Knill, 1994;Mingolla & Todd, 1986;Todd & Mingolla, 1983;Todd, Norman, Koenderink, & Kappers, 1997) and making convex surfaces appear concave (Blake & Bülthoff, 1990,1991; but also see Nefs, Koenderink, & Kappers, 2006). Similarly, it has been shown that shape can affect judgments of surface reflectance (Nishida & Shinya, 1998). Whether or not the presence of a specularity improves shape perception, the systematic patterns of distortions created by specular highlights on a surface provide a substantial amount of information about 3D surface curvature, and human observers use this information to derive 3D shape (Fleming, Torralba, & Adelson, 2004;Norman, Todd, & Orban, 2004).

Figure 2 illustrates how perceived gloss and mesotexture interact. Figure 2b has the same surfaces as Figure 2a with specular highlights greatly reduced. The surfaces look less glossy, as one might expect, but also less bumpy. This interaction is not unexpected given recent research. Techniques have been developed to successfully recover mesotexture from specularities (Chen, Goesele, & Seidel, 2006;Wang & Dana, 2006). This suggests that a surface's specular content can provide strong cues to local shape geometry, i.e., mesotexture. The effect of reducing specularities in Figure 2 suggests that two properties—gloss and mesotexture—interact in judgments of glossiness and bumpiness.

We have two goals in this study. First, we wish to estimate perceptual scales for two material properties, gloss and mesotexture. Second, we wish to model and understand their evident interaction. We employ a particular method, conjoint measurement, that allows us to achieve both goals with one experimental design (Krantz, Luce, Suppes, & Tversky, 1971, Chaps. 6,7; Luce & Tukey, 1964; Roberts, 1979, Chap. 5). We will demonstrate that an additive conjoint measurement model sufficiently describes the psychophysical mapping of each property to an internal scale of gloss and bumpiness. Although this is not the first study to examine the perceptual scaling of a material property (i.e., gloss, Ferwerda et al., 2001), this is the first study to simultaneously estimate the perceptual scaling for each of two material properties of a given surface and provide a simple model of their interaction.

Methods

Stimuli

Stimuli were 3D mesotextured surfaces positioned in a frontoparallel plane 70 cm in front of the observer. For convenience we will refer to the mesotexture in the surfaces we employ as "bumpiness." Each stimulus was assigned one of five possible bumpiness levels (mesotexture) and one of five gloss levels (specular reflectance) for a total of 25 possible surfaces. We denote the physical value of glossiness as g_i and bumpiness as b_j ($i, j = 1,2,\ldots, 5$).

We defined the stimuli using a Cartesian coordinate system whose *x*- and *y*-axes lay within the fronto-parallel plane of the stimulus. The *z*-axis was parallel to the observer's line of sight. Four hundred points in the stimulus plane forming a 20 × 20 cm square grid were jittered in the *x*- and *y*-directions by random values ranging over ±0.2 cm. Ellipsoids (possibly intersecting) were centered on each of these 400 points with principal axes parallel to the *x*-, *y*-, and *z*-axes. The radii in the *x*- and *y*-directions were 1 cm. For a surface texture with bumpiness level b^j, the *z*-radii of the ellipsoids were chosen randomly from the range $[0,b_j]$, where $b_j = (j + 1)^2/10$ cm.

The gloss value g_i was used to set two parameters associated with specular reflectance in the Ward reflectance model as implemented by the Radiance rendering software (Ward, 1994): the specular reflectance parameter ρ_s, which controls the proportion of incoming light reflected off of the surface at an angle close to the angle of incidence, and the microroughness parameter α, which controls the amount of blurring of the specular lobe. Gloss level g_1 corresponded to a matte (Lambertian) surface reflectance (i.e., $\rho_s = 0$, making α irrelevant). Gloss levels g_2 through g_5 corresponded to four logarithmically spaced levels ρ_s ranging from 0.007 to 0.053 and α ranging from 0.178 to 0.032. A combination of a high value of ρ_s and low value of α yielded a surface of high gloss and sharp highlights, while a low value of ρ_s and high value of α yielded a surface of low gloss with blurred highlights. These gloss levels were chosen such that each level of gloss was approximately equally discriminable from the next (see Fleming et al., 2003; Pellacini, Ferwerda, & Greenberg, 2000). The diffuse component (ρ_d) or surface albedo was fixed at red = 0.1, green = 0.2, and blue = 0.1, yielding a dark green surface color. All surfaces were rendered with interreflections (up to two ambient bounces) as well as occlusions and vignetting.

Each surface was rendered under a rectangular light source with dimensions 92 × 52 cm positioned above and to the left of the observer. These scene and object parameters provided the observer with several cues to gloss including the color and shape of the specular highlights. Each stimulus was rendered from the right and left eye's viewpoints, and viewed binocularly, providing a binocular disparity cue to depth. Four random surfaces were generated for each combination of gloss and bumpiness levels to minimize the chance of observers using idiosyncratic patterns in the distribution of ellipsoids to aid their judgments. Figure 3 shows a single stimulus stereo pair and the entire set of stimuli varying in physical gloss and bumpiness.

Apparatus

We presented the left and right images to the corresponding eyes of the observer on two 21-in. Dell LCD monitors placed to the observer's left and right and viewed through a mirror stereoscope. Lookup tables were used to correct the nonlinearities in the gun responses and to equalize the display values on the two monitors based on luminance measurements made with a Photo Research PR-650 spectrometer. The maximum luminance achievable on either screen was 114 cd/m^2. The stereoscope was contained in a box whose side measured 124 cm. The front face of the box was missing and that is where the observer sat in a chin/head rest. The interior of the box was coated with black flocked paper (Edmund Scientific, Tonawanda, NY) to absorb stray light. Only the stimuli on the screens of the monitors were visible to the observer. The casings of the monitors and any other features of the room were hidden behind the non-reflective walls of the enclosing box. Additional light baffles were placed near the observer's face to prevent light from the screens reaching the observer's eyes directly. The optical distance from each of the observer's eyes to the corresponding computer screen was 70 cm. The stimuli were rendered to be 70 cm in front of the observer to minimize any conflict between binocular disparity and accommodation. The observer's eyes were approximately in line with the center of the scene being viewed.

Software

The experimental software was written in the C programming language. We used the X Window System, Version 11R6 (Scheifler & Gettys, 1996), running under Red Hat Fedora Core 2 for graphical display. The computer was a Dell Optiplex GX 270 Workstation with a Matrox G450 graphics card. The rendered stereo image pair was represented by floating point RGB triplets for each pixel of the image. These triplets were the relative luminance values of each pixel. We translated the output relative luminance values to 24-bit graphics codes, correcting for nonlinearities in the monitors' responses by means of measured lookup tables for each monitor.

Procedure

Two sets of observers participated, the first judging bumpiness (Experiment 1) and the second judging glossiness (Experiment 2). All observers first participated in a screening test that consisted of two blocked conditions in which they were required to make judgments of bumpiness or glossiness within each given gloss or bumpiness level, respectively. We performed the screening test to ensure that observers could order the stimuli in bumpiness for fixed gloss and vice versa. In doing so, we tested one of the necessary conditions for an additive conjoint representation, monotonicity (Krantz et al., 1971, p. 249).

On each screening trial, observers viewed two surfaces in succession and judged which appeared bumpier or glossier, depending on the condition. In the condition in which observers judged bumpiness, only the comparisons between stimuli with the same gloss level were tested, resulting in a total of 50 trials (one trial per comparison). Likewise, glossiness judgments were only made for pairs of stimuli having the same level of bumpiness.

On each trial of the main experiment, observers viewed one of the 325 possible pairs (including self-comparisons) of the 25 surfaces in Figure 1. Like the screening test, the observer's task was to judge which of the two surfaces on each trial appeared bumpier (Experiment 1) or glossier (Experiment 2). Each pair was presented three[1] times.

The trial sequence proceeded as follows: first, a central fixation point was presented for 200 ms. Next, the first surface was presented for 400 ms followed by a 200 ms interstimulus interval (blank frame). Then, the second surface was presented for 400 ms. The observer indicated by key press whether the first or second surface appeared to be bumpier (or glossier). The next trial was initiated immediately after the response.

Observers

A total of 12 observers participated in this study (six each in Experiments 1 and 2). One additional observer failed to pass the screening test (i.e., obtained less than 90% correct responses) and was excluded from the main study. All observers were unaware of the purpose of the study and had normal or corrected-to-normal vision.

An additive conjoint measurement model of material perception

To determine whether observers could ignore cues to gloss when making judgments of bumpiness and vice versa, we fit an additive conjoint model to our data. For the model, it was assumed that the physical gloss level g_i and bumpiness level b_j of surface s_{ij} separately and additively contribute to perceived bumpiness and gloss. We modeled perceived

[1]This value was calculated to be the smallest number of repetitions needed to produce reliable parameter estimates for gloss and bumpiness based on simulations of the additive conjoint model.

bumpiness B^A_{ij} for the additive model as the sum of contributions ("cues") to bumpiness from physical bumpiness $B^b(b_j)$ and physical gloss $B^g(g_i)$,

$$B^A_{ij}=B^g(g_i)+B^b(b_j)=B^g_i+B^b_j. \quad (1)$$

Perceived gloss was modeled similarly,

$$G^A_{ij}=G^g(g_i)+G^b(b_j)=G^g_i+G^b_j. \quad (2)$$

Note that this is not the typical weighted linear cue combination model commonly discussed in the literature (e.g., Landy, Maloney, Johnston, & Young, 1995). We assume only that cues combine additively after we scale them by functions $B^g(.)$, $B^b(.)$ and $G^g(.)$, $G^b(.)$.

As written, the two equations model interactions by additive "contamination" of the estimate of one property by cues to the other. If the two surface properties do not interact, then B^g_i and G^b_j should equal zero for all *i* and *j*. When analyzing the data, we also test this simple additive model against a model that allows more complex, non-additive interactions between the surface properties.

In comparing the bumpiness of surfaces S_{ij} and S_{kl}, we assume that the observer forms the noise-contaminated decision variable,

$$\Delta=B^A_{ij}-B^A_{kl}+\varepsilon,\ \varepsilon\sim\text{Gaussian}(0,\sigma^2) \quad (3)$$

and judges surgace S_{ij} as bumpier precisely when $\Delta > > 0$. The parameter σ represents the observer's precision in judgment.

If we simultaneously scaled all of the values of B^A_{ij} and σ by a positive constant, or added a constant to all of the values of B^A_{ij}, the predictions of the model would not be affected. For convenience, we anchor the scales by setting $B^b_1=B^g_1=0$ and scale them so that σ = 1. We then estimate the remaining 8 free parameters $B^g_2,\ldots,B^g_5$ and $B^b_2,\ldots,B^b_5$ using maximum likelihood estimation (Mood, Graybill, & Boes, 1974). We fit a similar model of comparisons of gloss.

Our model makes no assumption about the direction of the cue effect, i.e., whether gloss increases or decreases bumpiness; estimated parameters can be positive or negative. Indeed, in a departure from the ordinary additive conjoint model (Krantz et al., 1971, Chap. 6) we do not force $B^g(.)$, $B^b(.)$ and. $G^g(.)$, $G^b(.)$ to be monotonic functions.

Results

Judgments of bumpiness and gloss from two typical observers for the 25 stimuli used in Experiments 1 and 2, respectively, are shown in the left panels of Figures 4a and 4b. The right panels show predicted results of an ideal observer with judgments of bumpiness and gloss uncontaminated by cues from the task-irrelevant property. Notice that although observers' performance is fairly close to ideal, there are obvious deviations. We estimated the perceived bumpiness B^A_{ij} and gloss G^A_{ij} parameters by maximum likelihood fit of the

additive model to determine the relationship (if any) between gloss and mesotexture. Figure 4c shows the averaged parameter estimates for all observers in Experiment 1 (top panel) and 2 (bottom panel). Parameter estimates for the observers shown in 4a and 4b are indicated by open circles. Note that the parameter estimates were first normalized to the maximum value of B_j^b (or G_i^g) for each observer to best illustrate the relative magnitude of cue contamination. The monotonically increasing form of $B^b(.)$ shows that perceived bumpiness increased with bumpiness level. Likewise, perceived gloss increased with gloss level. This is not surprising given that observers were screened in advance to ensure that their ordering of bumpiness and gloss was close to veridical. What *is* surprising is that the functions of perceived bumpiness and gloss were strikingly similar across all observers in each respective experiment (data not shown, except indirectly in Figure 5) suggesting that observers used one common perceptual scale for bumpiness and one for gloss in each experiment. The function $B^b(.)$ increases approximately linearly with increasing bumpiness level and $G^g(.)$ is sigmoidal in form. Thus, the internal scaling of bumpiness and glossiness can be described by fairly simple transformations of the corresponding physical properties.

Were observers' judgments of bumpiness and gloss contaminated by cues to the irrelevant property? Clearly, all parameter estimates B_j^g and G_i^b were greater than zero for $i, j = 2,\ldots,5$ for the two typical observers highlighted in Figure 4c. We tested whether there was significant contamination of bumpiness judgments by changes in gloss and vice versa for all observers using a nested hypothesis test (Mood et al., 1974, p. 440 ff.)..

To do this, we fit an *independent-property model* in which these task-irrelevant contributions (B_i^g and G_j^b) were fixed at zero. Thus, for perceived bumpiness, the independent-property model has only four parameters (B_2^b, B_3^b, B_4^b, and B_5^b) and likewise for the gloss model. The fits of these independent-property models were compared to the additive-model fits by the same likelihood-ratio test. The independent-property model was rejected for four out of six observers in both Experiments 1 and 2 at the Bonferroni-corrected level ($\chi^2 \geq 0.089$, $p < 0.008$). In other words, most observers could not ignore cues to gloss in making judgments of bumpiness and, similarly, could not ignore cues to bumpiness in making judgments of glossiness. An increase of gloss increased perceived bumpiness on average for all observers by as much as 11% of the range of bumpiness levels we used. Likewise, glossiness judgments for surfaces with higher bumpiness level increased by as much as 27% of the range of gloss levels used here.

We plot the same curves shown in Figure 4c in a form commonly used in analysis of variance in Figures 5a and 5b (top rows, solid lines) for all observers in Experiments 1 and 2, respectively. The solid lines correspond to the fit of the additive model. The additive model forces these level contours to be parallel. The clean separation of the contours confirms that perceived bumpiness increased with physical bumpiness and perceived gloss increased with physical glossiness for all observers (identical to the upper curves in Figure 4c). Almost all observers perceived an increase in bumpiness with increasing gloss level as shown by the slight upward trend in almost all of the bumpiness level contours (Figure 5a). Similarly, most observers could not ignore bumpiness when making judgments of glossiness (Figure 5b), although in this case the trend was less clearly monotonic again confirming the trend observed in Figure 4c.

We next tested whether the interaction between bumpiness and gloss could be sufficiently described by the simple additive "contamination" model defined in Eq. 1 and Eq. 2. We compared the fit of the additive model to that of a *full model* that allows nonlinear

interactions between the two cues. In the full model, we model the perceived bumpiness B_{ij}^F as an unconstrained value

$$B_{ij}^F = B^F(g_i, b_j), \tag{4}$$

and we model perceived gloss similarly,

$$G_{ij}^F = G^F(g_i, b_j). \tag{5}$$

As before, we assume that, in comparing the perceived bumpiness of surfaces S_{ij} and S_{kl}, the observer forms the noise-contaminated decision variable

$$\Delta = B_{ij}^A - B_{kl}^A + \varepsilon,\ \varepsilon \sim \text{Gaussian}(0, \sigma^2) \tag{6}$$

and judges surface S_{ij} as bumpier precisely when $\Delta > 0$. Again, without loss of generality, we anchor the scale by setting $B_{11}^F = 0$ and set $\sigma = 1$. We then estimate the remaining 24 values B_{ij}^F. We fit the glossiness judgments analogously.

The full model parameter estimates of B_{ij}^F and G_{ij}^F are plotted as dashed lines in Figures 5a and 5b, respectively. We performed a nested hypothesis test (Mood et al., 1974, p. 440 ff.) to determine whether the full model resulted in a significantly better fit to the choice data than the more constrained additive model. The nested hypothesis test revealed that the additive model performed just as well as the full model for five out of the six observers in Experiment 1 and three out of six of the observers in Experiment 2 at the Bonferroni-corrected level ($\chi^2 \geq 32.61$, $p < 0.008$). Table 1 reports the log-likelihood and corresponding p values. We also regressed the additive model predictions against the full model predictions. R^2 values ranged from 0.97 to 0.99 (median 0.98) for the six observers in Experiment 1 and 0.83 to 0.99 (median 0.96) for Experiment 2 suggesting that the additive and full model predictions were similar. Finally, we computed the residual differences between the two models for both judgments (i.e., $B_{ij}^F - B_{ij}^A$ and $G_{ij}^F - G_{ij}^A$), and normalized the residuals for each subject to have mean 0. Gray-scale plots of the normalized residuals are shown in the bottom rows of Figures 5a and 5b. The residual values are small and show no obvious common pattern. Thus, we conclude that the additive model is adequate to model the data.

Discussion

In this study, we used conjoint measurement to derive scales for the perceptual correlates of two surface material properties, gloss and bumpiness. An ideal observer judging glossiness should ignore variations in surface texture, and vice versa. In contrast, human observers perceived physically glossier surfaces to be bumpier and physically bumpier surfaces to be glossier, suggesting that cues to mesoscale and microscale properties are inextricably linked. In terms of the estimated perceptual scales, we found that we could model the interaction of the two properties as a simple additive “contamination” of each by the other. The degree of contamination, while significant, is small on average for all observers: about 11% for contamination of bumpiness by gloss and 27% for contamination of gloss by bumpiness, relative to the range produced by the relevant cue.

What might explain these interactions? The bas-relief ambiguity (Belhumeur, Kriegman, & Yuille, 1999) may have contributed to the observed misestimates of bumpiness. But, the availability of binocular disparity should have helped observers to overcome this ambiguity as well as to maintain gloss constancy across variations in mesotexture (Blake & Bülthoff, 1990, 1991; Norman et al., 2004). As for misestimates of glossiness, previous studies on the perception of gloss suggest that the changes in cues to gloss like the shape, size, and distribution of specular highlights affect judgments of glossiness (e.g., Beck & Prazdny, 1981; Berzhanskaya, Swaminathan, Beck, & Mingolla, 2005). Thus, it is possible that observers' judgments of the glossiness of surfaces used in this study were affected by local changes on each surface bump produced by changes in physical bumpiness.

There is a growing body of evidence that the human visual system makes use of simple, image-based statistics to evaluate material properties. Nishida & Shinya (1998) found visual estimates of surface reflectance can be modeled by a luminance histogram matching algorithm. Similarly, Fleming, Dror, & Adelson (2003) found that although highlights alone may be used to distinguish between matte and glossy surfaces, perceived gloss is a function of the statistics of the responses of band pass spatial filters to the image. Recently, Motoyoshi, Nishida, Sharan, & Adelson (2007) found that visual estimates of gloss correlate well with the skew of the distribution of luminance values and that changing the skewness of the distribution affected perception of both lightness and glossiness of the surfaces. Similarly, Ho et al. (2006; 2007) found that human observers' estimates of surface roughness are affected by image statistics such as the proportion of pixels in shadow.

Judgments of glossiness, bumpiness and global shape all provide information about the structural properties of the orange in Figure 1 but at very different spatial scales. The conjoint measurement procedure we employed here allowed us to simultaneously estimate the mapping between information at two of these scales and assess how information at one scale affected perception at the other. We found that both the derived perceptual scales and the interaction could be modeled in a remarkably simple form. Conjoint measurement is one type of scaling procedure that is potentially a powerful tool for analyzing surface material perception and, more generally, the perception of more complex visual scenes.

Acknowledgments

This research was supported by National Institutes of Health Grants EY16165 and EY08266. We thank Victoria Sconzo for help with this study.

References

Beck J, Prazdny K. Highlights and the perception of glossiness. Perception & Psychophysics. 1981; 30:407–410. [PubMed: 7322822]

Belhumeur PN, Kriegman D, Yuille A. The bas-relief ambiguity. International Journal of Computer Vision. 1999; 35:33–44.

Berzhanskaya J, Swaminathan G, Beck J, Mingolla E. Remote effects of highlights on gloss perception. Perception. 2005; 34:565–575. [PubMed: 15991693]

Billmeyer FW, O'Donnell FXD. Visual gloss scaling and multidimensional scaling analysis of painted specimens. Color Research Applications. 1987; 12:315–326.

Blake A, Bülthoff H. Does the brain know the physics of specular reflection? Nature. 1990; 343:165–168. [PubMed: 2296307]

Blake A, Bülthoff H. Shape from specularities: computation and psychophysics. Philos Trans R Soc Lond B Biol Sci. 1991; 331:237–252. [PubMed: 1674154]

Braje WL, Knill DC. Apparent surface shape affects perceived specular reflectance of curved surfaces. Investigative Ophthalmalogy and Visual Science Supplement. 1994; 35:1628.

Chen, T.; Goesele, M.; Seidel, H-P. Mesostructure from specularity. Proceedings of IEEE, Computer Vision and Pattern Recognition; New York, NY. 2006. p. 1825-1832.

Ferwerda JA, Pellacini F, Greenberg DP. A psychophysically-based model of surface gloss perception. Proceedings SPIE Human Vision and Electronic Imaging. 2001:291–301.

Fleming RW, Dror RO, Adelson EH. Real-world illumination and the perception of surface reflectance properties. Journal of Vision. 2003; 3:347–368. [PubMed: 12875632]

Fleming RW, Torralba A, Adelson EH. Specular reflections and the perception of shape. Journal of Vision. 2004; 4:798–820. [PubMed: 15493971]

Ho YX, Landy MS, Maloney LT. How direction of illumination affects visually perceived surface roughness. Journal of Vision. 2006; 6:634–648. [PubMed: 16881794]

Ho YX, Maloney LT, Landy MS. The effect of viewpoint on perceived visual roughness. Journal of Vision. 2007; 1:1–16. [PubMed: 17461669]

Hunter, RS.; Harold, RW. The Measurement of Appearance. 2nd ed.. New York: Wiley; 1987.

Koenderink JJ, Van Doorn A. Illuminance texture due to surface mesostructure. Journal of the Optical Society of America A. 1996; 13:452–463.

Krantz, DH.; Luce, RD.; Suppes, P.; Tversky, A. Foundations of Measurement, Vol. 1: Additive And Polynomial Representations. New York: Academic Press; 1971.

Landy MS, Maloney LT, Johnston EB, Young M. Measurement and modeling of the depth cue combination: In defense of weak fusion. Vision Research. 1995; 35:389–412. [PubMed: 7892735]

Luce RD, Tukey JW. Simultaneous conjoint measurement. Journal of Mathematical Psychology. 1964; 1:1–27.

Mingolla E, Todd JT. Perception of solid shape from shading. Biological Cybernetics. 1986; 53:137–151. [PubMed: 3947683]

Mood, AM.; Graybill, FA.; Boes, DC. Introduction to the Theory of Statistics. 3rd ed.. New York: McGraw-Hill; 1974.

Motoyoshi I, Nishida S, Sharan L, Adelson EH. Image statistics and the perception of surface qualities. Nature. 2007; 446 in press.

Nefs HT, Koenderink JJ, Kappers AML. Shape-from-shading for matte and glossy objects. Acta Psychologica. 2006; 121:297–316. [PubMed: 16181604]

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

Norman JF, Todd JT, Orban GA. Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. Psychological Science. 2004; 15:565–570. [PubMed: 15271003]

Obein G, Knoblauch K, Viénot F. Difference scaling of gloss: Nonlinearity, binocularity, and constancy. Journal of Vision. 2004; 4:711–720. [PubMed: 15493965]

Pellacini F, Ferwerda JA, Greenberg DP. Toward a psychophysically-based light reflection model for image synthesis. Computer Graphics. 2000; 34:55–64.

Pfund AH. The measurement of gloss. Journal of the Optical Society of America A. 1930; 20:23–26.

Pont SC, te Pas SF. Material-illumination ambiguities and the perception of solid objects. Perception. 2006; 35:1331–1350. [PubMed: 17214380]

Roberts, FS. Measurement Theory. In: Rota, G-C., editor. Encyclopedia of Mathematics and Its Applications. Vol. Vol. 7. Reading, MA: Addison-Wesley; 1979.

Scheifler, RW.; Gettys, J. X Window System: Core Library and Standards. Boston: Digital Press; 1996.

Sève R. Problems connected with the concept of gloss. Color Research & Application. 1993; 18:241–252.

Todd JT, Mingolla E. Perception of surface curvature and direction of illumination from patterns of shading. Journal of Experimental Psychology: Human Perception & Performance. 1983; 9:583–595. [PubMed: 6224894]

Todd JT, Norman JF, Koenderink JJ, Kappers AM. Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. Perception. 1997; 26:807–822. [PubMed: 9509135]

Wang J, Dana KJ. Relief texture from specularities. IEEE Transactions on pattern analysis and machine intelligence. 2006; 28:446–457. [PubMed: 16526429]

Ward GJ. The RADIANCE lighting simulation and rendering system. Computer Graphics. 1994; 28:459–472.

Zaidi Q. Color constancy in a rough world. Color Research & Application. 2001; 26:S192–S200.

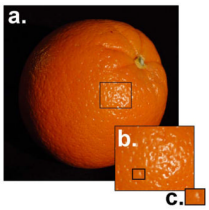

Figure 1.
An example of a typical object (an orange) that has (a) megascale, (b) mesoscale, and (c) microscale properties.

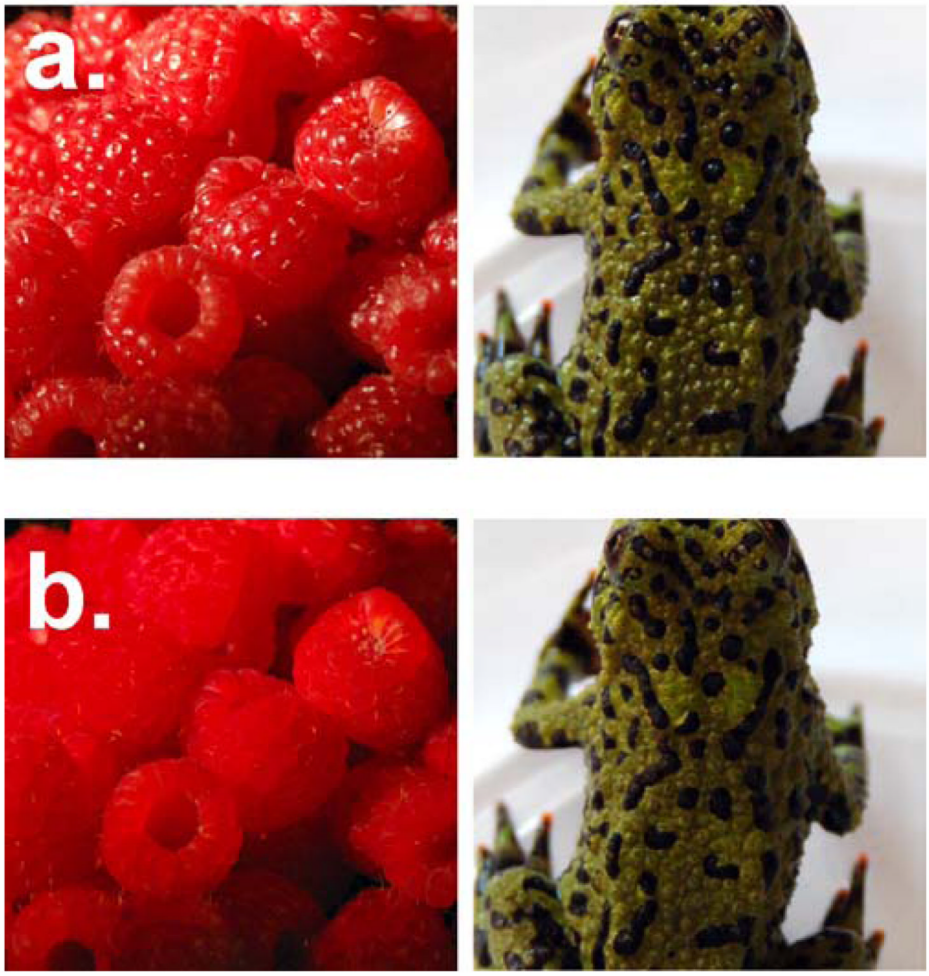

Figure 2.
Examples of real world mesoscale texture (a) with and (b) without specular highlights. Images of the raspberries without specular highlights were created by using a polarizing filter and the image of the toad was created by digitally removing highlights using the Adobe Photoshop CS™ software.

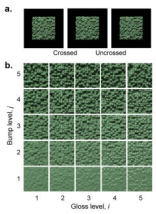

Figure 3.

(a) Example stimulus stereo pair. Stimuli were green surface patches composed of 20 × 20 intersecting ellipsoids. This example has physical bumpiness level $i = 3$ and gloss level $j = 2$. The left and right image pairs are for crossed and uncrossed viewing, respectively. (b) One representative set of stimuli showing all combinations of gloss and bumpiness level.

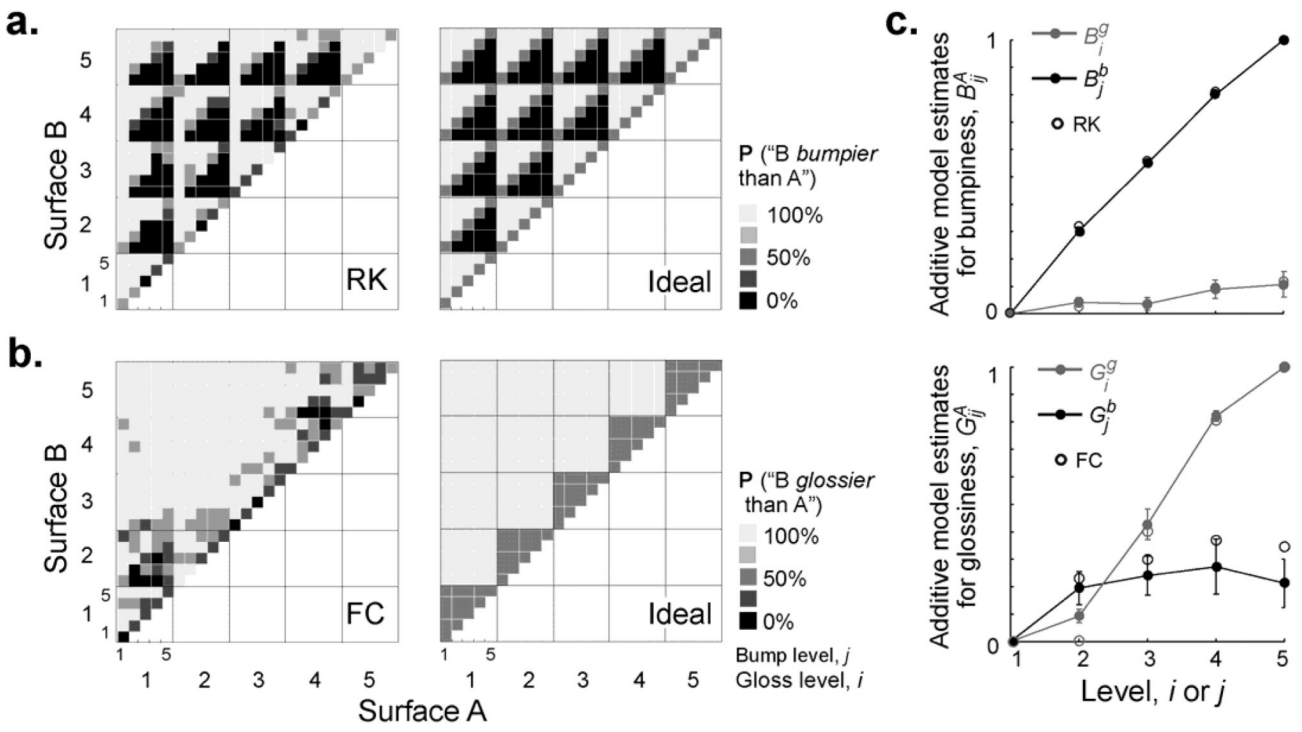

Figure 4.

Results for Experiments 1 & 2. (a) Bumpiness judgments made by one typical observer RK in Experiment 1 (left panel) and predicted judgments of bumpiness uncontaminated by changes in physical gloss for the ideal observer (right panel). The gray level of each square in the matrix represents the proportion of time that Surface B was perceived bumpier than Surface A for each pairwise comparison. (b) Glossiness judgments made by one typical observer FC in Experiment 2 (left panel) and predicted judgments of glossiness uncontaminated by changes in physical bumpiness for the ideal observer (right panel). (c) Averaged normalized additive model parameter estimates for all observers are shown for Experiment 1 (top panel) and 2 (bottom panel). Parameter estimates for bumpiness or glossiness judgments as a function of gloss level—B_i^g or G_i^g, respectively—are indicated in gray and as a function of bumpiness level—B_j^b or G_j^b, respectively—are indicated in black. Parameter estimates for observers RK and FC are indicated by the open circles in each respective plot. Standard errors across observers are plotted for each experiment (N=6).

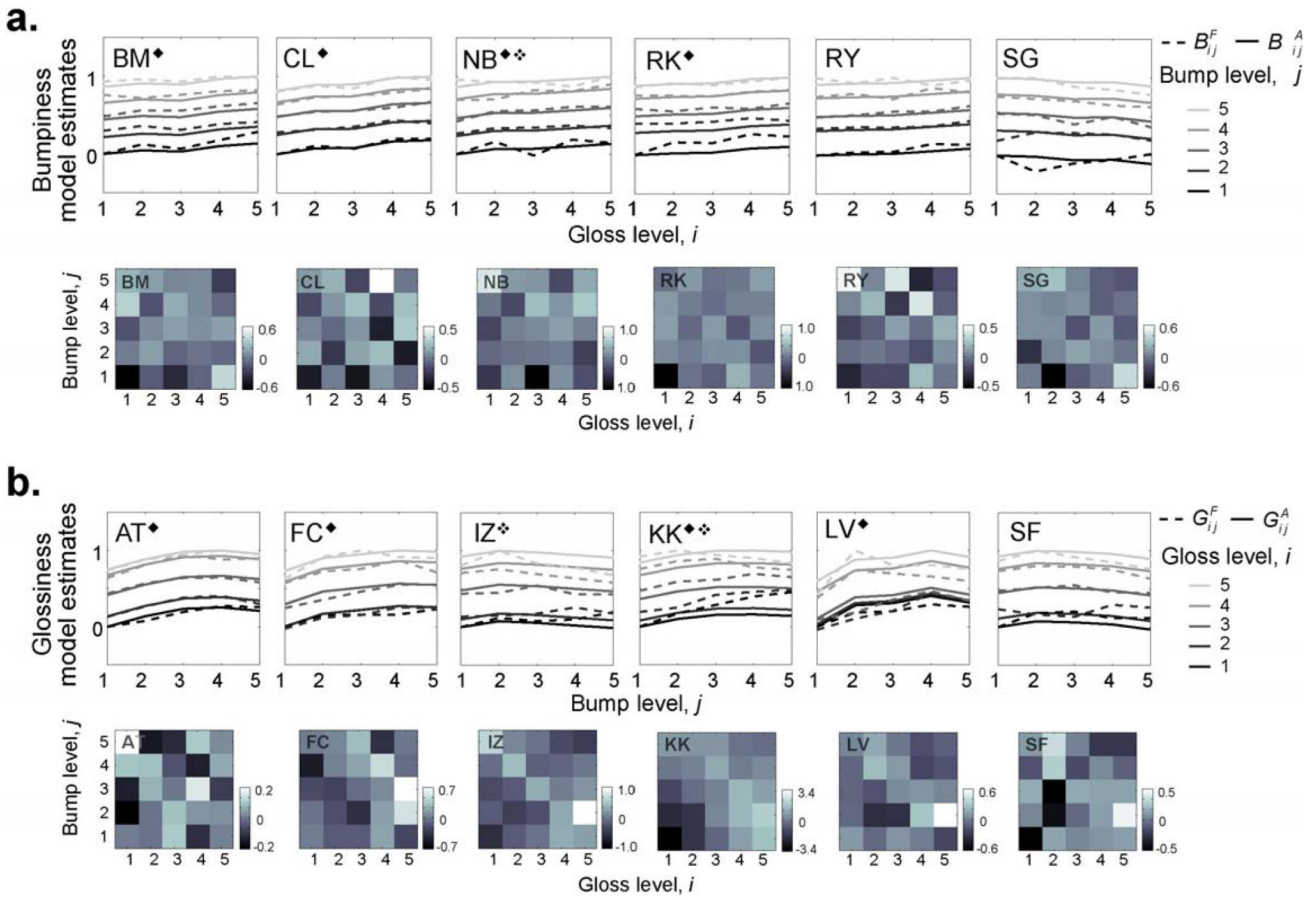

Figure 5.

Full and additive model predictions and residual differences for all observers in Experiments 1 & 2. (a) Estimates of perceived bumpiness from the fits of the full (B^F_{ij}, dashed curves) and additive (B^A_{ij}, solid curves) models are plotted as a function of gloss level, with bumpiness level as the parameter. Corresponding residual differences $B^A_{ij} - B^F_{ij}$ are shown as grayscale plots in the bottom row. (b) Similar plots of perceived gloss estimates from the full (G^F_{ij}, dashed curves) and additive (G^A_{ij}, solid curves) models as a function of bumpiness level with gloss level as the parameter. Again, corresponding residual differences $G^A_{ij} - G^F_{ij}$ are shown in the bottom row. Significant differences between the additive and independent-property model fits are indicated with the ♦ symbol. Significant differences between the full and additive model fits are indicated with the ❖ symbol. (Note: Residuals were forced to have mean 0; an additive shift of scale values leaves the predictions of both models unchanged.)

Table 1

Model comparisons. Log-likelihood values for various models and *p*-values for nested hypothesis tests for selected model comparisons. (**Bold** indicates *p*-values that were significant at the Bonferroni-corrected alpha level, 0.008.)

Model/Test	**No. of Parameters**	**Observers**					
Experiment 1.		**BM**	**CL**	**NB**	**RK**	**RY**	**SG**
I. Full Model	24	−241.95	−182.05	−236.88	−230.77	−258.31	−370.78
II. Additive	8	−248.857	−188.56	−254.30	−238.24	−267.17	−380.73
Model							
I vs. II	--	0.612	0.671	**0.004**	0.528	0.340	0.225
III. Independent-Property Model	4	−267.11	−239.92	−267.75	−250.06	−273.67	−385.27
II vs. III	--	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	0.011	0.059
Experiment 2.		AT	FC	IZ	KK	LV	SF
I. Full Model	24	−337.27	−259.48	−277.45	−177.17	−479.28	−322.36
II. Additive	8	−340.71	−273.73	−302.84	−250.38	−495.82	−337.49
Model							
I vs. II	--	0.976	0.028	**< 0.0001**	**< 0.0001**	**0.007**	0.017
III. Independent-Property Model	4	−379.09	−335.90	−308.10	−278.37	−544.04	−343.08
II vs. III	--	**< 0.0001**	**< 0.0001**	0.033	**< 0.0001**	**< 0.0001**	0.024