Gloss Perception in Painterly and Cartoon Rendering

Adrien Bousseau, James P. O'Shea, Frédo Durand, Ravi Ramamoorthi,
Maneesh Agrawala

▶ **To cite this version:**

HAL Id: hal-00811637
https://hal.inria.fr/hal-00811637

Submitted on 10 Apr 2013

Gloss Perception in Painterly and Cartoon Rendering

Adrien Bousseau [1,2], James P. O'Shea [2], Frédo Durand [3], Ravi Ramamoorthi [2], Maneesh Agrawala [2]

[1] REVES/INRIA Sophia Antipolis

[2] University of California, Berkeley

[3] MIT CSAIL

Depictions with traditional media such as painting and drawing represent scene content in a stylized manner. It is unclear however how well stylized images depict scene properties like shape, material and lighting. In this paper, we describe the first study of material perception in stylized images (specifically painting and cartoon) and use non photorealistic rendering algorithms to evaluate how such stylization alters the perception of gloss. Our study reveals a compression of the range of representable gloss in stylized images so that shiny materials appear more diffuse in painterly rendering, while diffuse materials appear shinier in cartoon images. From our measurements we estimate the function that maps realistic gloss parameters to their perception in a stylized rendering. This mapping allows users of NPR algorithms to predict the perception of gloss in their images. The inverse of this function exaggerates gloss properties to make the contrast between materials in a stylized image more faithful. We have conducted our experiment both in a lab and on a crowdsourcing website. While crowdsourcing allows us to quickly design our pilot study, a lab experiment provides more control on how subjects perform the task. We provide a detailed comparison of the results obtained with the two approaches and discuss their advantages and drawbacks for studies like ours.

1. INTRODUCTION

One of the main goals of painting and drawing is to suggest scene content in a simplified or stylized manner. Such stylized depictions are often surprisingly effective despite their departure from realism. Our goal is to better understand of how well stylized images depict scene properties. As a first step we focus on the evaluation of gloss perception in painting and cartoon images.

Existing work focus on the evaluation of shape depiction in stylized images [Winnemöller et al. 2007; Cole et al. 2009] and no study exists on the evaluation of material depiction, despite the variety of materials that one may wish to depict in an illustration. What makes an object look shiny in a painting? Can we depict a diffuse object in a cartoon? Artists often rely on their experience of their media to answer such questions and depict materials in different styles [Cooke 1967; Johnson 1992; Ott and Kuseno 2005]. However, this artistic knowledge is often implicit and while high level rules exist to depict light and shade in a given style, no guidelines exist to vary low level material properties such as the amount of gloss. In this paper we explore the use of non photorealistic rendering (NPR) as a tool to systematically study the effects of style parameters on material perception. Our aim is to build an explicit set of guidelines for depicting material in stylized images and we first investigate how painterly and cartoon styles influence the perception of gloss.

We build on Pellacini et al.'s [2000] psychophysical model of gloss perception which identifies contrast and sharpness of highlights as the two dimensions that people are most sensitive to when viewing glossy materials. As stylization directly alters both of these dimensions, we expect stylization to also alter gloss. In painterly rendering, large brush strokes eliminate or spread out the small specular highlights that contribute to the appearance of shininess. But opaque strokes also increase the number of sharp edges in diffuse regions of the image (Figure 1b) and may exaggerate the perception of gloss. Semi-transparent strokes primarily reduce local contrast making the material appear more diffuse (Figure 1c). Cartoon rendering quantizes colors and replaces smooth variations with sharp boundaries making the surface appear shinier (Figure 1d).

In this paper we present a series of quantitative perceptual studies that examine how such artistic style parameters affect gloss perception. We focus on painterly rendering and cartoon rendering which have received great attention in the computer graphics literature [Haeberli 1990; Meier 1996; Litwinowicz 1997; Hertzmann 1998; Hays and Essa 2004; Zeng et al. 2009; DeCarlo and Santella 2002; Winnemöeller et al. 2006]. In industry, numerous video games (*Jet Set Radio*, *Zelda: The Wind Waker*, *XIII*) and movies (*What Dreams May Come*, *Tarzan*, *Waking Life*, *A Scanner Darkly*) rely on painterly and cartoon styles similar to the ones we study. While our results are not directly relevant to other NPR algorithms, they are indicative of the types of effects that one can observe in related styles such as watercolor [Curtis et al. 1997].
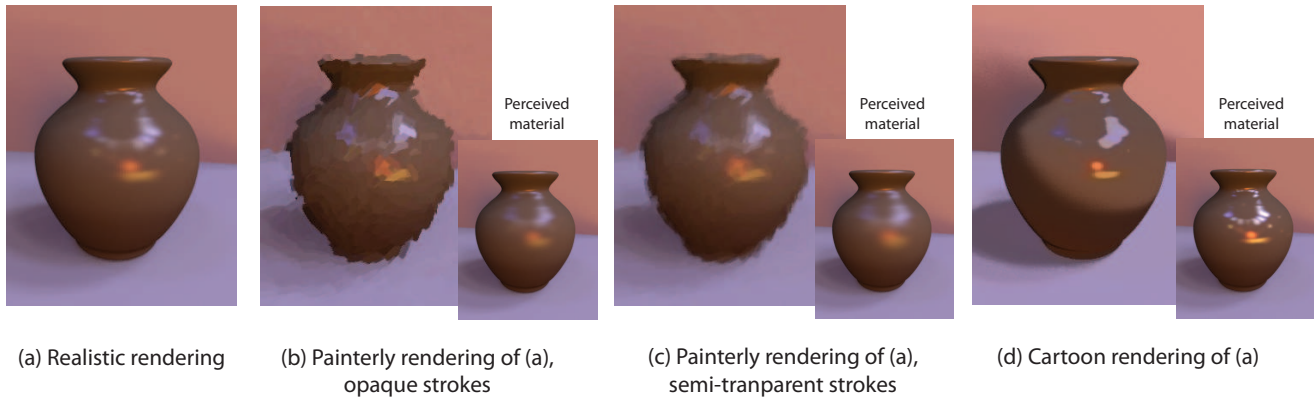
(a) Realistic rendering   (b) Painterly rendering of (a), opaque strokes   (c) Painterly rendering of (a), semi-tranparent strokes   (d) Cartoon rendering of (a)

Fig. 1: *Each stylization affects gloss perception differently. In painterly rendering, opaque strokes (b) removes some highlights and semi-transparent strokes (c) blend colors, making shiny materials appear more diffuse. In contrast, cartoon rendering exaggerates shininess (d). In this paper, we evaluate how people perceive gloss in stylized images, and we derive the function that predicts for a given gloss how it will be perceived after stylization, as shown here in insets.*

For painterly rendering we measure the effect of brush size, brush opacity and Hertzmann's [2002] brush bump mapping which simulates texture due to brush bristles. Out of many parameters, these three have the strongest impact on contrast and sharpness in the image and are shared by most algorithms. For cartoon rendering we consider the effect of quantization softness. While most cartoon rendering algorithms perform a hard color quantization, a soft quantization produces more subtle stylizations [Winnemöeller et al. 2006]. Finally we compare the effect of these non-photorealistic styles to the effect of a simple Gaussian blur and show that while both painterly rendering and blur remove details in the image, painterly rendering offers a better preservation of gloss variations.

Our study yields a number of key insights on the perception of gloss in cartoon and painterly images. First, we observe a compression of the range of perceivable gloss as stylization increases. We measure this compression and deduce the range of gloss that can be depicted in each of the styles we study. In particular, we find that painterly rendering cannot accurately depict shiny materials, especially when semi-transparent brush strokes are used. In contrast, cartoon rendering increases the perception of shininess for diffuse materials. Our study also reveals counter intuitive perceptual effects; although bump mapping introduces small-scale highlights over a painterly image, these additional variations reduce the perceived shininess. Finally our study yields novel insights on the perception of gloss in realistic renderings as we observe a correlation between perceived contrast and sharpness for materials in the mid-gloss range. This result differs from that of previous work [Pellacini et al. 2000; Fleming et al. 2003] which suggests that these two parameters are perceptually independent.

We leverage the low cost and scalability of crowdsourcing to design and conduct the pilot study of our experiment. We then replicate this study in a lab to validate our results. We discuss the pros and cons of the two approaches. Although crowdsourcing allows us to quickly identify general trends, the lab data reveal less variance and a more accurate perception of contrast due to additional control on the viewing conditions.

As an application of the data collected in our study, we estimate the function that maps realistic gloss descriptions to their perceptual values according to style parameters. This mapping predicts how materials will be perceived when rendered in a given style. The

inverse mapping indicates which style best depicts a given material, or how to exaggerate gloss to obtain a desired perception.

To summarize, this paper makes the following contributions:

—We conduct the first evaluation of material perception in stylized rendering.

—We compare the effect of brush size, brush opacity, brush bump mapping, cartoon quantization and blur.

—We measure how these different style parameters reduce the range of perceivable gloss.

—We compute from our measurements the mapping that predicts the perception of gloss in a painterly or cartoon image as a function of style parameters.

2. RELATED WORK

While guidelines on material depiction exist in art books [Cooke 1967; Johnson 1992; Ott and Kuseno 2005], these guidelines are often very high level, such as "Apply a white highlight to suggest shininess." We have not found lower level instructions explaining how to vary style parameters such as brush size of opacity to depict material variations like gloss. Our study represents a first step in that direction as we relate material perception in stylized images to controlled BRDF and style parameters used in common rendering engines. We design this study by taking inspiration from previous work on the perception of materials in realistic images and on the perception of shape and faces in stylized images.

Material Perception in Realistic Images. Pellacini et al. [2000] conduct a study to estimate the dimensionality of gloss perception. They use multidimensional scaling (MDS) to derive a perceptually uniform space expressed as a reparameterization of Ward's BRDF model [1992], with two parameters corresponding to the contrast and sharpness of highlights. The goodness of fit of a confirmatory MDS measures the independence of these two dimensions. Wills et al. [2009] present a similar experiment to derive a perceptual embedding of measured BRDFs. Complementary to these studies, Nishida and Shinya [1998] and Vangorp et al. [2007] measured that the accuracy of material perception is influenced by shape. Among the shapes Vangorp et al. use, a blob was the most descriminative.

Fleming et al. [2003] show that the recognition of surface reflectance is improved when objects are illuminated under natural environments. These results suggest that natural image statistics such as color and derivative histograms provide strong cues for material perception [Dror et al. 2001]. Ramanarayanan et al. [2007] evaluate if transformations of the lighting environment such as blurring and warping are perceivable given various geometries and materials. They observed that blurring the illumination is harder to perceive for diffuse materials, and that warping is harder to perceive for bumpy surfaces. They deduce from these observations a *visual equivalence* metric between images. While the stylizations studied in our paper could be seen as forms of blurring or warping, they occur on the final image, not on the reflected environment.

Kozlowski and Kautz [2007] and Křivánek et al. [2010] evaluate how approximations of the rendering equation alters appearance for various shapes and materials. Křivánek et al. deduce from their study the range of parameters of the Virtual Point Light algorithm that produce renderings that are visually equivalent to reference solutions. Kozlowski and Kautz conclude that approximations in the rendering are less noticeable for complex geometry and diffuse materials. In this paper we vary material and style parameters and leave the study of geometric variations for future work.

Perception in Non Photorealistic Rendering. A standard approach to evaluate the effectiveness of NPR depictions is to measure their performance on recognition tasks. Winnemöller et al. [2007] evaluate different shape cues (shading, textures, contours, motion) for shape recognition, and Cole et al. [2009] compare the ability of several line drawing algorithms to depict shape. They conclude that line drawings depict certain shapes almost as well as shaded images. Xue et al. [2010] generate patterns that enhance the shape details of an object and measure the effectiveness of different patterns in a psychophysical experiment. Gooch et al. [2004] show that faces depicted as illustrations or caricatures are faster to learn than photographs and equally recognizable. On the same topic, Wallraven et al. [2007] study the impact of several styles on the recognition of facial expressions. Among the different styles evaluated in the study (painting, cartoon, illustration), painterly images result in the worst recognition but the best preservation of facial expression intensity for increasing brush sizes. Smith et al. [2010] derive the parameters of a pen-and-ink algorithm from material parameters (tone, gloss, texture). They validate their approach with a user study, but do not evaluate how variations in the style parameters affect the perception of materials. In this paper we use a matching task to evaluate how glossy materials are perceived under varying styles.

3. BACKGROUND ON GLOSS IN REALISTIC IMAGES

Pellacini et al. [2000] have shown that the space of gloss is two dimensional. The first dimension, called contrast gloss c, corresponds to the perceived relative brightness of the diffuse and specular components. The second dimension, called distinctness-of-image gloss d, corresponds to the perceived sharpness of the specular highlights. In the remainder of this paper, we refer to c as *contrast* and d as *sharpness*. We illustrate material variations along the c and d dimensions in Figure 2.

Pellacini et al. define c and d with respect to the Ward isotropic BRDF [Ward 1992] as:

$$c = \sqrt[3]{\rho_s + \rho_d/2} - \sqrt[3]{\rho_d/2} \qquad (1)$$
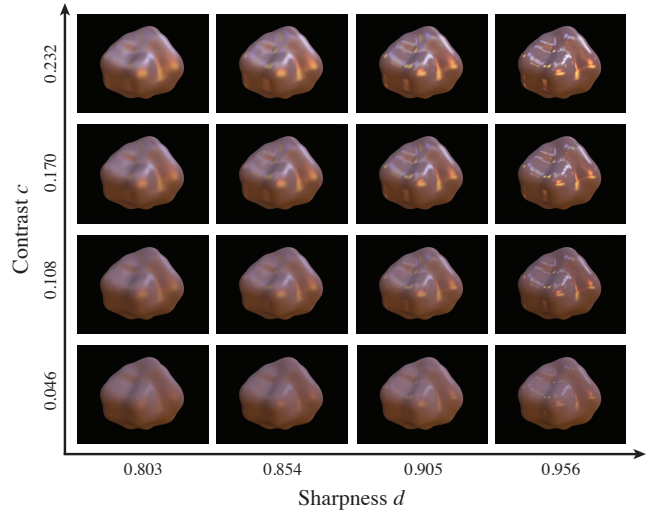$$d = 1 - \alpha \qquad (2)$$

Fig. 2: *Set of target materials used in our study, here rendered without stylization. Note that a larger set of materials is used for the match sliders.*

where ρ_d, ρ_s and α correspond respectively to the diffuse reflectance, the specular reflectance and the surface roughness of Ward's model:

$$f(\theta_i, \theta_o) = \frac{\rho_d}{\pi} + \frac{\rho_s}{4\pi\alpha^2 \sqrt{\cos\theta_i \cos\theta_o}} e^{-\frac{\tan^2\theta_h}{\alpha^2}} \qquad (3)$$

with θ_i and θ_o the incoming and outgoing radiance directions and θ_h the angle between the surface normal and the half-vector. The perceptual distance between two materials in gloss space is then:

$$D_{ij} = \sqrt{[c_i - c_j]^2 + [1.78(d_i - d_j)]^2} \qquad (4)$$

where the scale factor 1.78 is required to make the space perceptually uniform. In this paper, we express the gloss value of a material as its perceptual distance to the most diffuse material of the space of materials we study.

Pellacini et al. also introduce the notion of *iso-gloss contours* that correspond to materials of the gloss space that are equidistant to a reference material. According to their model, iso-gloss materials are perceived as equivalent in gloss as compared to the reference material: a material with high contrast blurry highlights will be perceived as equally glossy to a material with low contrast sharp highlights. Pellacini et al. support this prediction by an informal ranking task, and our results confirm this finding. In addition, Pellacini et al. report that the c and d axes are independent, i.e. that perceived contrast is not a function of sharpness and vice versa. The data collected by Fleming et al. [2003] support this finding since they found no statistical dependence of contrast over perceived sharpness nor of sharpness over perceived contrast. However, our findings differ from these previous observations as we identify a correlation between the two dimensions for materials in the mid-gloss range (Section 6.2).

Ferwerda et al. [2001] measured the just-noticeable differences (JND) for the two dimensions of the gloss space as $\Delta c = 0.031$ and $\Delta d = 0.017$.
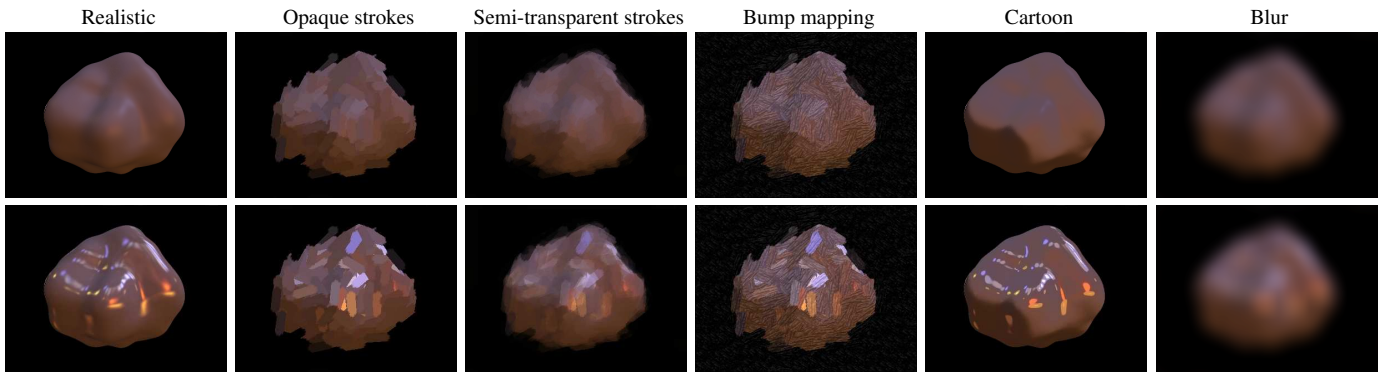
Realistic	Opaque strokes	Semi-transparent strokes	Bump mapping	Cartoon	Blur

Fig. 3: *Subset of the images used in the experiment. Notice the difference between the various stylizations on a diffuse object (top row, $c = 0.046$ and $d = 0.803$) and a shiny object (bottom row, $c = 0.170$ and $d = 0.956$). In particular, painterly rendering makes the material appear more diffuse, while cartoon increases shininess.*

4. METHODOLOGY

Mechanical Turk Study. Inspired by recent online perceptual studies (e.g. [Cole et al. 2009; Heer and Bostock 2010]), we used the crowdsourcing website Amazon Mechanical Turk to accelerate the design of our study. The Mechanical Turk is an internet service on which workers are paid to perform small tasks for requesters. A task is often paid between \$0.01 and \$0.20, making experiments like ours inexpensive to conduct. In addition, because workers complete tasks in parallel, a large number of tasks can be performed quickly. In our case, the experiments were performed in a day or two, which allowed us to design the experiment iteratively. As an example, in an early iteration of our experiment we used a smaller range of values for our interface sliders and quickly discovered that this leads to floor and ceiling effects in the results: many subjects set the sharpness and contrast values to the extremes of the sliders because they could not select higher or lower values that may correspond to their perception. We describe the final design of our experiment in Section 6.

15 to 30 different Mechanical Turk subjects performed each of our tasks. Each subject can only perform a task once, but nothing enforces the same subject to perform all the tasks of an experiment. Subjects were paid \$0.03 per task and had 3 minutes to enter their settings, although they completed the task in 30 seconds on average. We used a qualification test to explain to subjects the concepts of painterly and cartoon rendering, and the notion of sharpness and contrast for glossy materials. We provide the qualification test as supplemental materials. The qualification test also contained a simplified version of the task to familiarize subjects with the space of gloss covered by the sliders of the interface.

Lab Study. The downside of crowdsourcing in comparison to a lab study is that experimenters have less control on how workers perform the task. The calibration of the monitor and lighting conditions, for example, are unknown and reflect the variety of viewing conditions encountered on the web. As a result, data obtained from the Mechanical Turk can contain more variance than data obtained from a lab study. However this reduced control is compensated by the larger quantity of data that we can collect. We provide an evaluation of the Mechanical Turk data by replicating the final design of our experiment in a lab. The lab data show a good agreement with the crowdsourcing data but reveal higher accuracy along the contrast dimension.

For each style, three subjects participated in the study and we collected 10 responses per task from each observer[1]. All subjects were students (21-26 years of age), novice in computer graphics and unaware of the experimental hypotheses. They had normal visual acuity and wore optical corrections during testing when needed. Subjects were also instructed to complete the same qualification test as the one we used on the Mechanical Turk.

5. HYPOTHESES

As shown in Figures 1 and 3 stylization modifies the contrast and sharpness of highlights in an image and thereby alters the perception of gloss. We expect the range of representable materials to differ as we change the style parameters because each parameter affects the appearance of highlights in different ways. Our study quantitatively evaluates how these style parameters affect gloss perception. We consider three style parameters for painterly rendering – brush size, brush opacity and brush bump mapping and one parameter for cartoon rendering – quantization. Most painterly and cartoon rendering algorithms give access to these parameters which have the strongest impact on sharpness and contrast. We also include a simple image blur as an additional style for comparison.

Our hypotheses are:

H1: In painterly rendering, brush strokes alter sharpness by either eliminating or spreading sharp highlights. We expect shiny materials to appear more diffuse in this style.

H2: Opaque strokes also introduce sharp edges in diffuse regions of the image. Thus, we expect diffuse materials to appear shinier.

H3: Semi-transparent strokes blend colors and reduces local contrast. We expect this reduction of contrast to make shiny materials appear more diffuse.

H4: Bump mapping introduces high frequency details that may be interpreted as specular highlights. We expect the increase in contrast to make materials appear shinier.

H5: The quantization used in most cartoon shading sharpens the image and we expect diffuse materials to appear shinier.

H6: Blur reduces both contrast and sharpness in the image. We expect blur to make objects to appear more diffuse.

---

[1]An exception is the painterly style with opaque stokes, for which we had five subjects and collected 15 responses.

Match the sharpness and contrast

In the following images, an abstract shape is rendered in a painterly style and a different shape is rendered realistically.
Click the arrows to increase or decrease the sharpness and contrast of the realistic image until it matches the painterly image.
The painterly image may display brush strokes of various sizes, from very fine to very coarse.
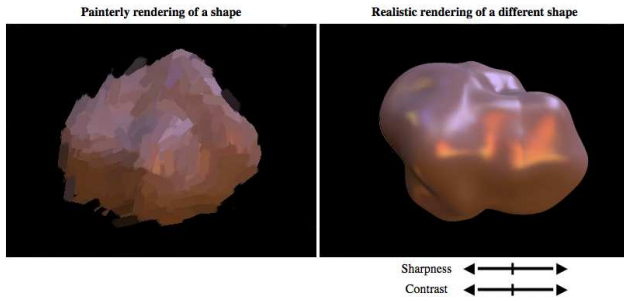
Painterly rendering of a shape   **Realistic rendering of a different shape**

Sharpness ◄ ──┼── ►
Contrast ◄ ──┼── ►

Fig. 4: *Screen capture of the user interface. Subjects adjust the contrast and sharpness of the highlights in the realistic image until it matches the contrast and sharpness in the painterly image.*

6. EXPERIMENTAL DESIGN

In order to evaluate our hypothesis, we asked subjects to assess gloss in stylized images using a single-interval matching task.

6.1 Task

We simultaneously present two images of different blobby shapes (Figure 4) in each trial of our matching task. One image is stylized (the "target") and the other image is rendered in a realistic manner (the "match"). We instruct subjects to adjust the contrast and sharpness parameters of the BRDF in the realistic image until it corresponds to the perceived material in the stylized image. We use the method of adjustment [Baird and Noma 1978] instead of a two-alternative forced choice (2AFC) matching task because a method of adjustment yields results using fewer trials and is better suited to subjective tasks like ours. Fleming et al. [2003] use a similar matching task to evaluate the perception of gloss under different illumination conditions.

Each image in our study represents an abstract blobby shape under a realistic lighting environment. We follow the approach of Vangorp et al. [2007] who show that people are more accurate in matching materials between blobby shapes as it factors out the influence of familiar shapes. A different blob is used for the match and target images, so that subjects cannot match images based on shape. Natural environment lighting also improves material perception [Fleming et al. 2001; 2003] and we use the *Grace* environment map for the experiment (http://www.debevec.org/Probes/).

For the online study, we displayed the task via a web browser using the Mechanical Turk interface. We assumed monitors to have a gamma $\gamma = 2.2$, which is the setting of most displays. We implemented a stand-alone version of the experiment for our lab study to avoid cluttering the display with the frame of a web browser. We conducted the experiment using a desktop computer and a 19-inch CRT display set to a resolution of 2048×1536. The background screen was gray apart from the stimulus images and the response controls. The observers head position was maintained using a chin rest. Observers were positioned directly in front of the display screen at a distance of 40cm, so that each pixel subtended approximately 1.7×1.7 arcmin. We gamma-corrected the display to linearize the luminance function for each color channel ($\gamma = 1.0$). The room was dark except for the light from the display screen.

Material Variations. The space of gloss that we study covers materials ranging from mirror-like to nearly diffuse. The sliders of the user interface vary from 0.015 to 0.263 for c and from 0.769 to 0.99 for d, with step sizes equal to one just-noticeable difference (JND) Δc and Δd respectively. The sliders are initialized to random values in these ranges.

Our stimulus set is made of a 4×4 regular sampling of this 2D gloss space. The contrast c varies from 0.046 to 0.232 with a step size equal to two JNDs Δc. The sharpness d varies from 0.803 to 0.956 with a step size equal to three JNDs Δd. We use a bigger step size in the sharpness dimension to compensate for the fact that $\Delta d < \Delta c$. Using a gray diffuse reflectance $\rho_d = 0.2$, this stimulus set corresponds to Ward parameters ρ_s and α varying from 0.0328 to 0.2374 and 0.044 to 0.197 respectively. Figure 2 shows the resulting set of materials. Note that the interface sliders cover values beyond the $\{c, d\}$ values of the stimuli to avoid floor and ceiling effects in the experiment, as explained in Section 4.

We precomputed all the images to provide immediate visual feedback to the user. We used PBRT [Pharr and Humphreys 2004] to render the images at a resolution of 400×300 pixels. We linearly scaled the dynamic range so that the brightest highlight of the most shiny material would map to 1. We then gamma corrected the renderings for display ($\gamma = 1.0$ in the lab, $\gamma = 2.2$ on the Mechanical Turk). For the lab study we bilinearly scaled the images by a factor of 1.2 to measure approximately 9.1×7.0 cm on the screen and subtend 13×10 degrees.

Style Variations. For painterly rendering, we stylize each image with three different brush sizes equal to $\{4 \times 12, 8 \times 24, 16 \times 48\}$ pixels, which spans styles from detailed to very coarse on the 400×300 images in our stimulus set, as shown in Figure 5(d,e,f). We use a variation of Haeberli's algorithm [1990] to create the painterly images because it is simple to implement and matches our needs for the study of the effect of brush size on perception. More advanced algorithms [Hertzmann 1998; Hays and Essa 2004] do not maintain a uniform brush size over the image and make use of stroke clipping and coarse-to-fine painting to preserve the image information. Our goal in contrast is to evaluate material perception when the image information is altered by the strokes. Our implementation distributes brush strokes over the image using stratified sampling, and orients the strokes along the image contours based on a smoothed edge tangent flow [Kang et al. 2007]. We sample colors in the image at the center of each brush stroke. For semi-transparent strokes the color is modulated by an opacity value of 0.5. Finally, we use Hertzmann's [2002] bump mapping technique to mimic the texture of brush strokes and varnish. We use the same set of stroke locations and orientations for every stimulus image. The only variables are the brush parameters (size, opacity, bump map) and material parameters.

For cartoon rendering, we apply Winnemöller's soft quantization [2006] on the luminance channel of the realistic rendering converted to CIE $L^*a^*b^*$ color space. A soft-quantized luminance L_q is expressed as:

$$L_q(x) = q_{nearest} + \frac{\Delta q}{2} \tanh(\varphi_q(L(x) - q_{nearest})) \quad (5)$$

where L is the input luminance value, Δq is the bin width (fixed to 15 in our images), $q_{nearest}$ is the bin boundary closest to $L(x)$, and the scalar φ_q defines the sharpness of the transition between two successive bins. A soft quantization produces less aliasing than hard quantization and allows us to also evaluate the perception of intermediate images, between purely realistic and purely cartoon
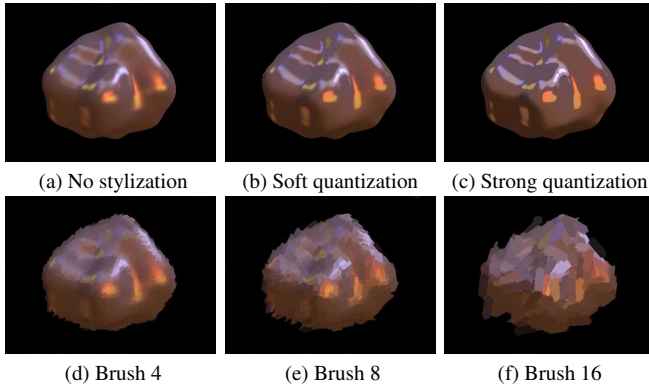
(a) No stylization   (b) Soft quantization   (c) Strong quantization

(d) Brush 4   (e) Brush 8   (f) Brush 16

Fig. 5: *(b,c) Effect of strong and soft quantization for the cartoon effect. (d,e,f) Effect of brush size in painterly rendering.*

images. The cartoon images are computed with two different levels of quantization: a strong quantization to create sharp edges, and a softer quantization for a more subtle stylization. These quantizations correspond to sharpness values φ_q equal to 0.3 and 0.6 respectively. We illustrate the effect of soft and strong quantizations in Figure 5(b,c). Finally, we compute the blurred images with Gaussian kernels of standard deviations equal to $\{4, 8, 16\}$. Figure 3 shows a subset of the stimulus images for each style.

In addition to the stylized images, we also show the subjects a set of images rendered without stylization, so that we can relate the effect of stylization to the perception of realistic images. Our experiment contains 64 matching tasks for the painterly styles and blur (4×4 materials, 3 brush or blur sizes and 1 realistic setting) and 48 matching tasks for the cartoon style (4×4 materials, 2 quantization softness and 1 realistic setting), for a total of 304 matching task that we repeated 10 times in the lab study.

6.2 Results

We summarize all our data at the end of the paper, and provide the data of the individual subjects as supplemental material. Figure 11 summarizes the data collected in our Mechanical Turk study and Figure 12 summarizes the data collected in our lab study. Arrow plots and ellipse plots visualize the mean and standard deviation of subjects' settings, respectively. The origin of each arrow corresponds to the c, d position of a reference material in the target set, and the endpoint corresponds to the mean of the subjects' settings for that material. Ellipses depict the standard deviation of subjects' settings along the two main axes of the covariance matrix. We assign a different color to each pair of arrow and ellipse to differentiate each reference material. The origin of the frame corresponds to the most diffuse material covered by the interface sliders ($c = 0.015$ and $d = 0.769$), and dashed curves represent iso-gloss contours with respect to this origin. From Equation 4, we express the gloss value of a material as the perceptual distance to the origin of the material space:

$$g(c, d) = \sqrt{[c - 0.015]^2 + [1.78(d - 0.769)]^2}. \quad (6)$$

We also show bar plots in Figures 11 and 12 that visualize the projection of subjects' settings along the gloss dimension. Bars correspond to the mean and standard deviation of the gloss value of subjects' settings. A dashed line indicates the ideal settings (perceived gloss = true gloss).

We first discuss the results obtained in the lab study since the lab data contain less variance. We then discuss the similarities and differences between the lab data and the Mechanical Turk data.

Perception of Iso-Gloss Materials. While our study focuses on the perception of gloss in stylized images, it also reveals valuable information about realistic images. In the realistic case (Figure 12, top left), distributions for mid-gloss materials are oriented along the iso-gloss diagonal from high-contrast low-sharpness to low-contrast high-sharpness. We further analyze this correlation between contrast and sharpness at the end of this section. For low-contrast low-sharpness materials, highlights are hard to perceive and subjects tend to not distinguish different values of sharpness (horizontal ellipses in the bottom left area of the space), while in the presence of high sharpness there is only uncertainty about contrast (vertical ellipses in the top right area of the space). In addition, arrows for mid-gloss materials are aligned with the iso-gloss diagonal (top and left of the space), indicating a tendency to favor median materials over more extreme iso-gloss counterparts.

We visualize the projection of subject's settings along the gloss dimension as bar plots in Figure 12. We observe a slight overestimation of gloss, although this deviation is in most cases smaller than the standard deviation. The standard deviation of perceived gloss in realistic images is equal to 0.036 on average, which corresponds to approximately one JND of contrast Δc or two JND of sharpness Δd (Ferwerda et al. [2001] do not provide the value of a JND in the gloss dimension). Note however that the standard deviation is higher for the low contrast materials, for which subjects cannot clearly distinguish highlights.

Painterly Rendering with Opaque Strokes. The strongest effect of painterly rendering that we observe is a compression of the range of perceivable materials as we increase brush size. The compression is stronger along the sharpness dimension than along the contrast dimension. A strong stylization (brush size 16) compresses the range of mean perceived gloss by a factor of 1.6, from $[0.07, 0.4]$ to $[0.16, 0.37]$ so that diffuse materials appear shinier while shiny materials appear more diffuse. We illustrate the range of perceivable materials for each style in Figure 6.

The average standard deviation of perceived gloss increases slightly with brush size, from a value of 0.045 for brush size 4 to 0.055 for brush size 16. Note also that as the space of perceived material is compressed, the effective perceived distance between two distinct materials reduces and eventually becomes smaller than the standard deviation. The distinction between materials would only be preserved if the standard deviation reduced at the same rate as the space compresses.

Painterly Rendering with Semi-Transparent Strokes. Subjects perceive most materials as more diffuse under this style, so that diffuse materials are better preserved than with opaque brush strokes. The range of perceivable gloss reduces from $[0.07, 0.4]$ to $[0.08, 0.28]$ with an average standard deviation of 0.06 at brush size 16.

Painterly Rendering with Bump Mapping. Bump mapping introduces small details and highlights over the image. Our results contradict our hypothesis and reveal that these variations make shiny materials appear more diffuse, with a range of perceived gloss compressed from $[0.07, 0.4]$ to $[0.11, 0.33]$ with brush size 16 (Figure 6). The standard deviation of perceived gloss remains on average equal to 0.04.
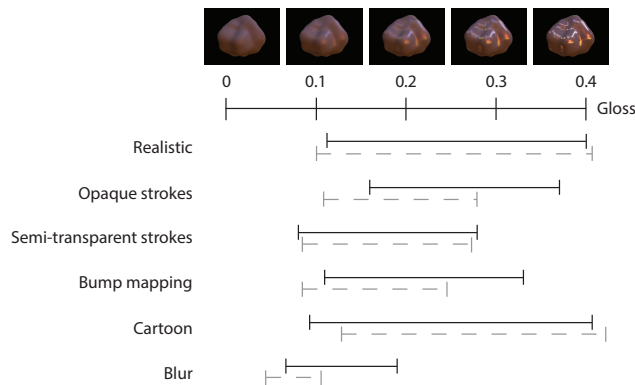
Fig. 6: *Range of mean perceived gloss for each style studied in our experiment (brush size 16 or strong quantization). Dark lines indicate the lab data while grey dashed lines indicate the Mechanical Turk data. Shiny materials cannot be depicted with painterly rendering, while diffuse materials are not well preserved by cartoon rendering.*

Correlation between material sharpness and perceived contrast					
Reference c		0.046	0.108	0.170	0.232
MTurk	correlation	**0.37**	**0.41**	**0.47**	**0.35**
	p-value	0.0000	0.0000	0.0000	0.0000
Lab	correlation	0.06	**0.30**	**0.25**	**0.11**
	p-value	0.11	0.0000	0.0000	0.003
Correlation between material contrast and perceived sharpness					
Reference d		0.803	0.854	0.905	0.956
MTurk	correlation	**0.39**	**0.54**	**0.58**	**0.42**
	p-value	0.0000	0.0000	0.0000	0.0000
Lab	correlation	0.05	**0.30**	**0.32**	**0.11**
	p-value	0.17	0.0000	0.0000	0.0026

Table I. : *Correlation between material sharpness and perceived contrast, for each reference contrast in our dataset (top), and correlation between material contrast and perceived sharpness, for each reference sharpness (bottom). The correlation is significant for all the materials in the Mechanical Turk data but is not significant for the low contrast and sharpness materials in the lab data. We highlight significant correlations in bold (p-value < 0.01).*

Cartoon Quantization. Cartoon rendering makes diffuse materials appear shinier by increasing their sharpness (top left of the space of materials). However, this effect does not occur for low-contrast materials (bottom left of the graph), for which quantization eliminates glossy highlights with very low contrast. In contrast with painterly rendering, cartoon rendering doesn't affect shiny materials significantly and the overall range of perceived gloss is well preserved, equal to $[0.09, 0.4]$. The average standard deviation in perceived gloss for strong quantization remains at 0.045.

Gaussian Blur. We compare the previous measurements with the effect of Gaussian blur to assess if painterly rendering is "just a blur". Our results confirm the intuition that blurring the image makes materials appear more diffuse, and that this effect is more dramatic than the one observed with painterly rendering. For a blur kernel of standard deviation 16, all the materials are perceived in a limited range of gloss equal to $[0.065, 0.19]$ with an average standard deviation equal to 0.04. In contrast, the various styles of painterly rendering can depict materials in a range of gloss of $[0.08, 0.37]$ at brush size 16. This result shows that even if painterly rendering significantly simplifies the image and removes details, it does it in a different way than blur and offers a better depiction of material variations.

Comparison with the Mechanical Turk. The data collected on the Mechanical Turk agree with the general trends observed with the lab data. In particular the Mechanical Turk data are accurate enough to confirm our main observations on iso-gloss materials (Figure 11, top left) and on the compression of the range of perceived materials for each stylization (Figure 6). We observe however a stronger compression along the contrast dimension in the Mechanical Turk data. The standard deviation in perceived gloss is also stronger on average, with a value of 0.05 instead of 0.034 for the realistic case.

Summary and Discussion. We have observed two forms of deviations in the subjects' settings: one along the iso-gloss contours, and one across them.

The first deviation occurs along the iso-gloss contours where the distributions of subjects' settings are oriented along the diagonal from high-contrast low-sharpness to low-contrast high-sharpness, even for realistic images. The center of the distributions also tend to move along the iso-gloss contours and away from the extreme materials. This deviation could be due to the fact that when subjects are uncertain, they prefer to avoid the interpretation that corresponds to the ends of the scales in the matching experiment. Fleming et al. [2003] report a similar bias in their experiment. We conclude from these observations that for mid-gloss materials, subjects confound an increase in contrast and decrease in sharpness with an increase in sharpness and decrease in contrast.

We performed an analysis of correlation to further evaluate the dependence between contrast and sharpness. We report in Table I (top) the correlation between material sharpness and the perceived contrast in realistic renderings, for each reference contrast in our dataset. Table I (bottom) reports the correlation between material contrast and perceived sharpness. In the Mechanical Turk data, we measure a positive correlation that varies between 0.35 and 0.47 for perceived contrast function of sharpness, and between 0.39 and 0.58 for perceived sharpness function of contrast. The correlation is weaker in the lab data but still statistically significant for the materials with medium and high sharpness and contrast (p-value < 0.01), with a correlation that varies between 0.11 and 0.32 for these materials. This correlation differs from the observation in prior work that the c and d parameters are perceptually independent. Further studies are needed to understand the dimensions of gloss.

The second deviation that we observed occurs in the direction normal to the iso-gloss contours and makes materials in stylized images appear shinier or more diffuse, compressing the range of perceivable materials. These results confirm our initial hypothesis that stylization affects the perception of gloss. In particular, we measure a reduction of shininess in painterly rendering as large brush strokes alter the small highlights that contribute to the appearance of shininess [Berzhanskaya et al. 2002] (hypothesis **H1**). In addition, semi-transparent strokes blend colors between neighboring strokes, which reduces local contrast and makes materials appear even more diffuse (hypothesis **H3**). Opaque strokes make diffuse materials appear slightly shinier (hypothesis **H2**) but this effect is mitigated by semi-transparent strokes.

The data we have collected also confirm our hypothesis that cartoon rendering makes materials appear shinier because of the sharpening of diffuse color variations (hypothesis **H5**). However, our data do not support our hypothesis that bump mapping would increase shininess (hypothesis **H4**). We observe instead an opposite trend as

bump mapping makes shiny materials appear as even more diffuse than with the painterly style without bump textures. An explanation for this perceptual effect is that, while subjects are able to distinguish the bump map specularities from the object specularities, the bristles texture masks the brush strokes' edges and alters the perception of sharp high contrast color variations over the object. Bump mapping can be seen as noise that corrupts the high frequencies of the original signal.

Finally, the results of our blur experiment show that painterly rendering with large brush strokes offers a better depiction of materials than blur. While blur averages the color values over the pixels, painterly rendering selects a subset of these colors and preserves more of the original contrast. In the context of image abstraction, that simplify image content by removing spurious details, our results suggest that painterly rendering is more effective than blur to preserve material appearance through the simplification.

The data we collected on the Mechanical Turk suggest most of the trends we observe in the lab data but contain a stronger compression along the contrast dimension for all stylizations. We explain this increase in accuracy for perceived contrast in the lab by the fact that the lab environment (calibrated monitor, dark room) offers much more contrast than the one experienced by the Mechanical Turk users on typical desktop and laptop computers. We also observed more variance in the Mechanical Turk data than in the lab data. However, although we have collected approximately the same number of responses in both setups, the variance on each platform can have different interpretations. Since the lab data of each style contains the responses of 3 subjects with 10 repeats, the observed variance may be mostly due to intra-subject variability. In contrast, the Mechanical Turk data corresponds to the responses of 15 to 30 subjects without repeat, so the variance is more representative of the inter-subject as well as viewing condition variability.

7. APPLICATIONS

7.1 Predicting Perception From Material and Style Parameters

Our results show that different stylizations compress the space of perceived materials in different ways. As an application of this result, we estimate the function that maps the gloss space to the compressed perceived space. This function predicts how a given material will be perceived in a given style, and the inverse function gives the material parameters required to obtain a target perception.

We use linear regression to fit a mapping function on the subjects' settings. A linear model proved to be sufficient for a good fit of our data. We fit a different function for each style (the three painterly styles and the cartoon style) and each style parameter (brush size or sharpness of quantization). Each function is expressed as:

$$c' = f_c(c, d) = \alpha c + \beta d + \gamma \qquad (7)$$
$$d' = f_d(c, d) = \delta c + \epsilon d + \zeta \qquad (8)$$

where $\{c, d\}$ are the input material parameters, and $\{c', d'\}$ the predicted perceived material. The inverse of this function predicts which material $\{c, d\}$ produces a desired perception $\{c', d'\}$, and is expressed as:

$$c = f_c^{-1}(c', d') = \frac{\epsilon c' - \beta d' - \epsilon \gamma + \beta \zeta}{\delta \beta - \alpha \epsilon} \qquad (9)$$
$$d = f_d^{-1}(c', d') = \frac{\delta c' - \alpha d' - \delta \gamma + \alpha \zeta}{\delta \beta - \alpha \epsilon}. \qquad (10)$$
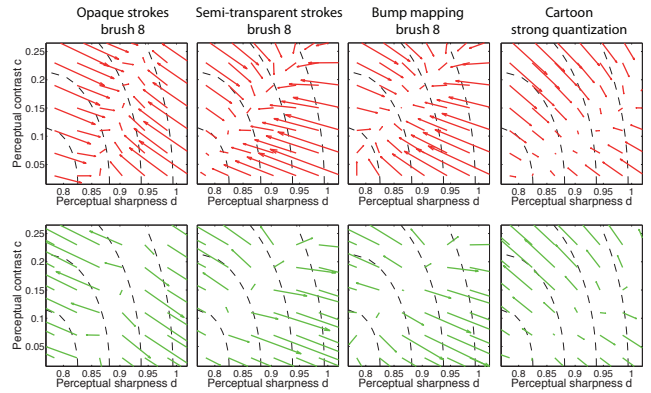
Fig. 7: *(First row) Mapping between the true and perceived materials. (Second row) Visualization of the inverse mapping, indicating the material parameters that are required to obtain a given material in gloss space. Note that the inverse mapping can potentially point to materials that exceed the physical limits of the BRDF model. This corresponds to materials that cannot be effectively depicted in a stylized image.*

We report the linear coefficients of our model computed from the lab data for painterly rendering with opaque strokes in Table II. We provide the coefficients of the other styles computed from the lab and Mechanical Turk data as supplemental materials. We interpolate between these coefficients for brush sizes that are not in our dataset. Figure 1 illustrates our prediction of perceived materials for various styles. For this figure we used the data collected on the Mechanical Turk that are more representative of the viewing conditions under which this paper is likely to be viewed.

	α	β	γ	δ	ϵ	ζ
No Brush	0.870	0.128	-0.097	0.069	0.892	0.094
Brush 4	0.807	0.318	-0.252	0.212	0.618	0.317
Brush 8	0.837	0.290	-0.230	0.235	0.494	0.424
Brush 16	0.722	0.285	-0.224	0.236	0.307	0.583

Table II. : *Mapping coefficients for the brush sizes in our dataset. Opaque strokes.*

The standard error of estimate of our model varies among styles between 0.032 and 0.039 for c and d and reflects the variance observed in the subjects' settings. Our model gives a coefficient of determination r^2 that varies between 0.5 and 0.8 for c and d on our lab data. The averaged L2 distance between the predictions of our model and the corresponding mean subjects' settings is equal to 0.01, which is much smaller than the standard deviation of subjects' settings.

Figure 7 visualizes the mapping estimated from the lab data for different styles, along with the corresponding inverse mapping. We observe two main directions of deviation in the estimated mapping, which reflects our earlier insights. A strong deviation occurs along iso-gloss contours, and a second deviation compresses the range of perceived materials across gloss values. Note that the inverse mapping can potentially lead to material parameters that cannot be represented with the BRDF model, indicating that these materials cannot be depicted effectively in a stylized image. For large brush strokes, the space of perceived materials collapses to a very small area of the gloss space, making the inverse mapping close to ill-posed especially along iso-gloss contours.

7.2 Enhancing Material Perception

Figure 8 (a) and (b) show the realistic and painterly rendering of two vases. The right vase is significantly more specular than the left one, yet the two materials are difficult to distinguish after stylization, as predicted by our mapping (c). We use our inverse mapping to estimate the material parameters that would produce the desired perception. The two materials point to values outside the range of materials that can be represented with the BRDF, so we clip the values on the boundary of the space. We use these parameters to exaggerate the contrast between the two materials so that the shiny vase can be distinguished from the diffuse vase after stylization (Figure 9(b)). Correcting material parameters according to style improves the rendition of materials in the stylized image. We used the Mechanical Turk data to generate these figures. This application is similar in spirit to the work of Vangorp and Dutré [2008] who correct material parameters to compensate for the influence of shape on gloss perception.

We have conducted a pilot study in the lab to validate our material exaggeration. We present subjects with a stylized image and two realistic images and instruct them to "Select the realistic image that looks most like the stylized image". We use two conditions to generate the images. In the first condition, one of the realistic image is used to generate the stylized image, while the second realistic image corresponds to our prediction of the materials perceived in the stylized image. We call this condition a "prediction". In the second condition we use the same two realistic images but we use our exaggerated materials to create the stylized image, so that the stylized image should look like the first realistic image. We call this condition a "exaggeration".

The images showed the same vase as in in Figure 1. We used 4 representative materials (low and high sharpness and contrast) and the four styles (three painterly styles with brush 8 and 16 and a cartoon style with strong quantization). Seven subjects participated in the study, yielding 112 trials for each painterly style and 56 trials for the cartoon style. Half of these trials corresponds to the prediction condition while the other half corresponds to exaggerations. Table III reports the success rates for predictions and exaggerations, where we consider a trial to be a success when subjects select our prediction in the first condition and the original image in the exaggeration condition. When pooled across all the styles, the success rates and 5% significance p-values reveal that subjects only perform slightly better than chance (43% of error) when asked to choose between the original material and our prediction, which suggests that the two materials are plausible interpretations of the stylized image. Our exaggeration is effective in helping users distinguish these materials in 65% of the trials. When broken across styles, the success rate of the exaggeration is only statistically significant for the painterly styles with semi-transparent strokes or bump mapping.

7.3 Compensating for Response Bias

As discussed in Section 6.2, the deviation along iso-gloss contours suggests a response bias as subjects tend to avoid the end of the scales when they are uncertain about the task. This bias results in a non-null mapping even for the realistic images. We propose to compensate for this potential bias by subtracting the contribution of the mapping fitted on the realistic settings from the other mappings. Denoting $\{f_c^0, f_d^0\}$ the mapping for the realistic case, we express

	All	Opaque	Semi-transp.	Bump	Cartoon
prediction	**0.43**	0.45	0.43	0.46	0.36
p-value	0.037	0.252	0.175	0.344	0.09
exaggeration	**0.65**	0.59	**0.79**	**0.64**	0.54
p-value	0.0000	0.114	0.0000	0.022	0.425

Table III. : *Success rate of our prediction and exaggeration. When pooled across the 4 styles, the results of this study suggest that subjects hardly make the distinction between the original material and our prediction (first row) and that our exaggeration helps to disambiguate these two materials in 65% of the cases (third row). We highlight values under the 5% significance level in bold (p-value < 0.05).*
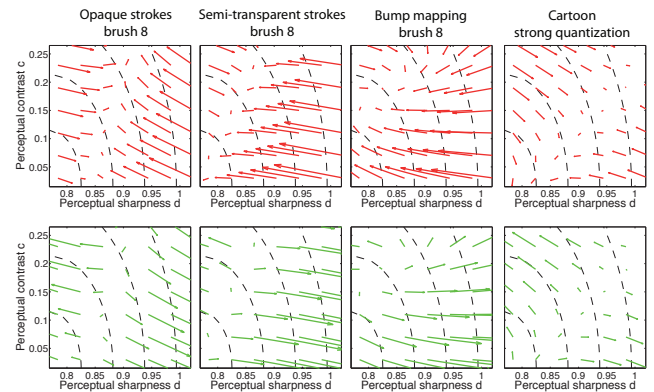

Fig. 10: *We subtract the contribution of the realistic settings to compensate for the response bias along iso-gloss contours. The differences in compression induced by each style are more visible in these compensated mappings.*

the compensated mapping $\{\hat{c}', \hat{d}'\}$ as

$$\hat{c}' = c + (f_c(c, d) - f_c^0(c, d)) \qquad (11)$$

$$\hat{d}' = d + (f_d(c, d) - f_d^0(c, d)). \qquad (12)$$

Figure 10 visualizes the compensated mapping fitted on the lab data for every style. The compensation effectively removes the deviation along the iso-gloss contours and helps distinguish the individual effect of each style. The overall compression appears stronger along the sharpness dimension than along the contrast dimension. While opaque strokes equally compress diffuse and shiny materials toward the center of the space, semi-transparent strokes and bump mapping mainly make the shiny materials appear more diffuse. In contrast, cartoon rendering increases shininess of high-contrast low-sharpness materials but reduces shininess at low-contrast.

8. DISCUSSION AND FUTURE WORK

Perception of Textured Objects, Dynamic Scenes and Contextual Cues. Our work is in the line of previous studies on the perception of materials in static images [Pellacini et al. 2000; Fleming et al. 2003; Ramanarayanan et al. 2007; Vangorp et al. 2007]. Nonetheless, the study of material perception in dynamic scenes represents an interesting direction for future research, especially for NPR styles where temporal coherence is a critical issue. Like previous work we also did not consider the effect of texture mapping on the object. Texture patterns can be hard to decorrelate from lighting reflections in a static image and studying the perception of these combined effects represent another challenging research
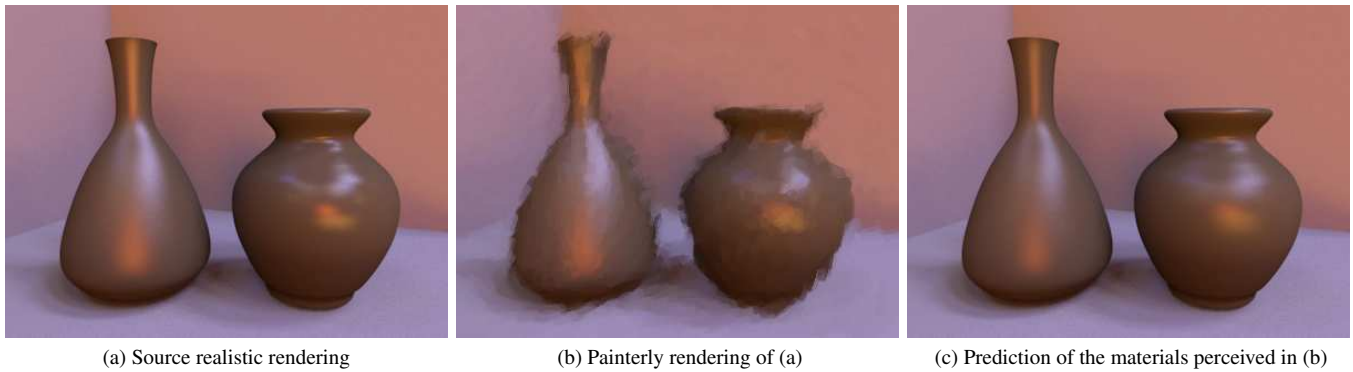
| (a) Source realistic rendering | (b) Painterly rendering of (a) | (c) Prediction of the materials perceived in (b) |

Fig. 8: *Prediction of the materials perceived in a painterly rendering. The right vase appears more diffuse in the painterly image than in the source realistic image.*
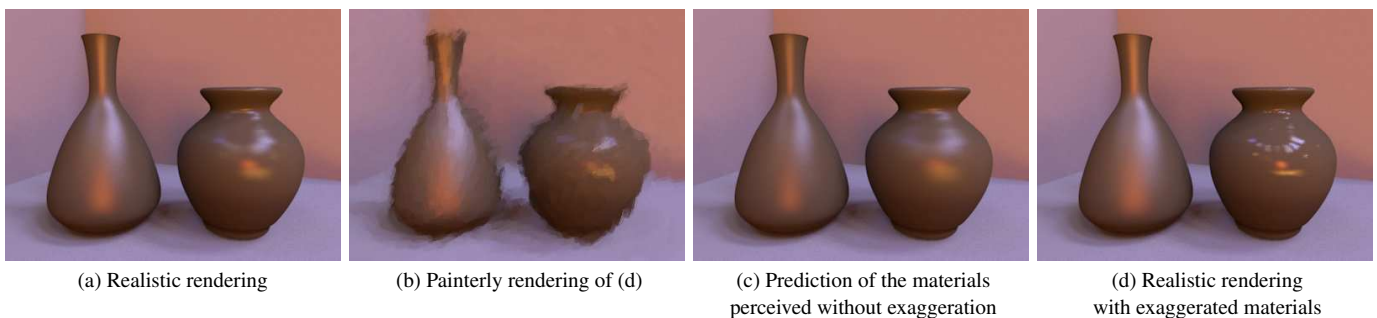

| (a) Realistic rendering | (b) Painterly rendering of (d) | (c) Prediction of the materials perceived without exaggeration | (d) Realistic rendering with exaggerated materials |

Fig. 9: *The inverse mapping exaggerates the materials in (a) to produce (d). The exaggeration adds contrast between the two materials to compensate for the compression of the space of perceived materials. As a result, the exaggerated painterly rendering (b) looks closer to the target image (a) than to the one perceived without exaggeration (c).*

direction for realistic and stylized rendering. Finally we chose to show the object over a black background to avoid contextual influence. However, most of our styles leave this background unaffected which may give the impression that the stylization modifies the object surface rather than the entire image. An additional study is needed to assess if the presence of background helps subjects to distinguish the effect of the stylization from the material variations.

Evaluating the Influence of Other Styles. This paper presents the first evaluation of material perception in NPR and focuses on painterly and cartoon styles. While our results do not directly apply to other styles, our experimental setup can be used to study many other stylization techniques. Existing work points to several guidelines regarding depiction of material such as the use of dark and light bands to depict metallic surfaces in technical illustrations [Gooch et al. 1998] or straight lines to depict glass in pen-and-ink [Winkenbach and Salesin 1996]. None of these guidelines have been evaluated formally and a better comprehension of their effects could lead to more effective stylization algorithms. Since black and white styles such as pen-and-ink and stippling use strokes to depict both texture and tone, a first challenge would be to evaluate how people distinguish tone from texture in these images.

Perception of Gloss. Our results on realistic images reveal a correlation between contrast and sharpness variations for materials in the mid-gloss range. This observation differs from the perceptual independence of contrast and sharpness reported in previous work [Pellacini et al. 2000; Fleming et al. 2003]. Further research is needed

to better understand how people perceive the various dimensions of gloss in realistic and stylized images.

Improving Painterly Rendering. Our study shows that the range of perceived materials is compressed as stylization increases. An interesting future research direction is to design painterly rendering algorithms that better convey materials. The material exaggeration that we describe in Section 7 is a first step. We observed that large brush strokes can remove small specular highlights and make shiny materials appear more diffuse. A possible improvement would be to impose that the brush strokes are located on the local maxima and minima in the image to preserve these highlights.

Evaluation of Crowdsourcing for Perceptual Studies. Our work demonstrates that crowdsourcing is a viable solution to iteratively design perceptual studies. In addition to speed and low cost, crowdsourcing gives us access to a large population working under a variety of viewing conditions that are more representative of the conditions under which images are viewed on the web. This variety also leads to additional variance in the data compared to the results obtained in a lab under a controled environment. Additional research in the spirit of [Heer and Bostock 2010] is necessary to understand the nature and magnitude of this variance. In particular, we plan to investigate how one could estimate or control the viewing conditions from a distance in order to improve the interpretation of studies conducted via crowdsourcing.

9. CONCLUSION

In this paper we used non photorealistic rendering algorithms to study the perception of materials in stylized images. We have measured the compression of the range of gloss that can be depicted in painterly and cartoon rendering and used these data to estimate the function that maps a realistic gloss description to its perception in a painterly or cartoon image. Our study also reveals novel insights on gloss perception in realistic images as we have identified a dependence between the contrast and sharpness of gloss. Finally we describe the use of crowdsourcing to conduct our perceptual study. We believe that crowdsourcing can greatly facilitate and accelerate research similar to ours and we plan to investigate how to make crowdsourcing more controllable for computer graphics studies.

REFERENCES

BAIRD, J. AND NOMA, E. 1978. *Fundamentals of scaling and psychophysics*. Wiley series in behavior. Wiley.

BERZHANSKAYA, J., SWAMINATHAN, G., BECK, J., AND MINGOLLA, E. 2002. Highlights and surface gloss perception. *Journal of Vision 2*, 3.

COLE, F., SANIK, K., DECARLO, D., FINKELSTEIN, A., FUNKHOUSER, T., RUSINKIEWICZ, S., AND SINGH, M. 2009. How well do line drawings depict shape? *ACM TOG (Proc. of SIGGRAPH 2009) 28*.

COOKE, H. L. 1967. *Painting Lessons from the Great Masters*. Watson-Guptill Publications.

CURTIS, C. J., ANDERSON, S. E., SEIMS, J. E., FLEISCHER, K. W., AND SALESIN, D. H. 1997. Computer-generated watercolor. *SIGGRAPH '97*.

DECARLO, D. AND SANTELLA, A. 2002. Stylization and abstraction of photographs. *ACM TOG (Proc. of SIGGRAPH 2002) 21*, 3, 769–776.

DROR, R. O., ADELSON, E. H., AND WILLSKY, A. S. 2001. Recognition of surface reflectance properties from a single image under unknown real-world illumination. In *Proc. of the IEEE Workshop on Identifying Objects Across Variations in Lighting*.

FERWERDA, J. A., PELLACINI, F., AND GREENBERG, D. P. 2001. A psychophysically-based model of surface gloss perception. In *Proc. of SPIE Human Vision and Electronic Imaging*.

FLEMING, R. W., DROR, R. O., AND ADELSON, E. H. 2001. How do humans determine reflectance properties under unknown illumination? *IEEE Workshop on Identifying Objects Across Variations in Lighting*.

FLEMING, R. W., DROR, R. O., AND ADELSON, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision 3*, 5, 347–368.

GOOCH, A., GOOCH, B., SHIRLEY, P., AND COHEN, E. 1998. A non-photorealistic lighting model for automatic technical illustration. *SIGGRAPH '98*, 447–452.

GOOCH, B., REINHARD, E., AND GOOCH, A. 2004. Human facial illustrations: Creation and psychophysical evaluation. *ACM TOG 23*, 1.

HAEBERLI, P. 1990. Paint by numbers: abstract image representations. *Computer Graphics (Proc. of SIGGRAPH '90)*, 207–214.

HAYS, J. AND ESSA, I. 2004. Image and video based painterly animation. In *NPAR '04*. 113–120.

HEER, J. AND BOSTOCK, M. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *ACM CHI '10: Proc. of the 28th int. conf. on Human factors in computing systems*.

HERTZMANN, A. 1998. Painterly rendering with curved brush strokes of multiple sizes. *SIGGRAPH '98*, 453–460.

HERTZMANN, A. 2002. Fast paint texture. In *NPAR '02*.

JOHNSON, C. 1992. *Creating Textures in Watercolor*. North Light Books.

KANG, H., LEE, S., AND CHUI, C. K. 2007. Coherent line drawing. In *NPAR '07*. 43–50.

KOZLOWSKI, O. AND KAUTZ, J. 2007. Is accurate occlusion of glossy reflections necessary? In *APGV '07*. 91–98.

KŘIVÁNEK, J., FERWERDA, J. A., AND BALA, K. 2010. Effects of global illumination approximations on material appearance. *ACM TOG (Proc. of SIGGRAPH 2010) 29*, 3.

LITWINOWICZ, P. 1997. Processing images and video for an impressionist effect. *SIGGRAPH '97*, 407–414.

MEIER, B. J. 1996. Painterly rendering for animation. *SIGGRAPH '96*.

NISHIDA, S. AND SHINYA, M. 1998. Use of image-based information in judgments of surface-reflectance properties. *Journal of the Optical Society of America A: Optics, Image Science and Vision 15*, 12.

OTT, J. AND KUSENO, Y. 2005. *Let's Draw Manga: Using Color*. Digital Manga Publishing.

PELLACINI, F., FERWERDA, J. A., AND GREENBERG, D. P. 2000. Toward a psychophysically-based light reflection model for image synthesis. *SIGGRAPH '00*, 55–64.

PHARR, M. AND HUMPHREYS, G. 2004. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann Publishers Inc.

RAMANARAYANAN, G., FERWERDA, J., WALTER, B., AND BALA, K. 2007. Visual equivalence: towards a new standard for image fidelity. *ACM TOG (Proc. of SIGGRAPH 2007) 26*, 3.

SMITH, K., SOLER, C., LUFT, T., DEUSSEN, O., AND THOLLOT, J. 2010. Automatic Pen-and-Ink Illustration of Tone, Gloss, And Texture. Research Report RR-7194, INRIA.

VANGORP, P. AND DUTRÉ, P. 2008. Shape-dependent gloss correction. In *APGV '08*. 123–130.

VANGORP, P., LAURIJSSEN, J., AND DUTRÉ, P. 2007. The influence of shape on the perception of material reflectance. *ACM TOG (Proc. of SIGGRAPH 2007) 26*, 3, 77.

WALLRAVEN, C., BÜLTHOFF, H. H., CUNNINGHAM, D. W., FISCHER, J., AND BARTZ, D. 2007. Evaluation of real-world and computer-generated stylized facial expressions. *ACM Trans. Applied Perception 4*, 3, 16.

WARD, G. J. 1992. Measuring and modeling anisotropic reflection. *SIGGRAPH '92 26*, 2, 265–272.

WILLS, J., AGARWAL, S., KRIEGMAN, D., AND BELONGIE, S. 2009. Toward a perceptual space for gloss. *ACM TOG 28*, 4.

WINKENBACH, G. AND SALESIN, D. H. 1996. Rendering parametric surfaces in pen and ink. *SIGGRAPH '96*, 469–476.

WINNEMÖELLER, H., OLSEN, S. C., AND GOOCH, B. 2006. Real-time video abstraction. *ACM TOG (Proc. of SIGGRAPH 2006) 25*, 3.

WINNEMÖLLER, H., FENG, D., GOOCH, B., AND SUZUKI, S. 2007. Using npr to evaluate perceptual shape cues in dynamic environments. In *NPAR '07*. 85–92.

XUE, S., CHEN, X., DORSEY, J., AND RUSHMEIER, H. E. 2010. Printed patterns for enhanced shape perception of papercraft models. *Computer Graphics Forum 29*, 2, 625–634.

ZENG, K., ZHAO, M., XIONG, C., AND ZHU, S.-C. 2009. From image parsing to painterly rendering. *ACM TOG 29*, 1.
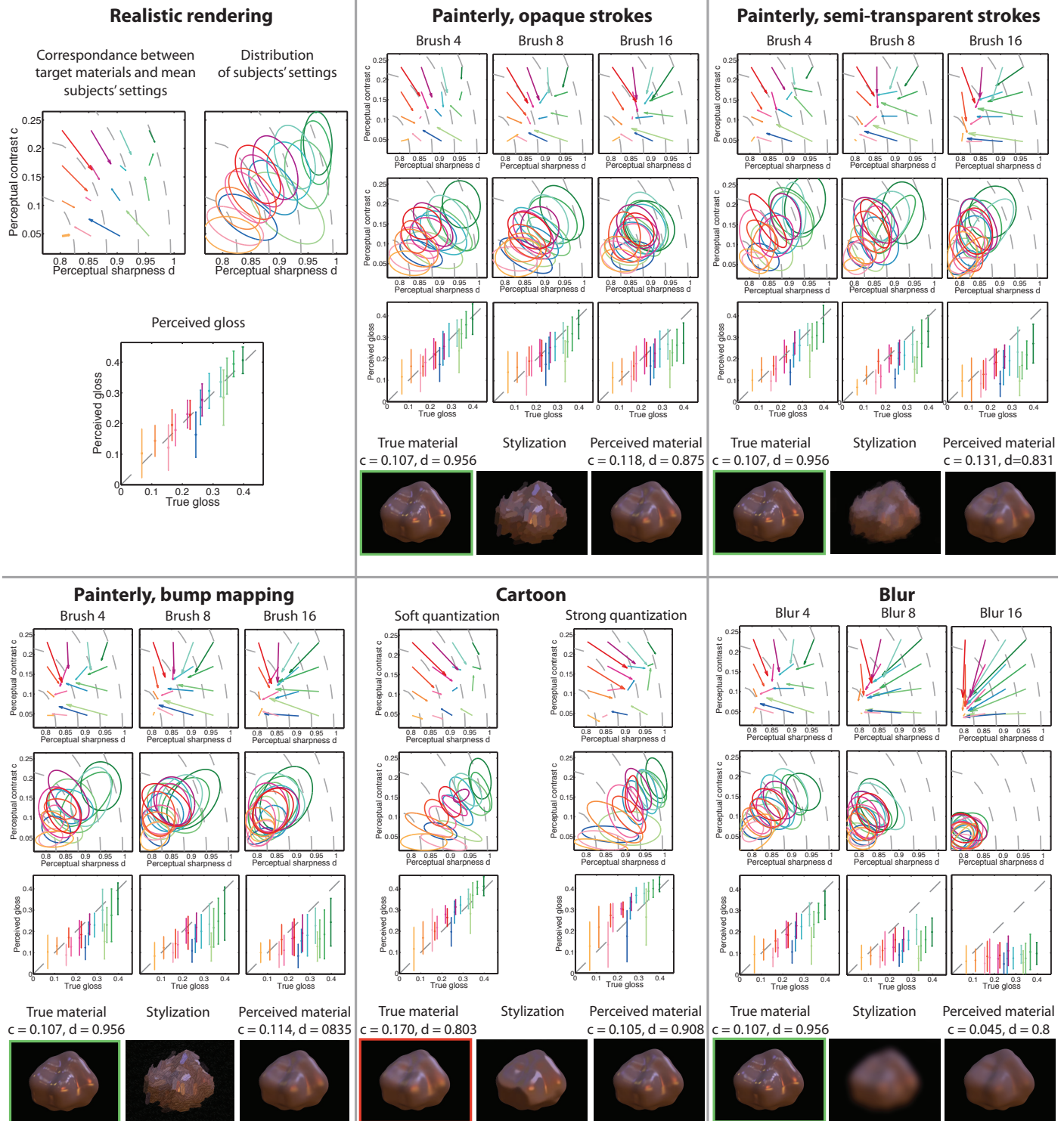
Fig. 11: *Summary of the data collected in our Mechanical Turk experiment. Arrows point from each reference material in the target set to the mean of the subjects' settings for this material. Ellipses represent the covariance of subjects' settings for each material. We use 16 different colors to distinguished the 16 target materials. The origin of the frame corresponds to the most diffuse material covered by the interface sliders ($c = 0.015$ and $d = 0.769$), and dashed curves represent iso-gloss contours with respect to this origin. Bar plots visualize the mean and standard deviation of subjects' settings projected along the gloss dimension, and a dashed line indicates the ideal settings (perceived gloss = true gloss). We observe two forms of deviation in the subject settings. The first deviation occurs along the iso-gloss contours and is even present in the realistic case. The second deviation occurs across the iso-gloss contours and compresses the range of perceived gloss in stylized images. However, the amplitude and direction of this second deviation differ among styles.*
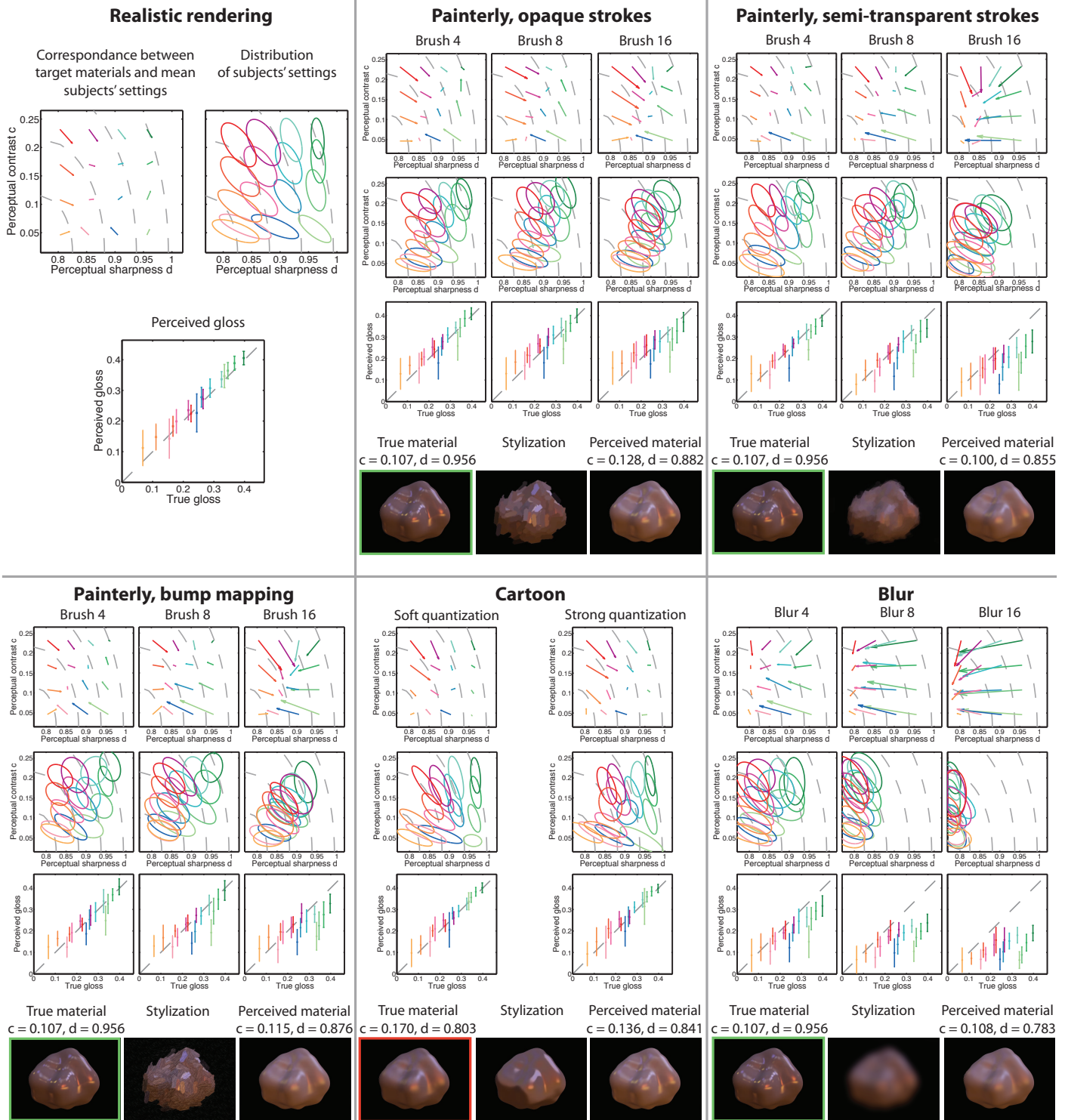
Fig. 12: *Summary of the data collected in our lab experiment. Arrows point from each reference material in the target set to the mean of the subjects' settings for this material. Ellipses represent the covariance of subjects' settings for each material. We use* 16 *different colors to distinguished the* 16 *target materials. The origin of the frame corresponds to the most diffuse material covered by the interface sliders (*$c = 0.015$ *and* $d = 0.769$*), and dashed curves represent iso-gloss contours with respect to this origin. Bar plots visualize the mean and standard deviation of subjects' settings projected along the gloss dimension, and a dashed line indicates the ideal settings (perceived gloss = true gloss). We observe in the lab data similar deviations as in the Mechanical Turk data. However, subjects perform more accurate settings of contrast in the lab.*