Research Article

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 and [2]Center for Neural Science, New York University

ABSTRACT—*The image of a material's surface varies not only with viewing and illumination conditions, but also with the material's surface properties, including its 3-D texture and specularity. Previous studies on the visual perception of surface material have typically focused on single material properties, ignoring possible interactions. In this study, we used a conjoint-measurement design to determine how observers represent perceived 3-D 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 quantified how changes in each surface-material property affected judgments of the other and found that a simple additive model captured visual perception of texture and specularity and their interaction. Conjoint measurement is potentially a powerful tool for analyzing perception of surface material in realistic environments.*

Surfaces can be analyzed at various spatial scales. Koenderink and Van Doorn (1996) delineated three distinct scale ranges in the perception of surface properties: *megascale*, *mesoscale*, and *microscale*. These scales can be understood with reference to the example of an orange (see Fig. 1). A megascale description of the orange refers to its global shape, which is spherical. Mesoscale refers to the properties of the orange skin, that is, the irregular and bumpy surface (mesotexture) of the orange, which is similar to the surface of a lemon, but dramatically different from the surface of a smooth rubber ball. Variations at the microscale level include changes in the microscopic surface structure of the orange skin that result in its glossy appearance. To visually identify an orange, one takes 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 the perception of material is that light interacts with surfaces in a complex way. As a consequence, it may be difficult for the 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 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 the perception of material has considered one perceived surface property at a time; however, most surfaces display several properties (e.g., the gloss and mesotexture of an orange, illustrated in Fig. 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 curved surfaces appear more curved (Braje & Knill, 1994; Mingolla & Todd, 1986; Todd & Mingolla, 1983; Todd, Norman, Koenderink, & Kappers, 1997) and by making convex surfaces appear concave (Blake & Bülthoff, 1990, 1991; but also see Nefs, Koenderink, & Kappers, 2006). Systematic patterns of distortions created by specular highlights on a surface provide a substantial amount of information about 3-D surface curvature, and human observers use this information to derive 3-D shape (Fleming, Torralba, & Adelson, 2004; Norman, Todd, & Orban, 2004). Similarly, it has been shown that shape can affect judgments of surface reflectance (Nishida & Shinya, 1998). Changes in curvature of glossy

Address correspondence to Yun-Xian Ho, New York University, Department of Psychology, 6 Washington Place, 8th Floor, New York, NY 10003, e-mail: yunxian.ho@nyu.edu.

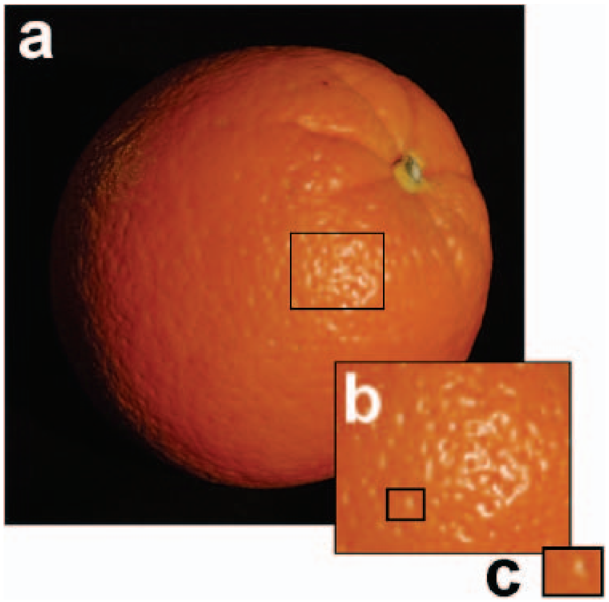

Fig. 1. An example of a typical object (an orange) that has (a) megascale, (b) mesoscale, and (c) microscale properties.

surfaces produce changes in shape, size, and distribution of specular highlights, and such changes, in turn, can affect judgments of glossiness (e.g., Beck & Prazdny, 1981; Berzhanskaya, Swaminathan, Beck, & Mingolla, 2005).

Figure 2 illustrates how perceived gloss and mesotexture interact. Figure 2b shows the same surfaces as Figure 2a with specular highlights removed. 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). Thus, a surface's specular content can provide strong visual cues that humans use to estimate local shape geometry (i.e., mesotexture). The effect of removing specularities in Figure 2 suggests that two properties—gloss and mesotexture—interact in judgments of glossiness and bumpiness.

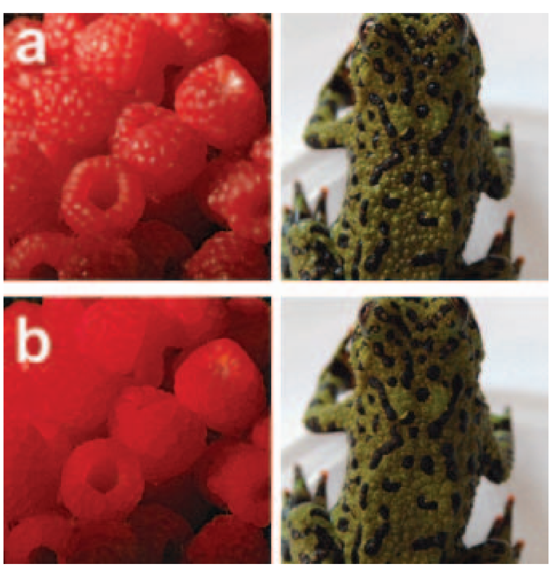

Fig. 2. Examples of real-world mesoscale texture (a) with and (b) without specular highlights. The image of the raspberries in (b) was created by using a polarizing filter, and the image of the toad in (b) was created by digitally removing highlights using Adobe Photoshop CS™ software.

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

METHOD

Stimuli

The stimuli were 3-D mesotextured surfaces positioned in a frontoparallel plane 70 cm in front of the observer. For convenience, we refer to the mesotexture in these surfaces as "bumpiness." Each stimulus was assigned one of five possible bump levels (mesotexture) and one of five gloss levels (specular reflectance), for a total of 25 possible surfaces. We denote the physical values 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 frontoparallel plane of the stimulus. The z-axis was parallel to the observer's line of sight. Four hundred points forming a 20-cm × 20-cm square grid in the stimulus plane were jittered in the x- and y-directions by random values ranging over ± 0.2 cm. An ellipsoid with principal axes parallel to the x-, y-, and z-axes was centered on each of these 400 points, and neighboring ellipsoids were allowed to intersect. The radii in the x and y directions were 1 cm. For a surface texture with bump 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 Ward's (1994) reflectance model as implemented by the Radiance rendering software: the specular reflectance parameter, ρ_s, which controls the proportion of incoming light reflected off 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, with ρ_s ranging from .007 up to .053 and α ranging from .178 down to .032. A combination of a high value of ρ_s and a low value of α yielded a surface of high gloss and sharp highlights, whereas a low value of ρ_s and a 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 = .1, green = .2, and blue = .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 that measured 92 cm × 52 cm and was 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 eyes' viewpoints, and viewed binocularly, so as to provide a binocular-disparity cue to depth. Four random surfaces were generated for each combination of gloss and bump 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 a representative set of stimuli showing all combinations of physical gloss and bumpiness.

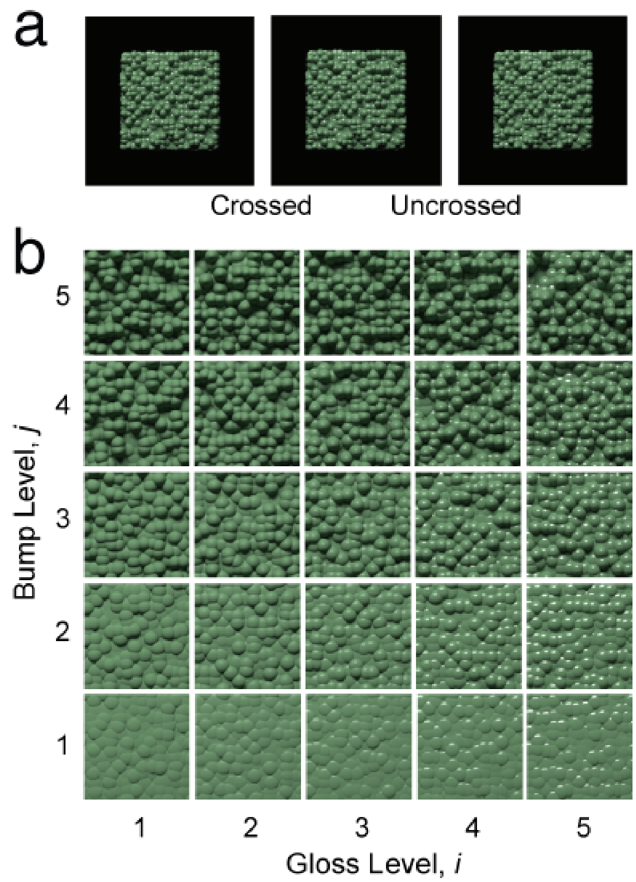

Fig. 3. Example of a stimulus stereo pair (a) and a representative set of stimuli showing all combinations of gloss level and bump level (b). The stimuli were green surface patches composed of intersecting ellipsoids in a 20 × 20 grid. The example in (a) has physical gloss level $i = 2$ and bump level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

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 on the basis of luminance measurements made with a Photo Research PR-650 spectrometer (Chatsworth, CA). 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 a chin-head rest was positioned there. 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 nonreflective walls of the box. Additional light baffles were placed near the observer's face so that light from the screens could not reach 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 (red, green, blue) triplets, one for each pixel of the image. These triplets were the relative luminance values of the pixels. We translated the output relative luminance values to 24-bit graphics codes, correcting for nonlinearities in the monitors' responses by means of a measured lookup table 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 level of gloss or bumpiness, respectively. We performed the screening test to ensure that observers could order the stimuli in bumpiness when gloss was fixed 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 (1 trial per comparison). Likewise, glossiness judgments were made only for pairs of stimuli having the same level of bumpiness.

On each trial of the main experiment, observers viewed 1 of the 325 possible pairs (including self-comparisons) of the 25 types of surfaces illustrated in Figure 3b. As in the screening test, the observer's task was to judge which of the two surfaces appeared bumpier (Experiment 1) or glossier (Experiment 2). Each pair was presented three[1] times.

The sequence on each trial was 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 (6 each in Experiments 1 and 2). One additional observer failed to pass the screening test (i.e., responded correctly on fewer than 90% of the trials) 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 bump level b_j of surface S_{ij} separately and additively contribute to perceived bumpiness and gloss. Perceived bumpiness was modeled in an additive model[2] $B_{ij}{}^{A}$ as the sum of contributions ("cues") to bumpiness from physical bumpiness $B^b(b_j)$ and physical gloss $B^g(g_i)$:

$$B_{ij}{}^{A} = B^g(g_i) + B^b(b_j) = B_i{}^g + B_j{}^b. \qquad (1)$$

Perceived gloss was modeled similarly:

$$G_{ij}{}^{A} = G^g(g_i) + G^b(b_j) = G_i{}^g + G_j{}^b. \qquad (2)$$

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 assumed only that cues combine additively after we scale them by functions $B^g(\,.\,)$, $B^b(\,.\,)$, and $G^g(\,.\,)$, $G^b(\,.\,)$.

As written, Equations 1 and 2 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 $B_i{}^g$ and $G_j{}^b$ should equal zero for all i and j. When analyzing the data, we tested this simple additive model against a model that allowed more complex, nonadditive 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

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

and judges surface S_{ij} as bumpier precisely for values of Δ greater than 0. The parameter σ represents the observer's precision in judgment.

If we simultaneously scaled all the values of $B_{ij}{}^{A}$ and σ by a positive constant, or added a constant to all of the values of $B_{ij}{}^{A}$, the predictions of the model would not be affected. For convenience, we anchored the scales by setting $B_1{}^b = B_1{}^g = 0$ and scaled them so that $\sigma = 1$. We then estimated the remaining eight free parameters $B_2{}^g, \ldots, B_5{}^g$ and $B_2{}^b, \ldots, B_5{}^b$ 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's effect, that is, 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 glossiness from 2 typical observers in Experiments 1 and 2, respectively, are shown in Figures 4a and 4b, along with predicted results of an ideal observer whose judgments are uncontaminated by cues from the task-irrelevant property. Although observers' performance is fairly close to ideal, there are obvious deviations. We estimated the perceived-bumpiness and perceived-gloss parameters, $B_{ij}{}^{A}$ and $G_{ij}{}^{A}$, 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 Experiments 1 and 2. Note that the parameter estimates were

[1]This value was determined to be the smallest number of repetitions needed to produce reliable parameter estimates for gloss and bumpiness (on the basis of simulations of the additive conjoint-measurement model).

[2]Note that we use superscript "A" to refer to our additive models and superscript "F" to refer to the full model, which we discuss in the Results section.

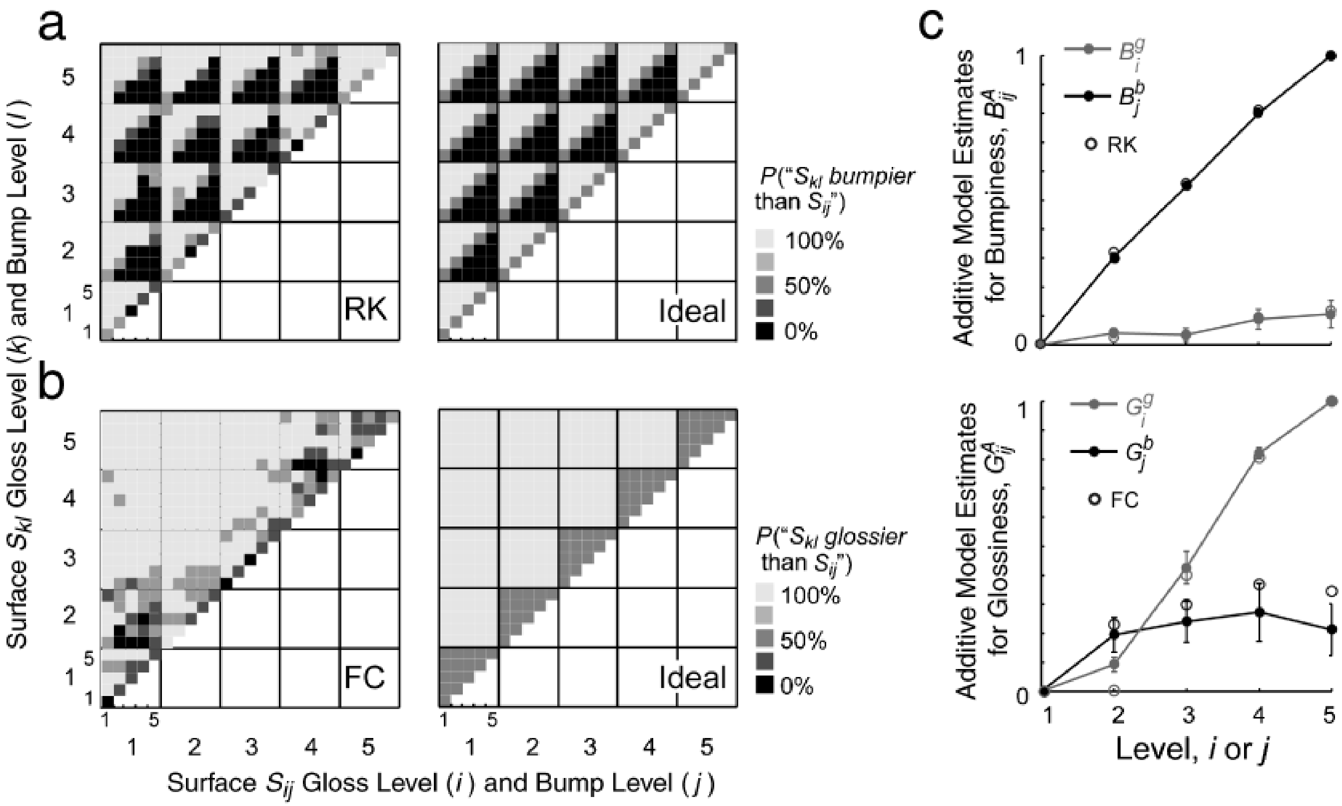

Fig. 4. Results for Experiments 1 and 2. In (a), bumpiness judgments made by 1 typical observer (R.K.) in Experiment 1 are shown along with predicted judgments of bumpiness uncontaminated by changes in physical gloss (i.e., judgments of the ideal observer). In (b), glossiness judgments made by 1 typical observer (F.C.) in Experiment 2 are shown along with predicted judgments of glossiness uncontaminated by changes in physical bumpiness (i.e., judgments of the ideal observer). The gray levels of the squares in the matrices represent the proportion of time that a surface S_{kl} was perceived to be (a) bumpier or (b) glossier than another surface S_{ij}, for each pair-wise comparison. Gloss level i (or k) is indicated by the large numerical labels (1, 2, . . . , 5), and bump level j (or l) is indicated by the small numerical labels (1, 2, . . . , 5). The graphs in (c) present the normalized parameter estimates for the additive model, averaged across all observers in Experiments 1 (top) and 2 (bottom). Parameter estimates for bumpiness and glossiness judgments as a function of gloss level—B_i^g and G_i^g, respectively—are indicated in gray, and estimates for bumpiness and glossiness as a function of bump level—B_j^b and G_j^b, respectively—are indicated in black. Parameter estimates for observers R.K. and F.C. are indicated by the open circles. Standard errors across observers are plotted for each experiment (N = 6).

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 bump level. Likewise, perceived gloss increased with gloss level. These results are not surprising given that observers were screened in advance to ensure that their ordering of bumpiness and glossiness was close to veridical. What *is* surprising is that the functions of perceived bumpiness and gloss were strikingly similar across all observers in each experiment (data not shown, except indirectly in Fig. 5), which suggests that observers used one common perceptual scale for bumpiness and one common scale for glossiness. The internal scaling of bumpiness and glossiness can be described by simple transformations of the corresponding physical properties.

Were observers' judgments of bumpiness and glossiness contaminated by cues to the irrelevant property? Clearly, all parameter estimates B_i^g and G_j^b were greater than zero for i, j = 2, . . . , 5 for the 2 typical observers whose data are shown in Figure 4. 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, pp. 440 ff.).

To do this, we fit the data to an *independent-property* model in which the task-irrelevant contributions (B_i^g and G_j^b) were fixed at zero. Thus, the independent-property model for perceived bumpiness has only four parameters (B_2^b, B_3^b, B_4^b, and B_5^b), as does the independent-property model for perceived gloss. The fits of these independent-property models were compared with the fits of the additive models by the same likelihood-ratio test. The independent-property model was rejected, at the Bonferroni-corrected level, for 4 out of 6 observers in each experiment ($\chi^2 \geq 0.089, p < .008$; see Tables 1 and 2). In other words, most observers could not ignore cues to gloss in making judgments of bumpiness, and, similarly, most observers could not ignore cues to bumpiness in making judgments of glossiness. An increase of gloss increased perceived bumpiness for all observers by an average of 11% of the range of bump levels we used. Likewise, glossiness judgments for surfaces with greater bumpiness increased by an average of 27% of the range of gloss levels used.

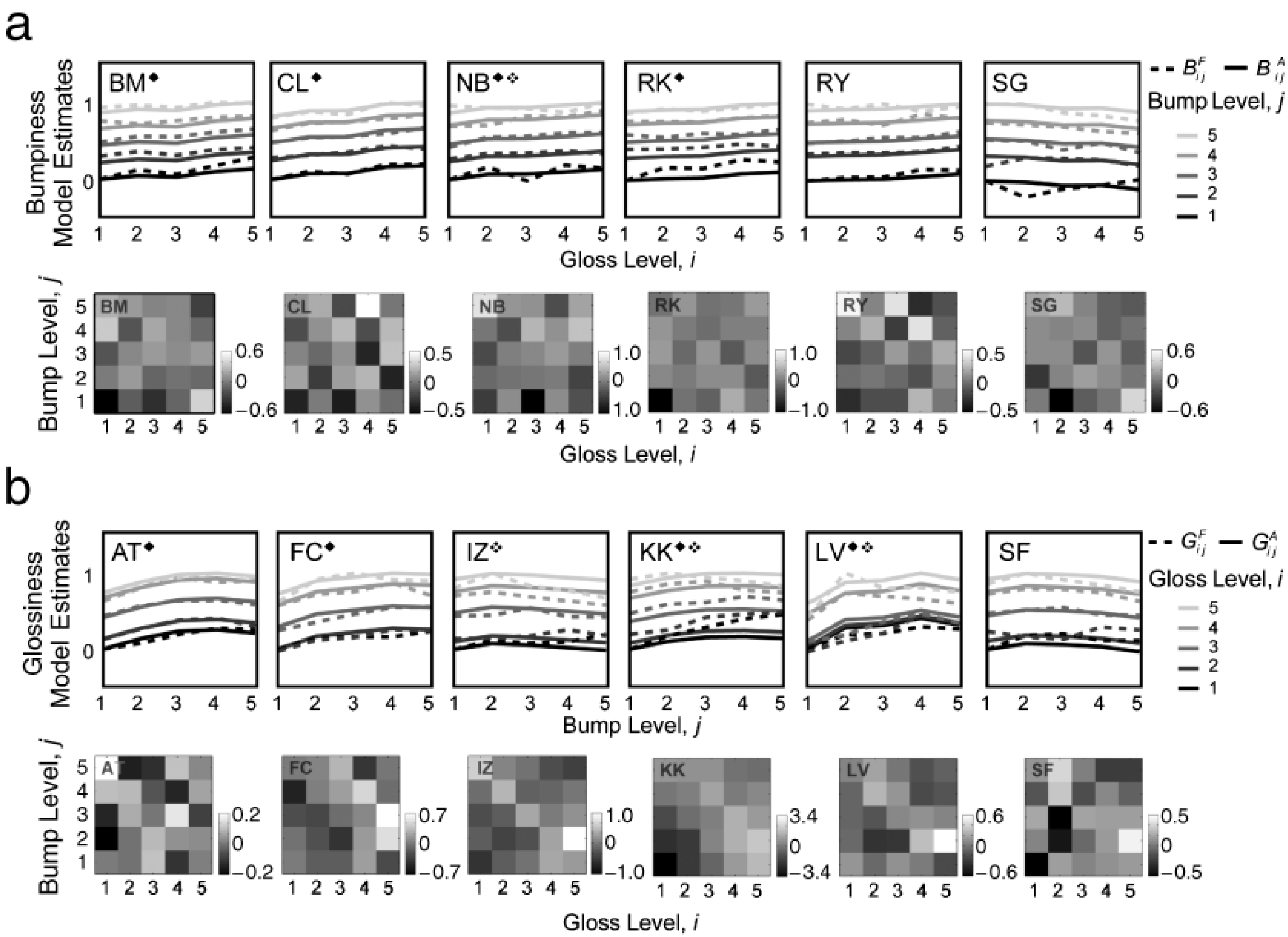

Fig. 5. Predictions of the full and additive models and residual differences for all observers in Experiments 1 and 2. In (a), estimates of perceived bumpiness from the fits of the full (B_{ij}^{F}; dashed lines) and additive (B_{ij}^{A}; solid lines) models are plotted as a function of gloss level, with bump level as the parameter. Corresponding residual differences, $B_{ij}^{F} - B_{ij}^{A}$, are shown as gray-scale plots in the bottom row. In (b), estimates of perceived gloss from the fits of the full (G_{ij}^{F}) and additive (G_{ij}^{A}) models are plotted as a function of bump level, with gloss level as the parameter. Again, corresponding residual differences, $G_{ij}^{F} - G_{ij}^{A}$, are shown in the bottom row. The ◆ symbol indicates participants who exhibited significant differences between the fits of the additive and independent-property models. The ❖ symbol indicates participants who exhibited significant differences between the fits of the full and additive models. (Note that residuals were forced to have a mean of 0; an additive shift of scale values leaves the predictions of both models unchanged.)

In the top rows of Figures 5a and 5b, we plot for all observers the same curves shown in Figure 4c, but in a form commonly used in analysis of variance. The solid lines correspond to the fit of the additive model, which forces these 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 (the same result shown by the upper curves in Fig. 4c). Almost all observers perceived an increase in bumpiness with increasing gloss level, as shown by the slight upward trend in almost all of

TABLE 1

Log-Likelihood Values for Three Models and p *Values for Nested Hypothesis Tests Comparing the Models: Experiment 1*

	Observer					
Model or test	B.M.	C.L.	N.B.	R.K.	R.Y.	S.G.
I. Full model (24 parameters)	−241.95	−182.05	−236.88	−230.77	−258.31	−370.78
II. Additive model (8 parameters)	−248.86	−188.56	−254.30	−238.24	−267.17	−380.73
Test: I vs. II	.612	.671	**.004**	.528	.340	.225
III. Independent-property model (4 parameters)	−267.11	−239.92	−267.75	−250.06	−273.67	−385.27
Test: II vs. III	**< .0001**	**< .0001**	**< .0001**	**< .0001**	.011	.059

Note. Boldface indicates p values that were significant at the Bonferroni-corrected alpha level, .008.

TABLE 2

Log-Likelihood Values for Three Models and p *Values for Nested Hypothesis Tests Comparing the Models: Experiment 2*

	Observer					
Model or test	A.T.	F.C.	I.Z.	K.K.	L.V.	S.F.
I. Full model (24 parameters)	−337.27	−259.48	−277.45	−177.17	−479.28	−322.36
II. Additive model (8 parameters)	−340.71	−273.73	−302.84	−250.38	−495.82	−337.49
Test: I vs. II	.976	.028	**< .0001**	**< .0001**	**.007**	.017
III. Independent-property model (4 parameters)	−379.09	−335.90	−308.10	−278.37	−544.04	−343.08
Test: II vs. III	**< .0001**	**< .0001**	.033	**< .0001**	**< .0001**	.024

Note. Boldface indicates p values that were significant at the Bonferroni-corrected alpha level, .008.

the bumpiness-level contours (Fig. 5a). Similarly, most observers could not ignore bumpiness when making judgments of glossiness (Fig. 5b), although in this case the trend was less clearly monotonic, again confirming the trend observed in Figure 4c.

To evaluate whether the additive conjoint model adequately fits the data, we compared the fit of that model with the fit of a full model that allows nonlinear interactions between the two cues. In the full model, we model the perceived bumpiness, $B_{ij}{}^{\mathrm{F}}$, as an unconstrained value,

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

and we model perceived gloss similarly,

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

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

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

and judges surface S_{ij} as bumpier precisely when the value of Δ is greater than 0. Again, without loss of generality, we anchored the scale by setting $B_{11}{}^{\mathrm{F}} = 0$ and $\sigma = 1$. We then estimated the remaining 24 values of $B_{ij}{}^{\mathrm{F}}$. We fit the glossiness judgments analogously.

The parameter estimates in the full model, $B_{ij}{}^{\mathrm{F}}$ and $G_{ij}{}^{\mathrm{F}}$, are plotted as dashed lines in Figures 5a and 5b, respectively. We performed a nested hypothesis test (Mood et al., 1974, pp. 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 at the Bonferroni-corrected level, the additive model performed just as well as the full model for 5 of the 6 observers in Experiment 1 and 3 of the 6 observers in Experiment 2 ($\chi^2 \geq 32.61, p < .008$; see Tables 1 and 2). This suggests that the additive model sufficiently describes the interactions in the data for 8 of the 12 observers.

We also regressed the additive model's predictions against the full model's predictions to examine any differences between the two models in more detail. R^2 values ranged from .97 to .99 (*Mdn* = .98) for the 6 observers in Experiment 1 and from .83 to .99 (*Mdn* = .96) for the 6 observers in Experiment 2. Thus, for all observers, including those who exhibited significant differences between the additive and full models' predictions, the predictions of the two models were strikingly similar. Finally, to determine if the predictions of the full model differed from those of the additive model in a systematic way, we computed the residual differences between the two models for both judgments (i.e., $B_{ij}{}^{\mathrm{F}} - B_{ij}{}^{\mathrm{A}}$ and $G_{ij}{}^{\mathrm{F}} - G_{ij}{}^{\mathrm{A}}$) and normalized the residuals for each subject to have a mean of 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 an ideal observer judging variations in surface texture should ignore glossiness. In contrast, most human observers in our study perceived physically glossier surfaces to be bumpier and physically bumpier surfaces to be glossier, despite the availability of binocular disparity, a cue potentially used to disambiguate bumpiness from gloss. These patterns of interactions are analogous to those found in previous studies concerning glossiness and megascale judgments of shape. In these studies, curved surfaces with high gloss appeared to have a greater degree of surface curvature than surfaces with low gloss (Braje & Knill, 1994; Mingolla & Todd, 1986; Nishida & Shinya, 1998; Todd & Mingolla, 1983; Todd et al., 1997).

In estimating the perceptual scales of glossiness and bumpiness, 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, although statistically significant, was small on average across observers: about 11% for contamination of bumpiness by gloss and 27% for contamination of gloss by bumpiness, relative to the range of the relevant cue in our stimulus set.

What might explain the interactions we observed? Researchers have proposed that the visual system makes use of simple, image-based statistics to evaluate material properties. In particular, Nishida and Shinya (1998) found that visual estimates of surface reflectance could be modeled by a luminance histogram-matching algorithm. Recently, Motoyoshi, Nishida, Sharan, and Adelson (2007) found that visual estimates of gloss correlate well with the skewness of the distribution of luminance values and that changing the skewness of the distribution affects perception of both surface lightness and glossiness. Motoyoshi et al. proposed a biologically plausible model in which the perception of surface reflectance properties is based on outputs of units in the visual system that are sensitive to image statistics.

If the visual system relies on simple, image-based statistics in judging a surface property such as gloss, then changes in other surface properties or viewing conditions that affect these image statistics may lead to errors in estimating the surface property in question. In a previous study, we found that human observers' estimates of surface roughness were affected by image statistics that varied not only with surface roughness, but also with lighting conditions and viewpoint (Ho et al., 2006, 2007). We proposed that the observed failures in "roughness constancy" were due to erroneous cue learning, and, indeed, it has been shown that associative learning can affect perceptual appearance (for a review, see Haijiang, Saunders, Stone, & Backus, 2006). We advance a similar hypothesis here, suggesting that the observed interactions were the result of imperfect cue learning, an error that could perhaps be reduced with further training involving both visual and haptic assessment of a range of bumpy, glossy surfaces.

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 in this study allowed us to simultaneously estimate the visual perception of two properties at two different scales, the meso- and microscales, 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 the perception of surface material 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. (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.

Billmeyer, F.W., & O'Donnell, F.X.D. (1987). Visual gloss scaling and multidimensional scaling analysis of painted specimens. *Color Research Applications*, *12*, 315–326.

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

Blake, A., & Bülthoff, H. (1991). Shape from specularities: Computation and psychophysics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *331*, 237–252.

Braje, W.L., & Knill, D.C. (1994). Apparent surface shape affects perceived specular reflectance of curved surfaces. *Investigative Ophthalmology and Visual Science*, *35*, S1628.

Chen, T., Goesele, M., & Seidel, H.-P. (2006). Mesostructure from specularity. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1825–1832). Washington, DC: IEEE Computer Society.

Ferwerda, J.A., Pellacini, F., & Greenberg, D.P. (2001). A psychophysically-based model of surface gloss perception. In B.E. Rogowitz & T.N. Pappas (Eds.), *Human vision and electronic imaging VI (proceedings volume)* (pp. 291–301). Bellingham, WA: SPIE Press.

Fleming, R.W., Dror, R.O., & Adelson, E.H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, *3*, 347–368.

Fleming, R.W., Torralba, A., & Adelson, E.H. (2004). Specular reflections and the perception of shape. *Journal of Vision*, *4*, 798–820.

Haijiang, Q., Saunders, J.A., Stone, R.W., & Backus, B.T. (2006). Demonstration of cue recruitment: Change in visual appearance by means of Pavlovian conditioning. *Proceedings of the National Academy of Sciences, USA*, *103*, 483–486.

Ho, Y.-X., Landy, M.S., & Maloney, L.T. (2006). How direction of illumination affects visually perceived surface roughness. *Journal of Vision*, *6*, 634–648.

Ho, Y.-X., Maloney, L.T., & Landy, M.S. (2007). The effect of viewpoint on perceived visual roughness. *Journal of Vision*, *7*, 1–16.

Hunter, R.S., & Harold, R.W. (1987). *The measurement of appearance* (2nd ed.). New York: Wiley.

Koenderink, J.J., & Van Doorn, A. (1996). Illuminance texture due to surface mesostructure. *Journal of the Optical Society of America A*, *13*, 452–463.

Krantz, D.H., Luce, R.D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement: Vol. 1. Additive and polynomial representations*. New York: Academic Press.

Landy, M.S., Maloney, L.T., Johnston, E.B., & Young, M. (1995). Measurement and modeling of the depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389–412.

Luce, R.D., & Tukey, J.W. (1964). Simultaneous conjoint measurement. *Journal of Mathematical Psychology*, *1*, 1–27.

Mingolla, E., & Todd, J.T. (1986). Perception of solid shape from shading. *Biological Cybernetics*, *53*, 137–151.

Mood, A.M., Graybill, F.A., & Boes, D.C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.

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

Nefs, H.T., Koenderink, J.J., & Kappers, A.M.L. (2006). Shape-from-shading for matte and glossy objects. *Acta Psychologica*, *121*, 297–316.

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.

Norman, J.F., Todd, J.T., & Orban, G.A. (2004). Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science*, *15*, 565–570.

Obein, G., Knoblauch, K., & Viénot, F. (2004). Difference scaling of gloss: Nonlinearity, binocularity, and constancy. *Journal of Vision*, *4*, 711–720.

Pellacini, F., Ferwerda, J.A., & Greenberg, D.P. (2000). Toward a psychophysically-based light reflection model for image synthesis. *Computer Graphics*, *34*, 55–64.

Pfund, A.H. (1930). The measurement of gloss. *Journal of the Optical Society of America A*, *20*, 23–26.

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

Roberts, F.S. (1979). Measurement theory. In G.-C. Rota (Ed.), *Encyclopedia of mathematics and its applications* (Vol. 7, pp. 210–232). Reading, MA: Addison-Wesley.

Scheifler, R.W., & Gettys, J. (1996). *X Window System: Core library and standards*. Boston: Digital Press.

Sève, R. (1993). Problems connected with the concept of gloss. *Color Research & Application*, *18*, 241–252.

Todd, J.T., & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 583–595.

Todd, J.T., Norman, J.F., Koenderink, J.J., & Kappers, A.M. (1997). Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. *Perception*, *26*, 807–822.

Wang, J., & Dana, K.J. (2006). Relief texture from specularities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 446–457.

Ward, G.J. (1994). The RADIANCE lighting simulation and rendering system. *Computer Graphics*, *28*, 459–472.

Zaidi, Q. (2001). Color constancy in a rough world. *Color Research & Application*, *26*, S192–S200.

(Received 4/25/07; Revision accepted 8/19/07)