Computational Aesthetics in Graphics, Visualization, and Imaging (2007)
D. W. Cunningham, G. Meyer, L. Neumann (Editors)

Perceptual reparameterization of material properties

D. W. Cunningham[1], C.Wallraven[2], R.W. Fleming[2], and W. Strasser[1]

[1]WSI/GRIS, University of Tuebingen, Germany
[2]Max Planck Institute for Biological Cybernetics, Germany

Abstract

The recent increase in both the range and the subtlety of computer graphics techniques has greatly expanded the possibilities for synthesizing images. In many cases, however, the relationship between the parameters of an algorithm and the resulting perceptual effect is not straightforward. Since the ability to produce specific, intended effects is a natural pre-requisite for many scientific and artistic endeavors, this is a strong drawback. Here, we demonstrate a generalized method for determining both the qualitative and quantitative mapping between parameters and perception. Multidimensional Scaling extracts the metric structure of perceived similarity between the objects, as well as the transformation between similarity space and parameter space. Factor analysis of semantic differentials is used to determine the aesthetic structure of the stimulus set. Jointly, the results provide a description of how specific parameter changes can produce specific semantic changes. The method is demonstrated using two datasets. The first dataset consisted of glossy objects, which turned out to have a 2D similarity space and five primary semantic factors. The second dataset, transparent objects, can be described with a non-linear, 1D similarity map and six semantic factors. In both cases, roughly half of the factors represented aesthetic aspects of the stimuli, and half the low-level material properties. Perceptual reparameterization of computer graphics algorithms (such as those dealing with the representation of surface properties) offers the potential to improve their accessibility. This will not only allow easier generation of specific effects, but also enable more intuitive exploration of different image properties.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation I.3.6 [Computer Graphics]: Methodology and Techniques J.4 [Computer Application]: Social and Behavioural Sciences—Psychology

1. Introduction

Experts from a number of fields often strive to create images that contain very specific effects. The tremendous increase in the power, flexibility, range, and fine control of computer graphics techniques has made them almost indispensable. Unfortunately, the perceptual effect of changing any one parameter in these algorithms, let alone combinations of parameters, is often not intuitive or easy to predict. To make matters worse, the effect of some parameters is non-monotonic: while increases in a parameter value in one range may increase a given effect, increasing the same parameter in a different range may decrease that effect – or produce an entirely different effect. This can make it very difficult to intentionally create a specific effect, hindering the wide-spread application of these techniques.

In order to improve the accessibility of these algorithms, one must first understand the perceptual consequences of the different parameters at both a quantitative and qualitative level. Early work on color perception and color reproduction struggled with a similar problem: How can one describe, quantify, and predict color appearance, and what parameter space can allow us to intuitively manipulate colors [JW63, WS82]. Today, most color researchers and computer graphics experts have a fairly clear notion of the appearance of a specific aperture color based on its CIE coordinates. This mapping was constructed, in part, by examining the relationship between several specific colors and the wavelength or combination of wavelengths needed to produce them.

Here, we describe a method for determining the mapping between parameter changes and their perceptual con-

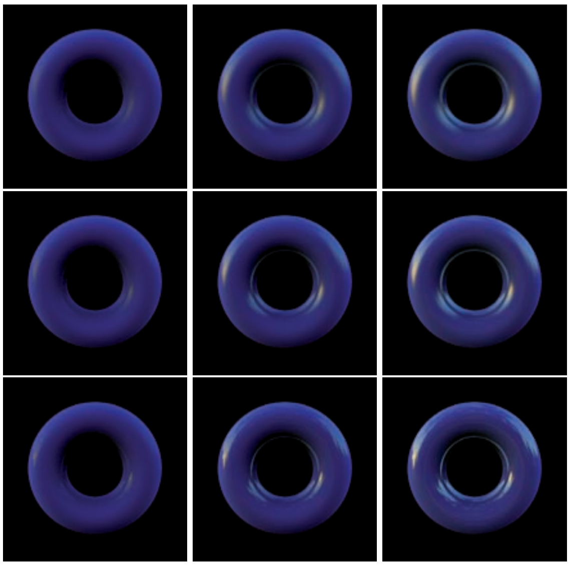

Figure 1: *The extremes and middle values for the 25 glossy objects*

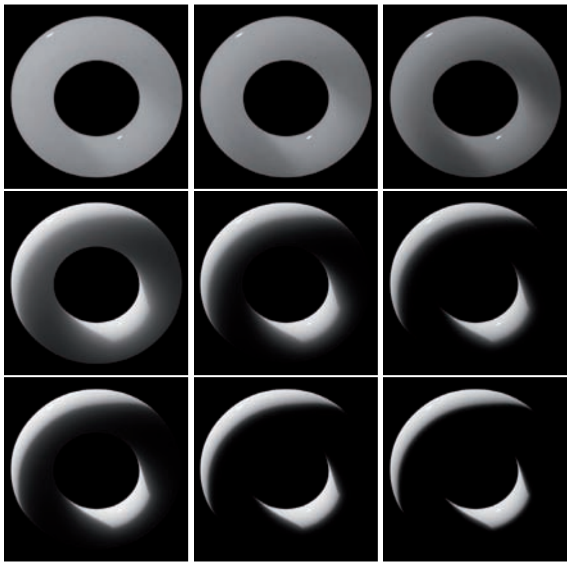

Figure 2: *The extremes and middle values of the 25 translucent objects*

sequences. The approach is similar to that outlined by [PFG00], and can be considered a generalization of their technique. In both approaches, the quantitive structure of the perceived similarity of objects within a given parameter range is recovered using Multidimensional Scaling (MDS). In MDS, the pairwise similarity ratings are converted into a, generally Euclidean, spatial representation such that the more similar objects are, the closer they are located in a multiple dimensional space. This provides an intuitive understanding of the perceptual distances between the objects, and the parameter values that create those effects. Thus, one should be able to generate specific objects in neighboring perceptual space by using the tested objects as landmarks. MDS does not, however, tell us anything about the semantic nature of this space.

The primary difference for this step between the present method and that outlined by [PFG00], lies in the nature of the similarity task. [PFG00] asked participants to rate the similarity of the objects based specifically on apparent surface gloss. In the method presented here, the participants are asked for general similarity judgments, allowing them to determine their own criteria.

The second step of the current method represents the largest change from [PFG00]'s approach. In this step, semantic differentials are used to obtain ratings of the objects along a number of scales. The data are then analyzed with factor analysis techniques to determine the minimum number of underlying factors that can explain the pattern of responses. This step determines the qualitative structure of many different perceptual and aesthetic aspects of the objects. In contrast to this, [PFG00] only had participants rate apparent surface gloss. By using a number of different scales, the present method can provide a more complete picture of the sematic nature of the objects. It is important to note that all scales range from one to seven, and that participants expand or contract their ratings to utilize much of the scale. Thus, similar changes in ratings for two different scales do not necessarily represent a similar magnitude perceptual change.

In a final step, the recovered semantic factors are related to the recovered similarity dimensions to determine their perceptual meaning. The result is a form of "colorimetric" map between exact parameter changes and the expected perceptual, semantic, or aesthetic results. This map can then be used to re-parameterize the underlying parameter space. Moreover, the entire procedure provides a novel method for examining image quality and fidelity.

2. Methods

2.1. Stimulus Generation

Two datasets were used, each of which contained 25 objects. The first set (see Figure 1) contained objects that were generated in RADIANCE using the isotropic version of the Ward reflectance model (i.e., the RADIANCE "plastic" material). The sum of diffuse and specular reflections were constrained not to exceed 1 (i.e., materials could not reflect more light than was incident on them). The Lambertian component was held constant at [0.1, 0.1, 0.3]. The parameters controlling the magnitude (rho_s) and spread (α) of the specular lobe were specified using [PFG00]'s perceptually-uniform re-parameterization (called 'c' and 'd' respectively). The magnitude parameter (Pellacini parameter 'c') ranged linearly from 0 to 0.19. The spread parameter (Pellacini parameter 'd') ranged linearly from 0.9 to 1.0

The second dataset (see Figure 2) consisted of objects that were generated in DALI using the Jensen and Buhler diffusion approximation for evaluating the BSSRDF [JB02]. The Refractive Index was set to 1.5, and the Heeney-Greenstein phase function was isotropic (i.e., g = 1). The scattering coefficient (K_s) ranged non-linearly from 0.8 to 150. The absorption coefficient (K_a) ranged non-linearly from 0.00001 to 0.8.

2.2. Psychophysical Methods

Twenty individuals participated in the experiment in return for financial compensation at standard rates. Half of the participants saw the glossy images, and the other half saw the translucent images. Each experimental session was divided into two parts: 1) Similarity ratings and 2) Sematic differentials. Half of the subjects in each group performed all the similarity ratings before starting the sematic differentials. The other half of the participants performed the two tasks in the opposite order. In all conditions, the participants sat at a distance of approximately 0.5 meters from the computer screen, and the images subtended approximately 15 by 15 degrees of visual angle. The experiment was performed in a darkened room, and the objects were presented on a black background.

2.2.1. Similarity Ratings and Multidimensional Scaling

For the similarity task, two images were presented side-by-side on the screen and participants were asked to rate the similarity of the two objects. They were asked to use a 7-point scale, with 1 meaning "low", and 7 meaning "high". After an initial 1 second presentation phase, the participants were free to respond. The images remained on screen until they responded. Each of the possible pairings of 25 images was presented 3 times, for a total of 975 trials. This portion of the experiment took approximately 45 minutes.

The ratings were then analyzed with Multidimensional Scaling (MDS). MDS was first introduced by Richardson [Ric38], and is a data reduction technique that is similar to PCA. The technique was further extended by Torgerson [Tor52] and subsequently has been used in a large range of application domains to explore an impressive variety of stimulus classes [SRY81].

There are many variants of MDS. We used the ALSCAL version, which uses the distance metric:

$$d_{ij} = \sqrt{\sum (x_{ia} - x_{ja})^2} \tag{1}$$

where x_{ia} and x_{ja} specify the positions of points i and j, respectively, on dimension a. It is important to note that the MDS model assumes that the appropriate metric for the psychological similarity space is Euclidean. If this assumption holds true, one expects low stress values for the overall MDS solution. Although establishing a threshold for acceptable values of stress is notoriously controversial, Monte Carlo studies suggest that stress values below 0.2 are indicative of an output configuration which provides a good fit to the similarity data [CC01]

2.2.2. Semantic Differentials and Factor Analysis

The semantic differential task was first introduced by Osgood [OST57]. Traditionally, participants are presented with some stimulus (e.g., an image, a scenario, a concept) and are asked to rate various properties of it using 7-point scales. Each property is represented with a pair of terms that are opposites, such as "good-bad" or "strong-weak". The sematic differential technique has been extensively researched in a number of application domains, from marketing research [Min61] to determining the aesthetic dimensions of baroque music performances [SF06]. It has been shown that three orthogonal dimensions serve to succinctly describe most of attitude space. The dimensions are often referred to as Evaluation (e.g., good-bad, heavy-light), Potency (e.g., strong-weak, powerful-powerless), and Activity (e.g., fast-slow, alive-dead). Most of the paired opposites scales that are traditionally used correlate very well with one of these dimensions, although some scales do correlate with more than one dimension. To improve the reliability of the measurements, each of the three dimensions is measured more than once using similar (but not identical) scales. To assess the aesthetic nature of the datasets, we used scales that are derived from the Evaluative and Potency dimensions. Since we were also interested in specific material properties, we added a few scales that are not traditionally used. More specifically, we used the following scales: unpleasant/pleasant, boring/interesting, ugly/beautiful, strange/normal, emotional/calm, dark/bright, transparent/opaque, shiny/matt, expensive/inexpensive, aesthetic/not aesthetic, hard/soft, 2D/3D, white/black, appealing/unappealing, artistic/not artistic. Finally, as a double check to see if the participants were using the scales in a meaningful manner, we also added a "like-dislike" evaluation scale using the sentence pair: "You want to drink a beer with it" - "You want to pour a beer over it". The 25 objects were presented one at a time, in random order. Since each of the 10 participants rated each of the 25 objects, we have 250 observations for each of the 16 scales. This portion of the experiment took approximately 20 minutes.

The resulting ratings were analyzed using factor analysis, a set of data reduction techniques which is also related to PCA and was originally introduced by Spearman [Spe04]. Briefly, factor analysis looks at the correlation matrix between the different variables and tries to determine whether the pattern of correlations can be explained by a smaller number of dimensions (factors). There are a number of methods for determining how many factors should be extracted. The most commonly used criterion, which was suggested by

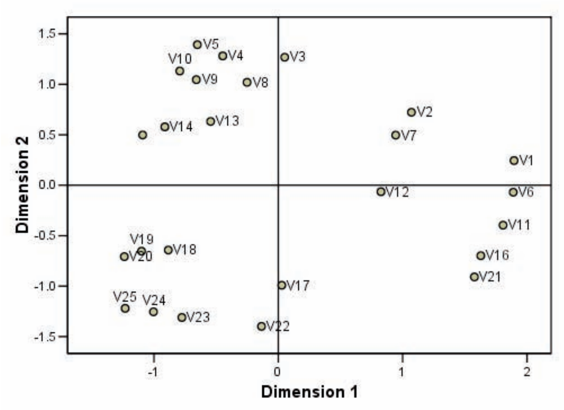

Figure 3: *The 2D similarity space for the glossy objects.*

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative%
1	4.77	29.8	29.8
2	1.74	10.9	40.7
3	1.55	9.7	50.3
4	1.15	7.2	57.5
5	1.09	6.8	64.3

Table 1: *The eigenvalues and variance explained for the glossy objects. Note that only factors whose eigenvalues are greater than one are listed.*

Kaiser [Kai60], is to use only those factors whose eigenvalue is greater than 1.

In general, factor analysis assumes that the pattern of variance can be explained by a linear combination of functions:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + ...a_{im}F_m + e_i \qquad (2)$$

where each X_i represents one of the i different scales. Each of the F_m terms represents some unknown function of one of the m factors. These functions are assumed to be independent. The e_i terms stand for the remaining error and the a_{ij} terms are the factor loadings. After obtaining the underlying factors, they are rotated to better improve their interpretability. Here, we use the most common rotation, varimax.

3. Results

Overall for the glossy objects, there were two dimensions in the similarity space, and these dimensions seem to correspond to those found by [PFG00]. For the transparent objects, there appears to be really only one dimension. This dimension controls the apparent shininess and opacity. The results are discussed in more detail below.

3.1. MDS for surface gloss

The stress for a two dimensional solution is 0.21. The data from one subject were eliminated due to failure to follow instructions. As can be seen in Figure 3, one of the obtained dimensions is mostly uniform, but the other is not. Since the parameter values were chosen based on [PFG00]'s work, the whole space should have been uniform if the participants were performing the same judgments as [PFG00]'s participants. The non-uniformity found here, then, suggest that when performing a general similarity judgment on this dataset, people base their decisions on more than just apparent surface gloss. The nature of some of these additional factors can be determined from the Semantic Differential ratings.

	Factors				
	1	2	3	4	5
Pleasant	**-0.74**	0.20	-0.06	-0.19	0.001
Interesting	**-0.57**	-0.22	0.12	-0.45	0.22
Beautiful	**-0.79**	0.02	0.04	-0.07	0.182
Normal	0.18	0.12	**0.57**	-0.23	-0.23
Calm	-0.01	0.34	**0.46**	0.20	0.006
Bright	-0.21	-0.08	0.007	**-0.62**	0.009
Opaque	-0.02	-0.03	**0.55**	0.38	0.09
Matt	**0.48**	0.33	0.12	0.15	0.08
Inexpensive	0.37	0.33	**0.43**	0.18	-0.41
Anaesthetic	**0.62**	0.50	0.23	-0.04	0.01
Soft	0.04	**0.44**	0.04	0.03	-0.12
3D	0.007	-0.11	-0.04	0.07	**0.62**
Black	0.07	0.03	0.12	**0.47**	0.06
Unappealing	**0.62**	0.32	0.06	0.08	0.111
Not artistic	0.41	**0.49**	0.25	0.07	-0.19
Pour	**0.43**	0.28	0.15	0.12	-0.03

Table 2: *Rotated Factor loadings of a six dimensional model for the glossy objects. Note that the each scale label represents a value of 7 for that scale.*

3.2. Semantic differentials for surface gloss

As can be seen in Table 1, five factors have eigenvalues greater than one. Table 2 shows the factor loadings. The five factors are:

- Factor 1: pleasant, interesting, beautiful, matt, anaesthetic, unappealing, and pour
- Factor 2: soft and not artistic
- Factor 3: normal, calm, opaque, and inexpensive.
- Factor 4: bright and black
- Factor 5: 3D

The first factor seems to represent the aesthetic aspects of the stimuli, and is similar to the traditional Evaluative factor. The fourth and fifth factors reflect material or object properties. The second and third factors seem to be combinations of emotional and material properties. These factors seem to be similar to the traditional Potency factor.

At first glance, it might be a bit surprising that more than the traditional three factors were extracted. This can be explained, however, by the facts that traditional semantic differential research rarely, if ever, measures material properties while the present work focused strongly on them. In this

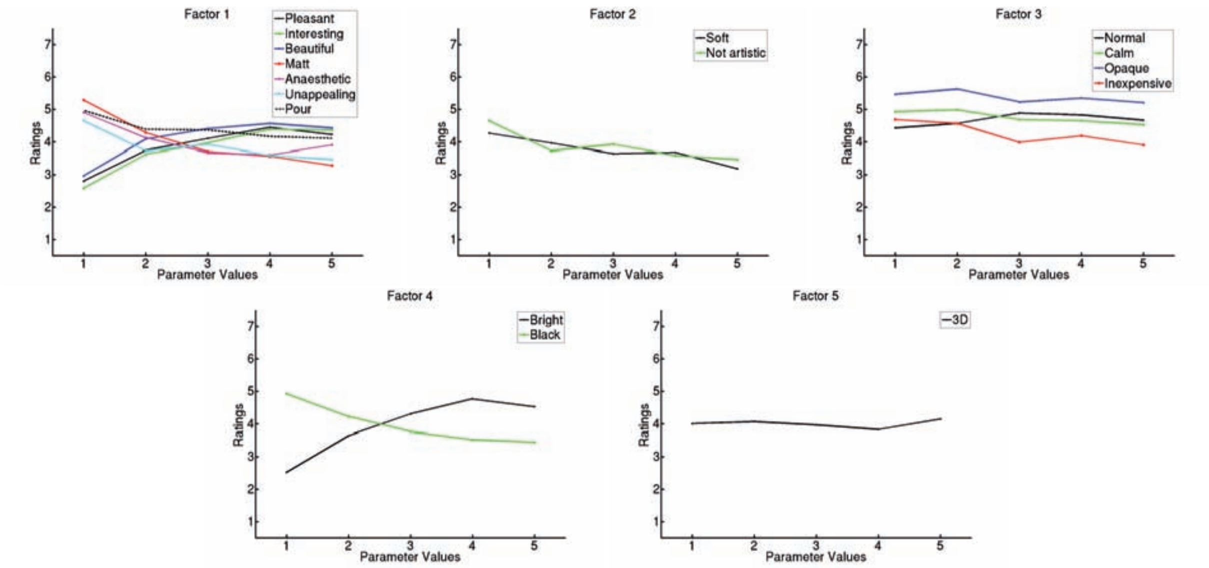

Figure 4: *Changes in the semantic differential ratings for the glossy objects as a function of parameter one, the spread of the specular lobe.*

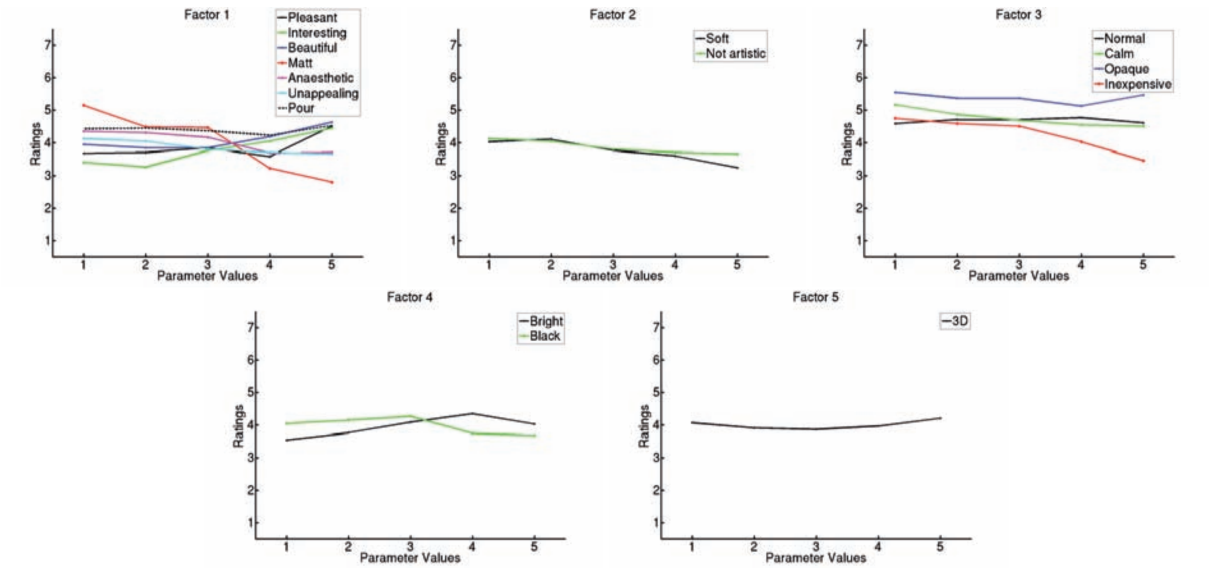

Figure 5: *Changes in the semantic differential ratings for the glossy objects as a function of parameter two, the magnitude of the specular lobe.*

regard, it is interesting to note that the first three factors, which account for just over 50% of the variance, nonetheless mostly represent aesthetic, emotional, or semantic properties.

Figure 4 plots the changes in the scales as a function of parameter one (i.e., spread of the specular lobe, seen as changes within in the rows of Figure 1, averaging across the columns). Based on their examination of their stimuli, [PFG00] suggested that this parameter affects the apparent sharpness or distinctness of the reflected image (Pellacini parameter d). The present semantic differential ratings are consistent with this: Increases in this parameter make the objects less glossy. Increases also tend to increase the Evaluative and Potency factors (i.e., make the object slightly harder, more

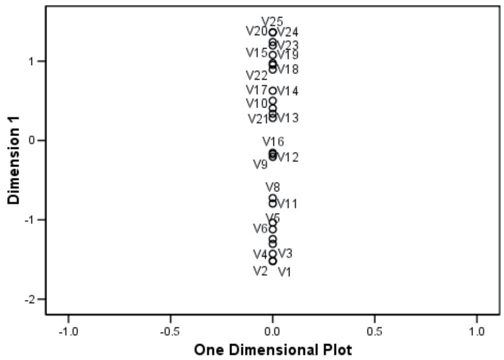

Figure 6: *The 1D similarity space for the transparent objects.*

interesting, slightly more beautiful, slightly more aesthetic and artistic, and more expensive).

Figure 5 plots the changes in the scales as a function of parameter two (the magnitude of the specular lobe, seen as changes within the columns). [PFG00] suggested that this parameter affects the apparent contrast of the reflected image (Pellacini parameter 'c'). Consistent with this, we found that this parameter strongly affected ratings of brightness and lightness (factor 4). Interestingly, increases in this parameter also strongly affected ratings of the objects "shininess". It also tended to strongly affect the Evaluative and Potency aspects of the objects (Factors 1 and 2). That is, increase in this parameter made the objects more beautiful, more pleasant, more interesting, more appealing, more aesthetic, harder, and more artistic.

3.3. MDS for translucency

The stress for a 1D solution is 0.23 which is just over the generally accepted limit (see Figure 6). The 2D stress was 0.18. As can be seen in Figure 7, however, the mapping really seems to be a nonlinear 1D manifold in 2D space, with parameter two (the absorption coefficient) controlling the majority of the differences in similarity.

3.4. Semantic Differentials for translucency

As can be seen in Table 3, six factors have eigenvalues greater than one. Table 4 shows the factor loadings. The six factors are:

- Factor 1: inexpensive, anaesthetic, artistic, and pour.
- Factor 2: pleasant, beautiful, normal, and calm
- Factor 3: interesting, 3D, and unappealing
- Factor 4: bright and black
- Factor 5: matt and soft
- Factor 6: opaque

While the first three factors are similar overall between the two datasets, there are some differences in the specific factor compositions. Interestingly, the first three factors, which also account for 50% of the variance here, contain almost no material properties (the sole possible exception is 3D). The difference in factor loadings between the two datasets might be the result of the small number of participants. It might also reflect the different nature of the datasets. This latter interpretation is supported by the fact that opacity, which was manipulated in the second dataset but not the first, has become its own factor in the second dataset.

Increases in either parameter decreases the brightness of the objects (see Figures 8 and 9). The other factors do not vary much with changes to parameter one (scattering coefficient), suggesting that none of the semantic dimensions measured were affected by this parameter. This is consistent with the mostly 1D nature of similarity space.

As for the second parameter (the absorption coefficient), two of the elements of Factor 1 (pour and not artistic) seem to show some co-variation, as do most scales of Factors 2 and 3. In other words, increases in the second parameter make the objects look less artistic, stranger, uglier, less pleasant, and slightly less appealing. At the same time, the objects become slightly more emotional, somewhat more interesting, and more 3D. Critically, increases in this parameter decrease apparent surface gloss and translucency, with the largest change being from the first parameter value to the second. In sum, parameter two changes the objects quite drastically, while parameter one barely affects the objects.

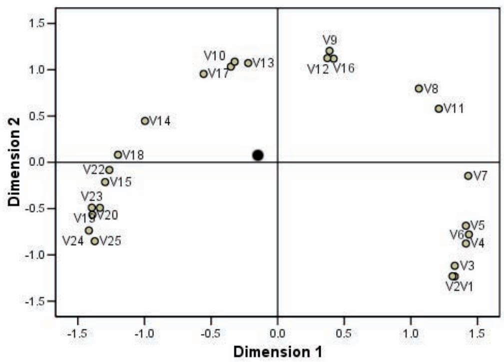

Figure 7: *The 2D similarity space for the transparent objects.*

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative%
1	3.73	23.3	23.3
2	2.64	16.5	39.8
3	1.64	10.2	50.0
4	1.37	8.6	58.6
5	1.23	7.7	66.3
6	1.03	6.4	72.8

Table 3: *The variance explained for the translucent objects. Note that only factors whose eigenvalues are greater than one are listed.*

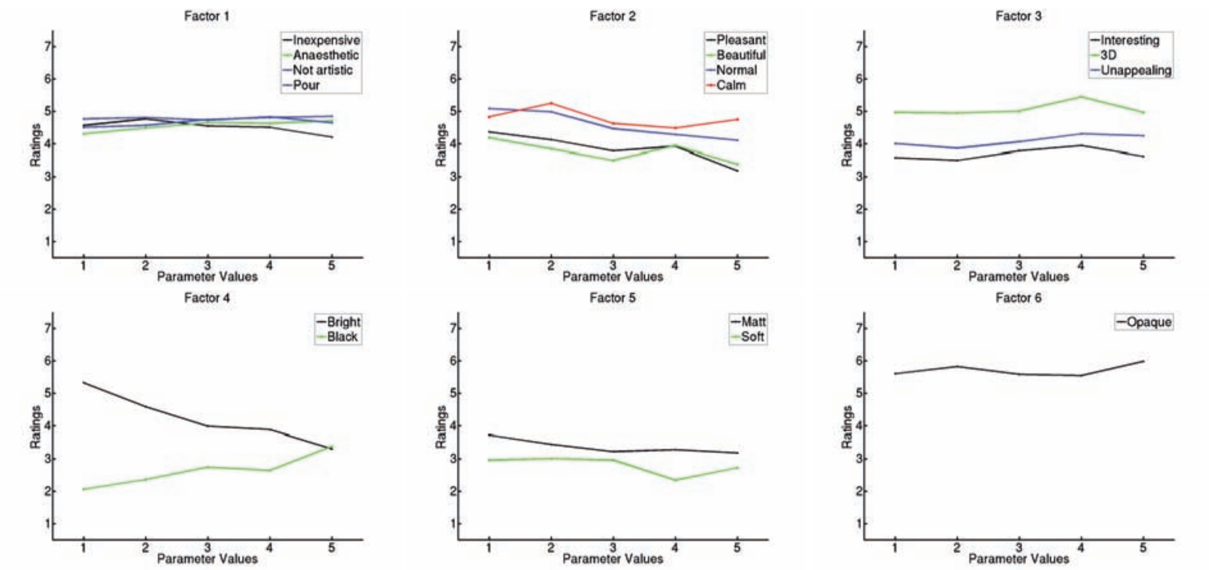

Figure 8: *Changes in the semantic differential ratings for the transparent objects as a function of changes in parameter one, the scattering coefficient.*

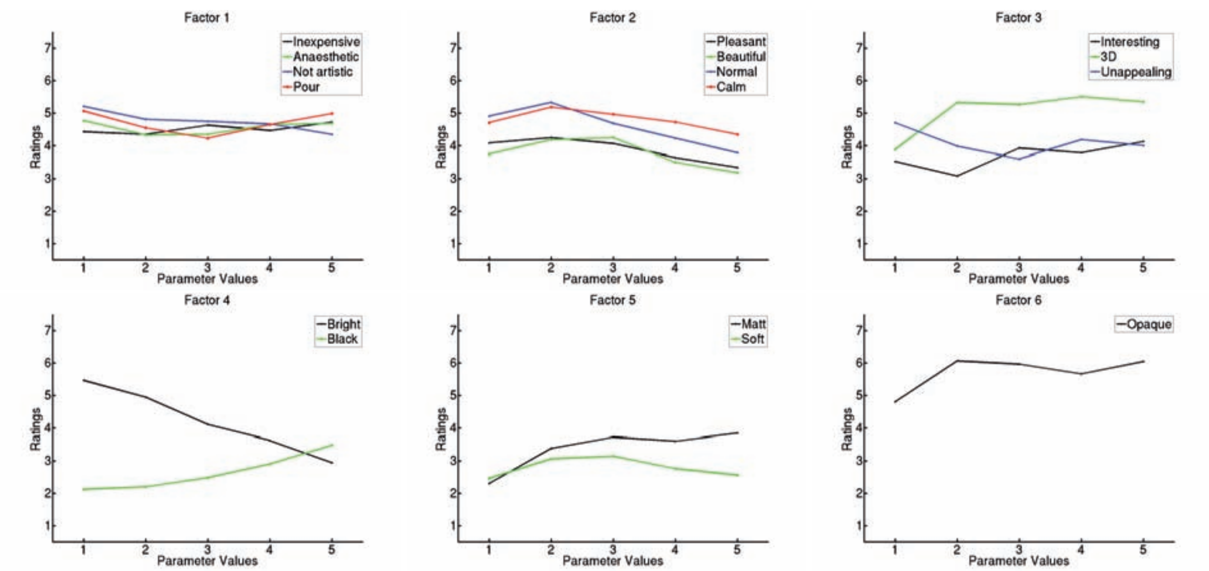

Figure 9: *Changes in the semantic differential ratings for the transparent objects as a function of changes in parameter two, the absorption coefficient.*

4. Discussion

The method presented here is a generalization of [PFG00]'s technique, and allows one to more completely determine the qualitative and quantitative mapping between the parameters of a computer graphics algorithm and its perceptual, semantic, and aesthetic effects. The method was demonstrated on two datasets. The first dataset, which varied the apparent surface gloss of an object, is similar to [PFG00]'s dataset, and the results are consistent with their results: The similarity space has two dimensions, which reflect changes in the apparent surface gloss and perceived brightness. The full semantic space of the objects seems to contain 5 factors, which equally reflect aesthetic and material properties. The two parameters also had specific effects on these factors.

	Factors					
	1	2	3	4	5	6
Pleasant	-0.25	**0.74**	0.14	0.04	-0.04	-0.12
Interesting	-0.17	0.08	**0.70**	-0.06	0.04	-0.26
Beautiful	-0.39	**0.72**	0.33	0.08	0.04	-0.07
Normal	0.14	**0.77**	-0.18	0.13	-0.01	0.17
Calm	0.30	**0.48**	-0.17	0.15	0.04	0.28
Bright	-0.04	0.12	0.05	**0.98**	0.05	-0.14
Opaque	0.09	0.05	0.02	-0.05	-0.01	**0.53**
Matt	0.07	-0.003	-0.01	-0.13	**0.89**	0.13
Inexpensive	**0.48**	0.19	-0.22	-0.13	0.16	-0.05
Anaesthetic	**0.76**	-0.16	-0.08	-0.02	-0.06	0.12
Soft	-0.05	0.02	0.07	0.05	**0.38**	-0.05
3D	-0.13	-0.09	**0.74**	0.11	0.12	0.31
Black	-0.13	-0.31	0.05	**-0.46**	0.36	-0.18
Unappealing	0.41	-0.40	**-0.45**	0.05	-0.18	-0.01
Not artistic	**0.60**	0.04	-0.14	0.14	-0.04	0.10
Pour	**0.82**	-0.17	-0.04	-0.03	-0.06	0.06

Table 4: *Rotated Factor loadings for a six dimensional model for the translucent objects. Note that the each scale label represents a value of 7 for that scale.*

Although the objects in the second dataset, the translucent objects differed along two distinct physical dimensions (scattering and absorption), participants' similarity ratings suggest that they judge the samples to be related to one another along a single, non-linear manifold embedded in this 2D space. While the changes in the scattering coefficient do affect the apparent surface gloss and transparency, the dominant semantic differential for these stimuli relates to their "brightness" and "blackness", suggesting that subjects paid close attention to the distribution of intensities in the image when making their judgments. These two findings are particularly interesting given previous work on the perception of translucency. Fleming and Bülthoff [FB05] suggested that human vision estimates translucency using a range of simple image measurements that correlate with physical changes, especially measures of the distribution of intensities in the image. For example, by modifying the intensity histogram of an image it is possible to make an object appear more opaque or more translucent. This could provide some insight into the pattern of our results: participants' judgments of similarity are dominated by their impression of the distribution of brights and darks in the image, which is a salient and reliable source of information about translucency.

Given that the majority of the changes in apparent translucency for the second dataset occurred in between the first two parameter values, future research on these objects should sample this range more densely. Furthermore, future work should probably reduce the number of redundant Evaluative scales, and increase the number of material property scales. Nonetheless, the method has already shed some light on exactly how one might produce specific desired effects with the two algorithms tested, and provided a few insights into how humans perceive surface gloss and translucency.

References

[CC01] COX T., COX M.: *Multidimensional scaling (2nd ed.).* Chapman & Hall, London, 2001.

[FB05] FLEMING R., BÜLTHOFF H.: Low-level images cues in the perception of translucent materials. *ACM Transactions on Applied Perception 2*, 3 (07 2005), 346–382.

[JB02] JENSEN H. W., BUHLER J.: A rapid hierarchical rendering technique for translucent materials. *ACM Trans. Graph. 21*, 3 (2002), 576–581.

[JW63] JUDD D., WYSZECKI G.: *Color in Business, Science and Industry, 2nd Edition*. Wiley, New York, 1963.

[Kai60] KAISER H. F.: The application of electronic computers to factor analysis. *Educational and Psychological Measurement 20* (1960), 141 – 151.

[Min61] MINDAK W. A.: Fitting the semantic differential to the marketing problem. *Journal of Marketing 25* (1961), 28 – 33.

[OST57] OSGOOD C., SUCI G. J., TANNENBAUM P. H.: *The Measurement of Meaning*. University of Illinois Press, Urbana, 1957.

[PFG00] PELLACINI F., FERWERDA J. A., GREENBERG D. P.: Toward a psychophysically-based light reflection model for image synthesis. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), pp. 55–64.

[Ric38] RICHARDSON M.: Multidimensional psychophysics. *Psychological Bulletin 35* (1938), 659 – 660.

[SF06] SCHUBERT E., FABIAN D.: The dimensions of baroque music performance: a semantic differential study. *Psychology of Music 34* (2006), 573 – 587.

[Spe04] SPEARMAN C.: 'General intelligence' objectively determined and measured. *American Journal of Psychology 5* (1904), 201 – 293.

[SRY81] SCHIFFMAN S., REYNOLDS M., YOUNG F.: *Introduction to Multidimensional Scaling*. Academic Press, New York, 1981.

[Tor52] TORGERSON W.: Multidimensional scaling: I. theory and method. *Psychometrika 17* (1952), 401 – 419.

[WS82] WYZECKI W., STILES W.: *Color Science: Concepts and Methods, Quantitative Data and Formulae, 2nd Edition*. Wiley, New York, 1982.