A Somewhat Fuzzy Color Categorization Model[1]

Johan M. Lammens

Scuola Superiore S. Anna
Advanced Robotics Technology and Systems (ARTS) Laboratory
via Carducci 40, 57127 Pisa, Italy
e-mail lammens@arts.sssup.it
URL http://www.cs.buffalo.edu/~lammens

Submitted to ICCV-95

Abstract

We present a computational model of color categorization using a normalized Gaussian function of perceptual color space coordinates as the basic category model, and an application based on the model that can *name, point out*, and *select* colors in images, and provide a *confidence* or (fuzzy) *membership* value for each categorial judgement. We quantify the performance of our model relative to existing data about human color naming behavior, and relative to the use of different color spaces, including a novel one derived from neurophysiological measurements.

The application we present is to some extent able to deal with the *color constancy* problem in real images made in an unmodified office environment with low-grade uncalibrated equipment. We present some empirical results with an example of this kind of image.

The novelty of the approach lies in the fact that it models *graded* or *fuzzy* color categories as found in anthropological, psychological and psychophysical color perception work, and relates these categories in a simple way to the symbolic labels that natural languages use for them. As such the approach can be seen as a particular case study in *symbol grounding*, *embodiment*, or *situated cognition*, based on a well-defined computational model.

[1]The research reported in this paper was done while the author was a member of the SNePS Research Group at the department of Computer Science, State University of New York at Buffalo, and was supported in part by Equipment Grant No. EDUD-US-932022 from SUN Microsystems Computer Corporation, and in part by NASA under contract NAS 9-19004.

1 Introduction

In this paper we present a computational model of color categorization using a normalized Gaussian function of perceptual color space coordinates as the basic category model, and an application of the model which is able to

1. *Name* real colors in response to real visual stimuli (camera images), and provide a confidence or "goodness of example" or (fuzzy) membership value.

2. *Point out* examples of named colors in real images and provide a confidence rating, and as a derivative of this capability, pick the *best example* of a named color from a set of color samples, or from an image in general.

The model is at present contrained to the eleven so-called *basic color categories* and their corresponding names in English, as defined first in the work of [1]. The model's performance on these tasks is quantitatively shown to be reasonably consistent with human performance. In particular, this means that

1. The model places the *foci* of the basic color categories in the same regions of the color space as human subjects do.

2. The model places the *boundaries* of the basic color categories in the same regions of the color space as human subjects do.

As such, the color categorization application can be seen as a (partly) *embodied* [13, 12, 9] or *grounded* [8] system, or as an instance of *situated cognition* [20].

The physical stimulus involved in color perception is light, or more precisely: electro-magnetic radiation in the visible wavelength range, approximately 380–770 nm. Representing the spectral energy distribution of the stimulus (the radiation striking the retina) as a function of wavelength $E(\lambda)$ (after [4]), the stimulus energy distribution of the radiation that is reflected off (or transmitted by) an object is determined by (1) the spectral distribution of the radiation incident on the object (the light source) $S(\lambda)$, and (2) the spectral reflectance characteristics of the object $R(\lambda)$, so we can write $E(\lambda) = S(\lambda)R(\lambda)$ in a simplified form. It is easily seen from this equation that changing the light source or changing the reflectance characteristics of the object changes the stimulus $E(\lambda)$. Were we informally to think of *object color* as $R(\lambda)$, then the task of the color vision system is to recover $R(\lambda)$ from $E(\lambda)$ (for some examples of this approach, see [17, 16]). The human visual system is to some extent able to identify the same "object color" (surface reflectance, under the view just mentioned) under widely varying lighting conditions, a phenomenon known as *color constancy*. We will not be concerned with color constancy in the first place (but see Section 4.4.1). What we are concerned with is the internal structure of the color categories that invariant object colors are perceived as belonging to. That is not to say that we will adhere to the Color Science aperture mode of color perception either; we will take a mainly psycho-physical approach to modeling color perception.

We define *color naming* as a mapping $\mathcal{N}$ from visual stimuli to pairs of color terms and numeric "confidence", "goodness", "typicality", or (fuzzy) "membership" values. More precisely, since we are not interested in the spatial characteristics of visual stimuli, we will represent stimuli as spectral distributions $E(\lambda)$ associated with single points in the visual field only. The domain of $\mathcal{N}$ is thus the set of all such distributions: $\mathbf{E} = \{E \mid E : \lambda \mapsto \mathbf{R}^+\}$, where $\lambda = [380, 770]$ represents wavelength in nm, and E represents a spectral distribution. The mapping $\mathcal{N}$ can then be defined as $\mathcal{N} : \mathbf{E} \mapsto \mathbf{C} \times \mathbf{I}$, where $\mathbf{C}$ is a finite set of undefined terms, and $\mathbf{I}$ is the closed unit interval $[0, 1]$.

Informally, the preferred interpretation of the model is that $\mathbf{E}$ represents the set of spectral distributions that function as input to the visual system, $\mathbf{C}$ represents a set of basic color terms, e.g. $\{$*white, black, red, green, yellow, blue, brown, pink, purple, orange, grey*$\}$, which is the set of English basic color terms described in [1], and $\mathbf{I}$ represents the set of "confidence", "goodness", "typicality", or (fuzzy) "membership" values.

If we ignore the i part of the $\langle c, i\rangle \in \mathbf{C} \times \mathbf{I}$, we might think of $\mathbf{C}$ as representing a partition of the set of all possible color percepts $\mathbf{E}$ into a number of equivalence classes, one for each color term. If we do take the i part into account, we could model the response to any given stimulus $E \in \mathbf{E}$, i.e. $\mathcal{N}(E)$, as a *fuzzy partition* of the set of color terms $\mathbf{C}$ [22, 12], with the additional constraint $\sum_{c\in\mathbf{C}} 2^{nd}\langle c, i\rangle = 1$. The mapping $\mathcal{N}$ then defines a set of *membership* or *characteristic functions* f_c, one for each color category $c \in \mathbf{C}$, on the universe $\mathbf{E}$.

The definition of $\mathcal{N}$ implies that we will not be concerned with recovering $R(\lambda)$ or $S(\lambda)$ from $E(\lambda)$. A practical consequence of this is that changing the lighting of a scene may yield a different value of $\mathcal{N}$. We don't consider this a problem as long as the change is consistent with human performance on the same task.

If we want to model the relation between the domain of $E(\lambda)$ and a set of color terms, i.e. if we want to model *color naming*, it is not sufficient to define the extensions of color terms as intervals on the wavelength range, which might be intuitively appealing. Even disregarding the fact that the spectral compostion of stimuli is usually not known, this approach could only work for pure monochromatic stimuli, which are very rare in real-world situations, and it would not explain the typical graded membership functions one finds when subjects are asked to identify best examples and maximal extensions of color terms [1]. This approach would also leave out non-spectral colors like *purple*, *brown*, or *white* altogether.[2]

A "color model", or "color space", or "color order system" is a way to organize the set of possible human color percepts in some systematic way. Four families of color models can be distinguished in the color vision literature at large (see, e.g., [3], [21], and [4] for some references, and [5] for an overview of color models and color theory). We use these categories merely as an organizational tool; they are not entirely mutually exclusive:

1. Physiologically inspired models using 3 primaries, based on the three types of cones in the human retina, starting with Hering's theory of color vision, e.g., the familiar RGB models used in computer graphics hardware.

2. Colorimetric color models, based on physical measurement of spectral properties, usually using three primary color filters and some kind of photometer, e.g., the CIE chromaticity diagram.

3. Opponent models, based on perception experiments, using pairwise opponent primary colors (yellow and blue, red and green), starting with the Young-Helmholtz theory, e.g., [10].

4. Psychological and psychophysical color models based on the *appearance* of colors to human observers, with data derived either in an impressionistic way (e.g., the Munsell and Ostwald color models) or in an experimental way (e.g., the Hue-Saturation-Brightness (HSB) family of color models, or the OSA uniform color space).

Elsewhere we have derived a colorspace from neurophysiological data [14], which we will refer to in what follows as the Neuro-Psycho-Physical or NPP color space. Other types of color spaces will be refered to with their usual names, and are defined in the references quoted.

2 Related Work

The notion of *basic color terms* originates in the anthropological and linguistic work of Brent Berlin and Paul Kay in the late sixties [1]. They studied color naming behavior with native speakers of a variety of languages, and the existing literature on the subject, and recorded two main findings:

1. There are substantial universal constraints on the shape of basic color lexicons.

2. Basic color lexicons change over time by adding basic color terms in a highly constrained, though not mechanically predictable, manner.

The research was carried out against a background of extreme linguistic relativism, also known as the Sapir-Whorf hypothesis [11], which holds that each language performs the coding of experience into language in a unique and arbitrary manner and that there are no semantic universals (principles of meaning that hold across all languages) in principle.

The only models of color naming we are aware of that attempt to explain the universality of foci and graded membership functions are [12] and [6]. Apart from not being defined or implemented as computational models, both of these have important drawbacks. Kay and McDaniel's model interprets (stylized versions of) the response functions of four types of color-sensitive cells in the LGN [7] as characteristic functions of four fuzzy sets corresponding to the categories red, green, yellow, and blue. As such, the model accomodates the existence of universal foci, which correspond to maxima of the characteristic functions, and graded membership functions, which correspond directly to the characteristic functions, of these four

[2] *Non-spectral colors* are colors that do not appear in the spectrum of white light, i.e. they are not single wavelength signals but complex signals. In the frequency domain, they are represented as non-impulse functions.

categories. But the model has to be tweaked to account for other categories, requiring the introduction of new and ad hoc fuzzy set operations and a nonlinear compression of the wavelength dimension that is without apparent external motivation, and it is not clear at all how non-spectral color categories like brown or purple are to be dealt with. In their model, a flat-spectrum white stimulus light should be a good example of *every* color category, since there is no opponent mechanism that cancels out opponent primaries. This prediction is clearly not born out by experience or by psychophysical experimentation. In addition, as we have pointed out before, the spectral composition of stimuli is usually not known, and certainly not easily derivable from the output of the cones in the human retina, nor from that of a typical color (RGB) sensor used in cameras or scanners.

Cairo's model of color naming [6] is also based on findings in the physiology of the pre-cortical visual system. It is four-dimensional: wavelength, intensity, purity, and adaptation state of the retina. All of these dimensions are defined as physical parameters of the stimulus, rather than as perceptual dimensions. Cairo introduces a "data-reduction model" of analog-to-digital conversion in the pre-cortical visual system, and represents Berlin and Kay's eleven basic color categories as specific combinations of quantized values on the four dimensions. In addition, he predicts four potential new basic color categories beyond Berlin and Kay's eleven, which he calls *sky-blue, turquoise, lime*, and *khaki*. Although this model is interesting for its attempt to take adaptation into account, it suffers from a number of important drawbacks. For instance, the discrete nature of the model and the use of wavelength as one of the dimensions forces complex stimuli to be treated as if they were monochromatic, and this is done using the "dominant wavelength" to represent an entire spectrum. While this may work in limited circumstances, it is clearly a gross simplification which cannot hold in general; e.g., which is the "dominant wavelength" of a flat spectrum or a perfectly symmetric bimodal spectrum? In addition, Cairo claims that Berlin and Kay's finding of universal foci of categories is a mere artifact introduced by the use of the Munsell color system [2] for the stimulus material used in the experiment. This conclusion does not seem to be supported by the wealth of research done since Berlin and Kay's initial study (cf. the bibliography of recent research in [1]). On the contrary, it seems that it is rather Cairo himself who is forced to deny the evidence because of the discrete nature of his model, which does not allow for graded categories. The discretization of the model itself seems to be fueled by a desire to make color notation the subject of algebraic analysis, requiring a complete ordering of an integer number of possible color values. While that may be a noble cause, it also requires most of the data on actual color naming behavior to be ignored, which is too much of a price to pay for a convenient analysis. Berlin and Kay's hierarchy of languages is explained in a rather circular fashion by Cairo in terms of the "perceptual salience" of the four underlying dimensions, but without explaining what their salience derives from.

3 From Color Space to Color Names

Conceptually, we can think of the mapping $\mathcal{N}$ as consisting of two parts: the first part takes us from visual stimuli to points in a color space, and the second takes us from a color space to a set of color names, each with a "typicality" measure associated. We now describe the procedure used for fitting a particular category model to the experimental color naming data of [1], and present a theoretical evaluation of the model.

3.1 The Normalized Gaussian Category Model

In [19], Shepard provides an elegant argument for his claim that the probability of generalization of an existing (known) category to a new (unknown) stimulus is a monotonic function of the normalized distance in psychological space of the unknown stimulus to known stimuli belonging to the category. He further specifies that this function can be approximated by a simple exponential decay or, under certain circumstances, a Gaussian function. The distance metric is either Euclidean, resulting in circular (or spheroid) contours of equal generalization, or a slight variant that results in elliptic (or ellipsoid) contours of equal generalization.

Following Shepard's suggestion, we have used a variant of the Gaussian normal distribution as a category model, which we will refer to as the *normalized Gaussian* model. The usual normal curve in one variable, as used in probability theory, is given by $G(x) = \frac{1}{\sigma\sqrt{2\Pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, where σ is the standard deviation, and μ is the mean or expected value. Since the normal curve is used as a probability density function, it has the special property that $\int G(x) = 1$. To derive the normalized Gaussian function, we drop the factor $\frac{1}{\sigma\sqrt{2\Pi}}$, since we don't need the interpretation as a probability density function, and we substitute the general

N-dimensional Euclidean distance function for the distance term $x - \mu$, which gives us

$$G_n(x) = e^{-\frac{1}{2}\left(\frac{\sqrt{\sum_{i=1}^{N}(x_i-\mu_i)^2}}{\sigma}\right)^2} \tag{1}$$

An example of a two-dimensional version of this function is shown in Figure 1. This function has a number

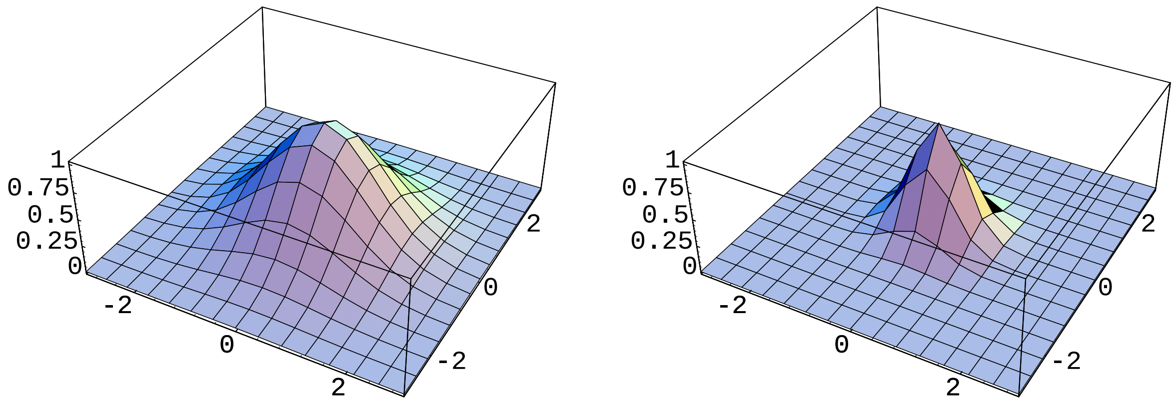

Figure 1: An example of a two-dimensional normalized Gaussian function with $\mu = (0, 0)$ and $\sigma = 1$ (left) or $\sigma = 0.5$ (right).

of interesting properties for use as a color category model:

- The maximum value occurs at μ and is unity, regardless of the value of σ.
- The "width" of the curve is a function of the parameter σ only.
- The value decreases monotonically, but not linearly, as a function of distance to μ.
- It is strictly positive everywhere except at infinite distance from μ.
- It has a simple mathematical definition, using only two parameters μ and σ.

These properties allow us to interpret μ as the location of the center or focus of the category, and the function value as the memberhip value of a stimulus represented by its color space coordinates. By modulating the value of σ we can affect the size of the volume of the color space that is included in the category, relative to some threshold membership value. The maximum membership value is unity, as is common in fuzzy set models, but we can get a membership value for any point in the color space, no matter how far removed from the focus. This is important for our purpose, as we will explain below. The non-linear decrease of membership with distance from the focus reflects the general shape of psychological categories as discussed by [19], and it can result in category extensions different from those obtained with a simple nearest-neighbor criterion, since neighboring categories may have different values of σ. The representation is also economical, since a category can be represented by only two parameters.

3.2 Locating Color Stimuli in the Color Space

Before we can apply the category model described in the previous section to Berlin and Kay's color naming data, we need to quantify the stimulus set they used. It contained 329 Munsell color chips consisting of 40 equally spaced hues at 8 equally spaced brightness (Value) levels each, all at maximum saturation (Chroma), and a gray scale consisting of nine equally spaced brightness (Value) steps. They asked subjects to point out both the extent and the foci of the basic color categories of their native language on the array of color chips, viewed under a CIE A light source (Figure 2).

In what follows we will always use the CIE XYZ, the CIE L*a*b*, and the NPP color spaces for comparative purposes. The XYZ space is in a sense "the mother of all RGB spaces", since the various

Figure 2: The extent and focus for each of the eleven basic color categories of (American) English, from [1, p.119]. The vertical bar on the left represents the nine values of the grey scale used, the rectangular array represents the set of 320 color chips: 40 equally spaced hues (horizontal) by 8 equally spaced brightness levels (vertical). The shaded areas represent the extent of the categories, the dots inside the shaded areas their foci (a category can have a focus extending over more than one chip).

RGB spaces are simple linear transforms of it. It is generally accepted as an approximation to the spectral sensitivities of the human cone photoreceptors, and thus a "primary" representation, as close to the sensor as we can hope to get. The L*a*b* space is defined by the CIE to be perceptually equidistant across (most of) the color gamut, and is often used as a reference in color work. It is a non-linear transform of the XYZ space. For mathematical derivations of these and other color spaces, see Appendix B in [14]. The NPP space is the one that we derived from neurophysiological measurements in Chapter 5 of [14], and is also a non-linear transform of the XYZ space. We used this space to attempt to link the category model to the underlying neurophysiology.

The conversion from Munsell coordinates, in which the Berlin & Kay stimulus set is defined, to CIE XYZ coordinates, which is the basis for the color spaces we are interested in, is non-trivial, and there is no simple closed-form formula to do the job. Fortunately, the Munsell set of standard color reference chips has been measured spectrophotometrically and converted to CIE xyY coordinates in the past [18]. After conversion from CIE xyY, we obtain unnormalized CIE XYZ coordinates for each of the stimuli contained in the set. To normalize the coordinates to the unit cube, with the gray axis going from $\langle 0,0,0\rangle$ to $\langle 1,1,1\rangle$, we used Von Kries adaptation: $VK(\overline{x}) = \frac{\overline{x}}{\overline{w}}$, where $\overline{x}$ is the vector representing the unnormalized stimulus values, and $\overline{w}$ is the vector representing the unnormalized white reference stimulus values. Although Von Kries adaptation cannot theoretically be shown to undo all the effects of a non-flat spectrum light source, it works well enough in practice to be allowable, especially with a light source that is as close to a flat spectrum as the CIE C source used in these measurements [21]. The obtained stimulus set is shown in Figure 3, represented in CIE XYZ, CIE L*a*b*, and NPP coordinates. There are some interesting things to note about this figure:

- In all color spaces, the stimuli lie near or on the OCS[3] surface, as is expected since they represent the maximum saturation levels available at each brightness and hue value. We computed the OCS surface completely independently, however, as described in [14]. This provides a nice cross-check for the consistency of both data sets.

- The spacing of the stimuli varies in the various color spaces. Although it is difficult to see from the figure, the spacing is most equal in the CIE L*a*b* space, which was created explicitly for this purpose. The NPP space is less than perfect in this respect.

- There are some irregularities in the spacing of the stimuli, particularly in the blue region. They had

[3] Optimal Color Stimuli; the theoretical limit of physically realizable surface colors [21].

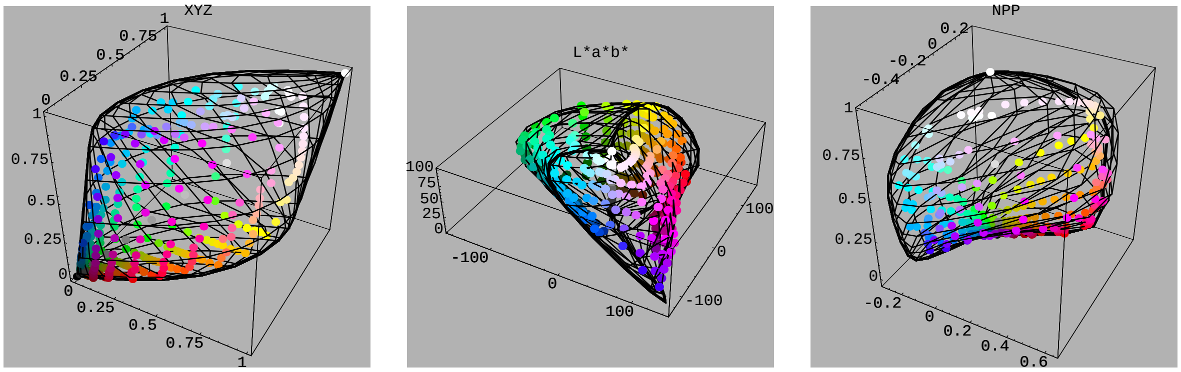

Figure 3: The Berlin and Kay stimulus set, represented in the CIE XYZ, CIE L*a*b*, and the NPP color spaces (left to right). Each stimulus is represented as a dot. The stimuli are shown together with a wire-frame rendition of the OCS surface.

already been noticed by [18], who contribute them to problems with the underlying CIE standard observer functions.

Combining the information from Figure 2 with the derived color space coordinates of the stimuli, we can now describe the boundary of a category as a polygon passing through the coordinates of each of the boundary stimuli, and the focus of a category as the center of mass of the points indicated as focal points. Figure 4 shows the boundaries and foci obtained in this way. Note that the shape and size of the polygons

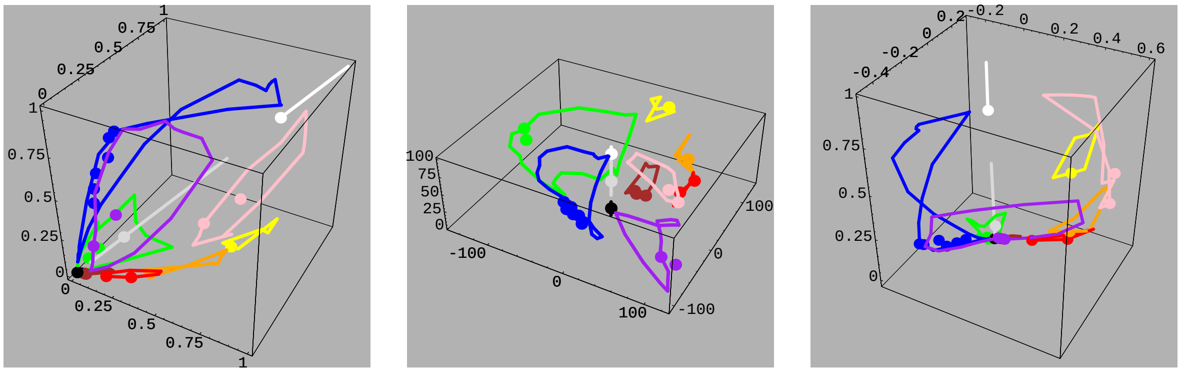

Figure 4: The boundaries and foci of the Basic Color categories for (American) English, represented in the CIE XYZ, CIE L*a*b*, and NPP spaces (left to right). Boundaries are represented as polygons and foci as dots.

is different in different color spaces, and that in general a straight line on the Berlin and Kay chart does not necessarily translate into a straight line in the color space, since the stimuli are lying on or near a curved surface.

3.3 Fitting the Model to the Data

The problem of fitting the model to the data is complicated by having a continuous-valued model (the normalized Gaussian) but boolean data (member of the category or not, focal example or not). We need to choose some threshold value to be associated with the category boundaries, and another value to be associated with the foci. Given our model it is natural to choose a value of unity for the focus, and for the threshold value we have arbitrarily chosen $\frac{1}{2}$. After some experimentation, we have obtained the best results with a fitting procedure that minimizes the following quantity (independently computed for each category):

$$E = \sum_{i=1}^{N_v} \left[G_n(\overline{v_i}, \overline{s}\,\overline{\mu}, \sigma) - \theta \right]^2 + w_f \sum_{i=1}^{N_f} \left[G_n(\overline{f_i}, \overline{s\mu}, \sigma) - 1 \right]^2 + \sum_{i=1}^{N_o} \left[T\left(G_n(\overline{o_i}, \overline{s\mu}, \sigma) \right) \right]^2 \tag{2}$$

which is of course a sum squared error criterion. In this equation, E is the total sum squared error for a particular category, composed of three separate terms. The first term quantifies the error of fit with respect to the vertices of the bounding polygon, with N_v representing the number of vertices, G_n the normalized Gaussian function from equation 1, $\overline{v_i}$ a boundary vertex vector, $\overline{s}$ a linear scaling vector for the initial focus $\overline{\mu}$ of the normalized Gaussian, σ the "width" parameter of the normalized Gaussian, and θ the threshold value for category membership, in our case equaling $\frac{1}{2}$. The second term quantifies the error of fit with respect to the focal stimuli, with w_f representing a fixed weight for this error term, set to $\frac{1}{3}$ by trial-and-error, N_f representing the number of focal stimuli (indicated by black dots in Figure 2), $\overline{f_i}$ a focal stimulus vector, and the remaining symbols as in the first term. The third term quantifies the error of fit with respect to other-category representatives, with N_o representing the number of other-category representatives, T a threshold function, $\overline{o_i}$ an other-category representative, and the other symbols as before. The other-category representatives are collected by selecting from the boundary or focus stimuli of each other category those that are closest (in terms of Euclidean distance) to the initial focus of the current category, for each dimension separately. The union of the sets of such stimuli over all dimensions becomes the set of other-category representatives. It is necessary to select the nearest points in each dimension separately, because the dimensions may be scaled independently in a subsequent step. The threshold function T is defined as

$$T(x) = \begin{cases} 0, & x < \eta \\ x, & otherwise \end{cases} \tag{3}$$

The effect of using T in the third term is to create a horizon for other-category representatives: as long as their function value for the *current* category's model function is below some value η, it is not counted at all. We have used $\eta = 0.9\,\theta$. The minimization of E was carried out with respect to σ and $\overline{s}$ (each component independently) by Mathematica's standard FindMinimum function, which finds local minima in a function by following the path of steepest descent from any point it reaches. Loosely speaking, this amounts to moving the focus of the model function around in the color space so as to bring the boundary stimuli's membership values as close as possible to θ, the focus stimuli's membership values as close as possible to 1, and the membership values for any other category's representatives below the horizon value η.

The use of other-category representatives in the error function may cause category boundaries to shift to some extent as a function of the distance, or presence or absence, of other categories nearby. There is some evidence in Berlin and Kay's work that the extent of categories may indeed be influenced by the total number of basic categories, and hence by their mutual distance, but we have not examined this any further.

Since the minima found by the FindMinimum function are local, the initial values for the variables are quite important. For the initial value of $\overline{s}$ we have used the identity vector $\langle 1,1,1 \rangle$, and for $\overline{\mu}$ we used the initial value that places it at the center of gravity of the focal stimuli. In order to determine an initial value for σ, we found the same-category stimulus $\overline{s_m}$ that is farthest away from $\overline{\mu}$ (using Euclidean distance), and then determine σ such that $G_n(\overline{s_m}, \overline{s}\,\overline{\mu}, \sigma) = \theta$. Since the normalized Gaussian is a monotonic function of distance to the focus, and $\overline{s_m}$ is the point in the set farthest away from the focus, any value of σ larger than that will result in a monotonically increasing square error over the complete set, and values smaller than that will lead to the nearest local minimum. Even if this minimum is not the global one, this procedure will ensure it is the "same" minimum for all categories. In principle the error landscape as a function of σ can have up to N local minima for N points over which the error is computed. In practice the minimum found in this way seems to be usually the global one, or at least very close to it. Figure 5 illustrates this for the 1-dimensional case.

After determining the values of $\overline{s}$ and σ through error minimization, a second minimization step is applied, which involves distorting the Euclidean nature of the space. The first step used a Euclidean distance metric, resulting in spheroid contours of equal membership value. The second step allows the dimensions of the space to be scaled individually for each category model, which can result in a non-spheroid contour of equal membership value. The procedure for this scaling involves an other error minimization comparable to the first, with an objective function which is basically the same as equation 2, but having

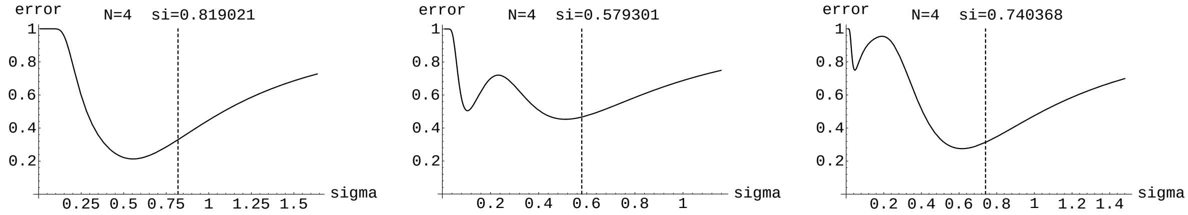

Figure 5: The value of the first term of the error E (equation 2) as a function of the parameter σ, for fixed values of $\overline{\mu} = 0$ and $N_v = 4$, with randomly chosen 1-dimensional points in $[0, 1]$. The initial value of σ is indicated by the dashed line. On the X axis: σ, on the Y axis: error.

both data points and category foci scaled by a vector $\overline{s'}$, and with $\overline{\mu'} = \overline{s}\,\overline{\mu}$ where $\overline{s}$ is the result of the first minimization step. Informally speaking, this amounts to stretching (or compressing) the dimensions of each category to make it fit the data better. Alternatively, one can think of it as stretching the dimensions of the color space to make it fit the categories better, but only locally. The obtained scaling vectors for the XYZ, L*a*b*, and NPP spaces are shown in Figure 6. One way to interpret this figure is that the

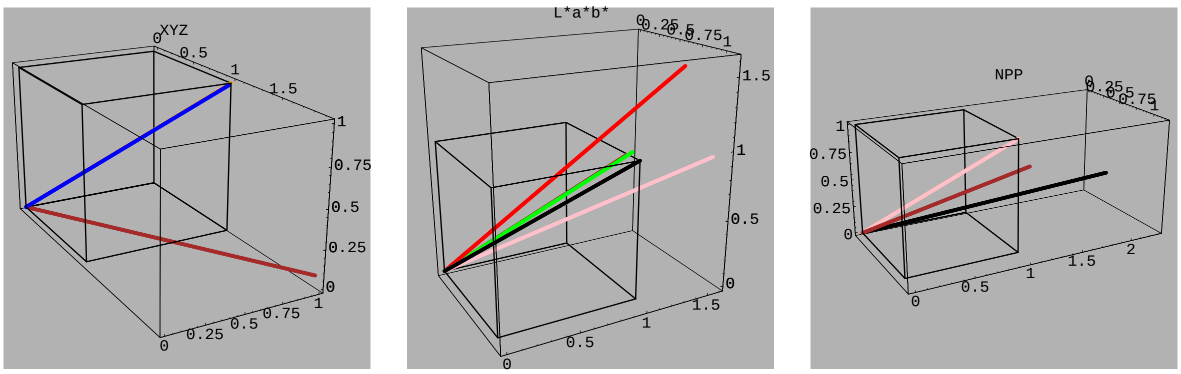

Figure 6: The scaling vectors obtained by minimizing with respect to σ' and $\overline{s'}$, for the XYZ, L*a*b*, and NPP spaces (left to right). For reference, the unit cube is displayed as well. The unit scaling vector, which leaves the space unchanged, is the main diagonal of the unit cube, going from $\langle 0, 0, 0\rangle$ to $\langle 1, 1, 1\rangle$. Note that only one vector shows in case of multiple overlapping ones.

closer the vectors cluster to the main diagonal of the unit cube, the more Euclidean the space is (or the categories are), since that corresponds to spheroid surfaces of equal membership value. There are clear differences between the spaces in this respect, but there does not seem to be one "best" space.

The result of fitting the category models to Berlin and Kay's data for (American) English, using the two-step method described above, is shown in Figure 7. This figure shows the locations of the category centers (the values of $\overline{\mu'}$ for each category model) only; the boundaries will be discussed below. It is interesting to note that some of the foci lie outside of the OCS volume, which means that they do not correspond to physically realizable surface colors.[4] Technically, this is the result of allowing the foci to "float", in the first part of the fitting procedure. While it is possible to constrain the foci to lie on or below the OCS surface, that results in a considerably poorer fit, and in degraded performance for the algorithms using the category models.

[4] Whether this is equivalent to "perceivable" colors is not clear. Could one dream in Extreme TechniColors that are not physically realizable?

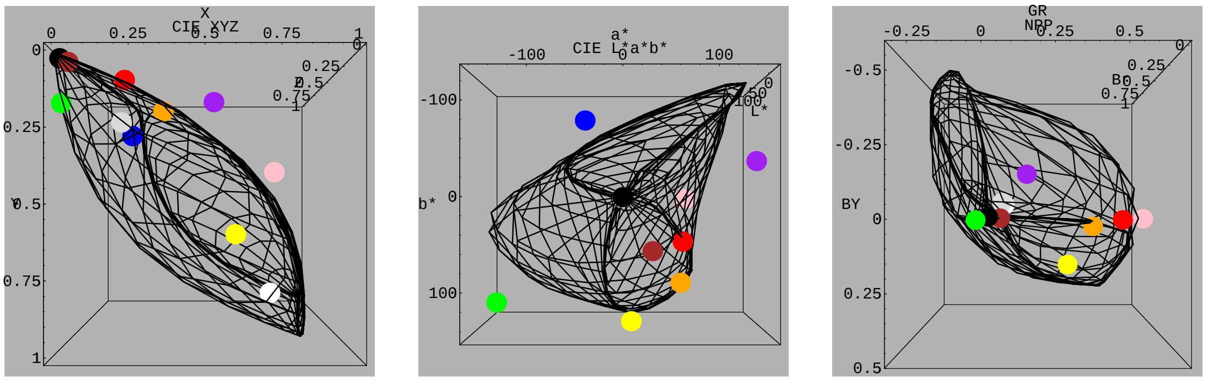

Figure 7: The locations of the category model foci (values of $\overline{\mu'}$) in XYZ, L*a*b*, and NPP spaces (left to right) are indicated by dots, relative to a wire-frame rendition of the OCS surface, in a view from "below", i.e. from the side where the dark colors are found. The location of the blue category focus is not shown for the NPP space, because it lies too far outside of the figure's bounding box.

3.4 Quantitative Evaluation of the Category Model

To evaluate the category model, we will compare its categorial judgments to the Berlin and Kay data. We will do this for each color sample in the Berlin and Kay stimulus set, which is actually a superset of the data the model was fitted to. We are thus to some extent testing the generalization capabilities of the model. Note that our model implies convex category regions, and while that is not always the case for the regions as depicted in Figure 2 (e.g. the green region), it is generally true for the regions as mapped onto the OCS surface in the various color spaces, perhaps with some minor local exceptions. Figure 8 illustrates this comparison.[5]

From visual inspection of the figure it is apparent that the categorization works better in some spaces than in others. In some spaces a certain category may "bleed" outside of the Berlin and Kay category boundaries, e.g. in the XYZ space, green intrudes into the brown region, and black into green; blue spills over into the purple region, purple somewhat into pink, white into pink, and green and blue slightly into gray. Other categories exceed their boundaries without intruding into another basic category's region, e.g. yellow in all three spaces (but particularly in L*a*b*). Another type of mismatch is a category that does not fill enough of its region, e.g. orange, brown, green, blue, and purple in XYZ. The same type of errors occur in all three spaces, but to varying extents. The L*a*b* space seems to perform best in general, followed by the NPP space and the XYZ space. Some errors do not seem as serious as others, e.g. the row of white judgments for the very pale blue and pink stimuli in the top row in the NPP space does not seem like a particularly serious mistake, but judging gray as green or blue in the XYZ space or as purple in the NPP space does. This judgment is qualitative in nature, of course, and somewhat subjective. In addition, performance on color-related tasks may be considered more important than this type of theoretical evaluation (see Section 4).

To try and arrive at a more objective measure of performance, we now turn to quantifying the "goodness of fit" of the category models. There are two concurrent criteria we can use: how well the extent of a model category fits the extent of the corresponding category in the data, and how close the model focus of each category is to the corresponding category focus in the data. The following error metric attempts to capture the first of these, and is computed over the complete set of Berlin and Kay stimuli:

$$E_x = \sqrt{\frac{1}{N_s}\sum_{i=1}^{N_s}(\alpha_p^i - \alpha_e^i)^2} \tag{4}$$

[5] Since a grey-scale rendition of this and other figures necessarily leaves out too much, a color version of the paper will be made available for ftp and/or WWW access.

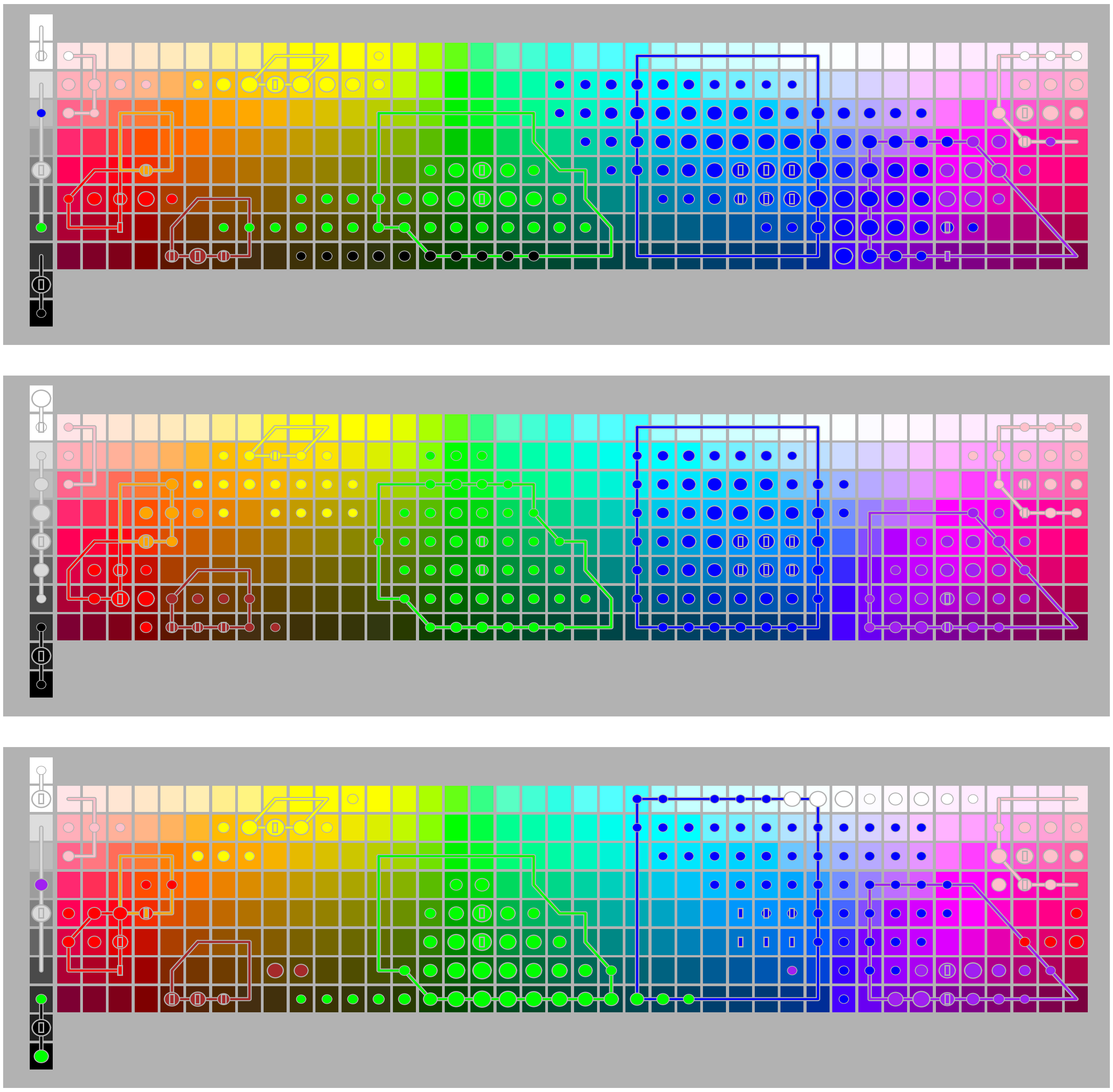

Figure 8: Categorization results for the XYZ, L*a*b*, and NPP color spaces (top to bottom). The category model judgments are shown as disks with a diameter proportional to the membership value, superimposed on a recreated stimulus chart, with Berlin and Kay's category boundaries and focal stimuli indicated by lines and small rectangles, respectively. If all category functions return a value below the threshold θ, no disk is shown.

$$
\begin{aligned}
\alpha_p^i &= \begin{cases} 0, & \alpha_m^i < \theta \wedge C_m^i \notin P^i \\ \alpha_m^i, & \alpha_m^i < \theta \wedge C_m^i \in P^i \\ 0, & \alpha_m^i \geq \theta \wedge C_m^i \notin P^i \wedge P^i \neq \emptyset \\ \alpha_m^i - \theta, & \alpha_m^i \geq \theta \wedge C_m^i \notin P^i \wedge P^i = \emptyset \\ \theta, & \alpha_m^i \geq \theta \wedge C_m^i \in P^i \end{cases} & (5) \\
\alpha_e^i &= \begin{cases} \theta, & P^i \neq \emptyset \\ 0, & otherwise \end{cases} & (6) \\
\alpha_m^i &= \max\{\alpha_1^i \ldots \alpha_{N_c}^i\} & (7) \\
C_m^i &= C_j \quad such\ that \quad \alpha_j^i = \alpha_m^i & (8) \\
P^i &= \{C_i \mid s_i \in C_i\} & (9)
\end{aligned}
$$

where E is the total error, N_s is the number of stimuli in the set, α_p^i is the predicted (model) and α_e^i is the expected (data) membership value for stimulus i, α_m^i is the maximum membership value for stimulus i over all category model functions, θ is the membership threshold as before, C_m^i is the category yielding the α_m^i, P^i is the set of all data categories that the stimulus belongs to (decided by being within Berlin & Kay's category boundaries), α_j^i is the membership value for stimulus i in model category C_j, and s_i is stimulus i. The complicated way of determining α_p is necessary to deal with various cases such as the predicted category matching or mismatching the expected category, and with boolean data versus continuous model. In addition to the error over the complete data set, the square error for each stimulus is added to the running total for category C_j iff $(C_j \in P^i) \vee (C_j = C_m^i \wedge \alpha_m^i \geq \theta)$, i.e. either when the data or the model says it should belong to category C_j, so we will get an error when the model category is either too small or too large, as compared to the same category in the data.

The error in the placement of the category model foci is determined as follows: for each category C_j, determine the stimulus $\overline{s_i}$ that is closest to the focus $\overline{\mu_j}$, using Euclidean distance. This is also the stimulus with the maximum membership value α_j^i, since that is a monotonic function of distance to the focus. Then the center of gravity $\overline{\mu_j'}$ of the data focal stimuli is determined, and finally the focal error is computed as the Euclidean distance between the two. This is approximately equivalent to the distance between the data focus and the orthogonal projection of the model focus onto the OCS surface (there may be small deviations because the stimuli and data focus may lie slightly below the OCS surface). The focal error is computed per category as well as averaged over all categories.

The computed errors for the color spaces of interest are shown in Figure 9. We get confirmation that the L*a*b* space performs best in terms of the extent of the color categories (leftmost bar of the left part of the figures, marked "ALL"). For each of the categories individually except yellow, the error is less in L*a*b* space than in the other two spaces, and for black, gray, and white (i.e. the gray axis), the error is zero in L*a*b* space but not in the other two spaces. Between the RGB and NPP spaces the differences are smaller, with the overall error virtually the same. For some categories such as pink and white the error is smaller in NPP space, for others such as red and yellow the error is smaller in XYZ space. In terms of the error of focus location, the three spaces are comparable overall (leftmost bar marked "AVG" in the figures), with some per-category variation among the spaces.

When comparing performance, it is important to note that both the XYZ and the L*a*b* spaces are psychophysical ones, based on color perception experimental data averaged over large populations. In contrast, the NPP color space is based on neurophysiological recordings made from a single individual (and a monkey at that). As such it should be representative of a particular individual, but not necessarily of a population mean. The color naming data of Berlin and Kay is also averaged over a number of experimental subjects, so it is likely that theirs would be a better fit with the CIE spaces than with the NPP space, when considering only the way both color spaces and naming data were constructed or collected. The categorial model itself is of course also a psychophysical model, based on experimental observations over a wide range of tasks and sizable populations [19], so again we would expect similarly constructed color spaces to be at an advantage. The model works best on the L*a*b* space, which is meant to be a perceptually equidistant space, meaning that a fixed distance between two points anywhere in the space should correspond to the same magnitude of perceptual difference. This is not surprising giving the categorial model, which assumes a Euclidean distance measure on the underlying space, i.e. the perceptual differences should be equal in all directions and in all areas of the space.

The error metric as presented does give us a quantitative basis for comparing the performance of different color spaces, but it is a less than perfect one. For instance, it does not take into account how

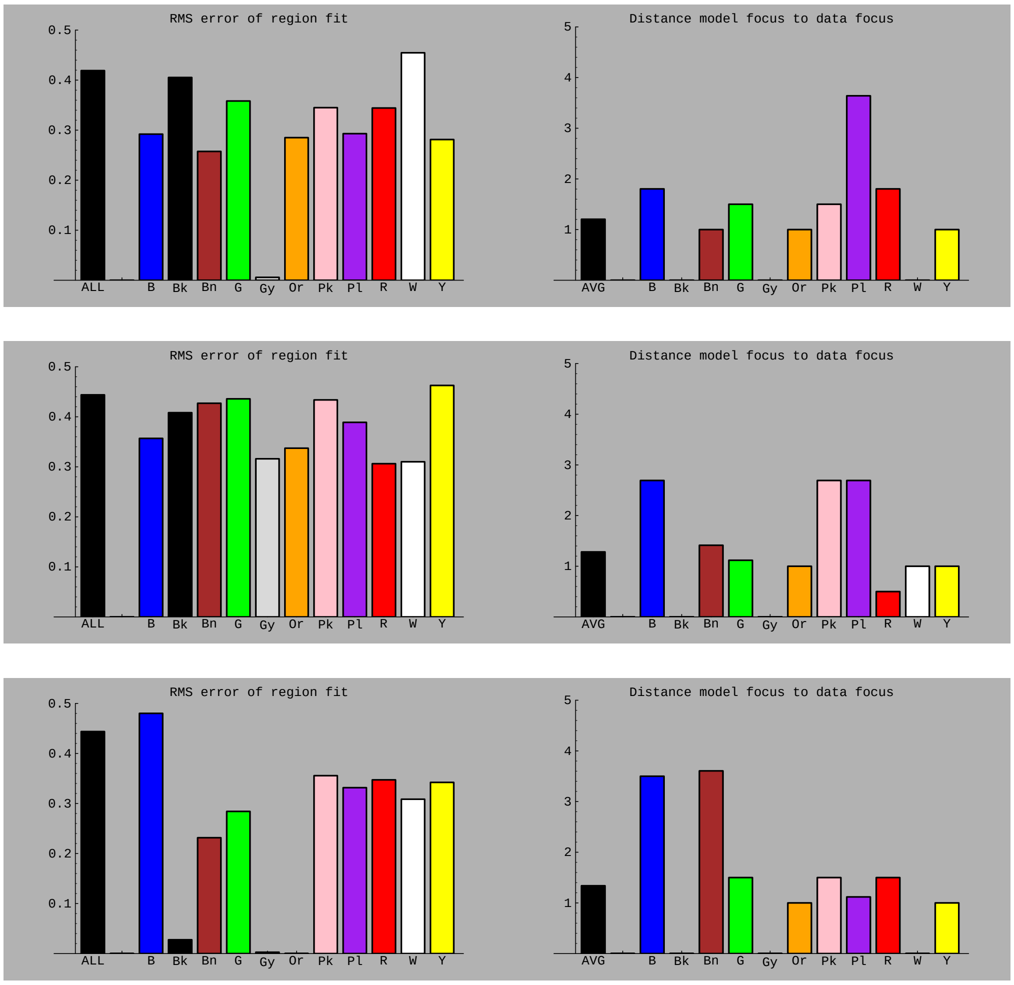

Figure 9: The error of fit of the model categories relative to the Berlin and Kay data, for the XYZ, L*a*b*, and NPP spaces (top to bottom). Left: RMS error of fit of the extent of the categories, right: distance of projected model foci to data foci. Averages over the set of stimuli/categories is shown in black, as the first bar in each plot.

"serious" a categorization error is: categorizing very pale blue as white is not as serious as categorizing green as red. It is difficult to include such qualitative judgments into the error metrics, but doing so might give somewhat different results.

4 An Application and Some Empirical Results

4.1 Naming Color Samples

The complete process required for naming the color of an object in one's field of view is schematized in Figure 10. The sensing device typically outputs RGB values for each pixel in its sensor array, which are transformed into the color space coordinates of choice. A blob of a uniform color in the field of view thus corresponds to a point in color space, $\overline{p}$. We then determine its membership value with respect to each of the defined color categories $C_i \in \mathcal{C}$:

$$\begin{aligned} \mathcal{A} &= \{\langle C_i, \alpha_i \rangle \mid C_i \in \mathcal{C}\} && (10) \\ \alpha_i &= G_n(\overline{p}, \overline{\mu_i}, \sigma_i) && (11) \\ C_i &= \langle \overline{\mu_i}, \sigma_i, l_i \rangle && (12) \end{aligned}$$

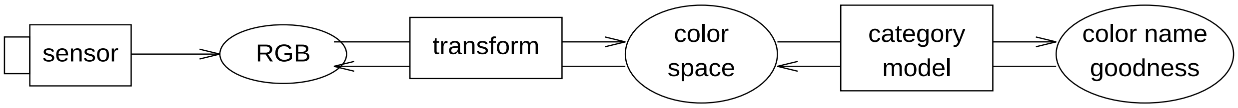

Figure 10: Schematic diagram of the color naming process. Boxes represent devices or operations, ellipses represent data structures. Directed line segments represent data transformations.

where $\mathcal{A}$ is the set of all category-membership value pairs, α_i is the membership value for category C_i, and $\mathcal{C}$ is the set of all defined (non-null) categories. G_n is the normalized Gaussian function with the corresponding parameters μ (the focus) and σ (the extent), and l_i is the label (name) associated with category C_i. Next we select all categories for which the membership value exceeds the threshold for category membership θ, and the corresponding membership values:[6]

$$\mathcal{B} = \{\langle C_i, \alpha_i \rangle \mid \langle C_i, \alpha_i \rangle \in \mathcal{A} \wedge \alpha_i \geq \theta\} \tag{13}$$

More than one membership value may exceed the threshold, or there may be none (as the combined extent of the categories does not cover the entire color space, at least when thresholded). The former is particularly the case for partly overlapping categories such as red and orange (Figure 2). Next we sort the candidates by decreasing membership value, and the first element of the sorted n-tuple is our categorization judgment with the corresponding membership value. The name corresponding to the perceived color is then simply $3^{rd}(C_1)$.

Using a particular threshold value θ amounts to doing a free choice naming experiment with 11 named basic color categories and one null category for everything that does not exceed any threshold. Using a zero threshold amounts to a forced-choice experiment where we will always assign one of the 11 named categories to the stimulus, regardless of how low the best membership value might be. This technique has been used by [10] for instance, using only the four primary colors red, green, blue, and yellow, to obtain psychophysical measurements of the color-opponent processes assumed to underlie color perception.

Once we have the sorted list of candidates $\mathcal{S}$, we can not only return the name of the best candidate category, but also a complex color name, constructed from the names of the top two candidates. A tentative algorithm for constructing complex color names based on multiple category membership values is presented in [14].

4.2 Pointing Out Examples of Colors

The complete process required for pointing out an object of a named color in one's field of view is schematized in Figure 11. We now also need a pointing device, a complete color image rather than just a blob,

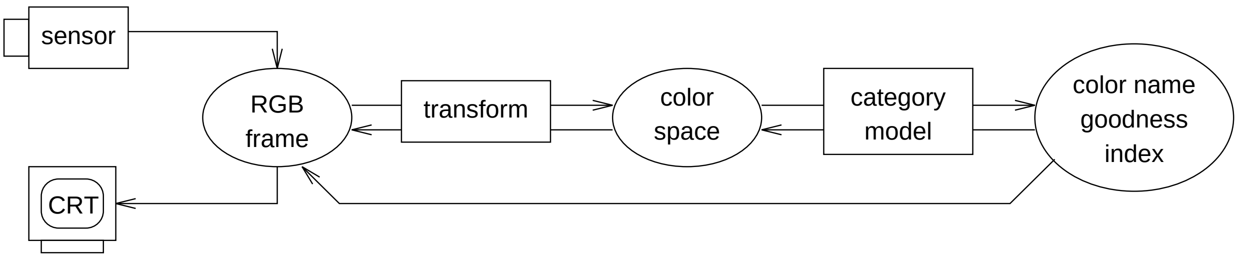

Figure 11: Schematic diagram of the color pointing-out process (cp. Fig. 10).

and an index into the image for each blob to be categorized.

When given a color name N to point out an example of, we sample the image using a predefined blob size, collecting the averaged device RGB values and the image coordinates of the center of each blob, and transform the RGB values to the color space of choice, which gives us a set of points in color space

[6] The description of the various steps is mainly for explanatory purposes, and need not reflect an actual implementation strategy. The first two steps can easily be combined into one, for instance.

with corresponding image coordinates. We then determine the membership value of the samples in the appropriate category, and select all samples with a membership value exceeding the threshold for category membership θ. As before, we sort the candidates by decreasing membership value. If we just want any referent for the category named N, we pick an arbitrary element $r \in \mathcal{S}$ from this tuple (or from set B), and if we want the best example we take the first element $r = 1^{st}(\mathcal{S})$. Finally, all that is left to do is point out the location of the sample in the image (or in the world giving rise to the image, which is harder to do), using the index vector $2^{nd}(r)$.

As is the case with color naming, we can do forced choice experiments if we use a zero threshold, in which case a referent will always be selected regardless of how good an example it is, or we can do free choice experiments using the standard threshold θ, in which case a referent will only be selected if it is a "good enough" example of the category in question.

4.3 Choosing Objects by Color

When combined with an image segmentation algorithm, we can use the same techniques to choose objects by color. An interesting observation in this respect is that the color perception mechanism need not be as good to perform this task (discrimination) as to perform the naming or pointing tasks (identification). The categories may be considerably "wider" and more overlapping, or the system's idea of what constitutes a particular color may vary to some extent compared to the agent issuing the request, as long as the colors of potential referents are distributed widely enough throughout the color space. This task is probably best performed with $\theta = 0$, to avoid quibbles over whether or not a particular object color is a good enough example of the requested kind, i.e. to be maximally cooperative.

4.4 A Simple Application

The application consists of two separate parts, one concerned with selection and display of samples from images, and the other concerned with the actual color perception and categorization model (Figure 12).

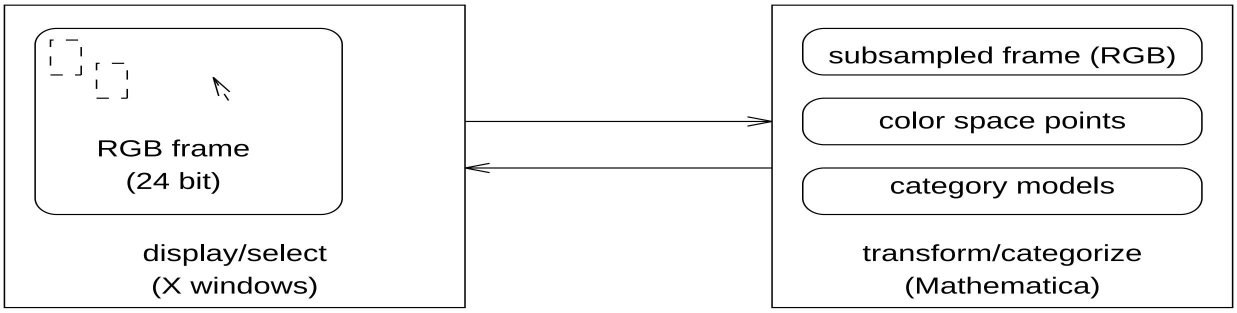

Figure 12: Outline of the color naming/pointing-out/selecting application, consisting of a display and select part (left), implemented as an X windows client, and a transformation and categorization part (right), implemented in Mathematica code. The two parts communicate asynchronously via a simple file protocol.

The display program runs under X windows, and allows one to display a 24-bit RGB image. It also lets one select samples of a certain pre-defined size (currently 12×16 pixels) from the image using the mouse, which will be passed to the categorization program. It can subsample the entire image using the same blob size, and pass the result to the categorization program. Finally, it can draw boxes around blobs, whose center coordinates it gets from the categorization program.

The categorization program is a collection of Mathematica functions that can (1) *Name* the color of a blob pointed to on the image, and provide a membership value, (2) *Point out* examples of a named color category on the image, and provide a membership value, and (3) *Select* one from a number of samples whose color best fits a named category. The names returned can be simple or complex, and the best n candidates can be returned. The category membership threshold θ may be specified, as well as the underlying color space to use. The names specified for the pointing-out function can be simple or complex as well, and the threshold θ can be specified, as well as the color space to use. The function can either return any n examples (the first exceeding the threshold) or the best n examples. The select function always points to the best example of the specified category within the set of samples provided, using a specifiable underlying color space. It will ask the user to provide it with a set of samples from the image first, to get around the absence of any image segmentation or object recognition algorithms.

4.4.1 Constancy

To make the color naming and pointing algorithms work in practice, with real images, we can no longer avoid the issue of color constancy. The approach we have adopted to deal with this problem is of a rather pragmatic nature, and we make no claims about its generality. Nevertheless, the results are interesting, and the approach seems to be fairly robust with respect to different lighting conditions, and to some extent even with respect to different sensing devices. It is presumably related to the "white balance" algorithms implemented in home video cameras and the like.

The first thing that needs to be done is to gamma correct the image, to make intensities perceptually linear. Video cameras (especially home camcorders) often incorporate gamma correction, but scanners may not, or may allow it to be specified under software control. This step is required since the CIE XYZ standard assumes perceptually linear intensity responses, and also for display on computer (or TV) monitors. We used a gamma correction factor of 2.2, if any.

Next we need to compensate for differences in lighting, both with respect to intensity and with respect to spectral characteristics. Although the spectrum of the light source can never be completely recovered from the image, we can use Von Kries adaptation with good results, as long as the spectrum of the light source is not too wildly skewed or irregular. In real camera (and scanner) images it turns out that not only is white not white (i.e. the RGB values are not equal and unity, viz. $\langle 1,1,1\rangle$), but black is often not black either (i.e. the RGB values are not equal and zero, viz. $\langle 0,0,0\rangle$). To compound the problem, lighting may vary considerably across a single image.

The "constancy" algorithm we have developed searches the image for likely representatives of white and black, and then uses those do to a modified Von Kries adaptation:

$$\overline{p'} = \frac{\overline{p} - \overline{b}}{\overline{w} - \overline{b}} \tag{14}$$

where $\overline{p'}$ is the "adapted" pixel value, $\overline{p}$ is the original pixel value, $\overline{b}$ is the black representative, and $\overline{w}$ is the white representative. Finding the black and white representatives works as follows (regarding an image as a set of pixels for convenience):

$$\begin{aligned}
\mathcal{I} &= \{\langle r,g,b\rangle\} && (15)\\
\mathcal{S} &= \{\langle\langle r,g,b\rangle, r+g+b\rangle \mid \langle r,g,b\rangle \in \mathcal{I} \wedge \sqrt{(r-g)^2+(g-b)^2+(b-r)^2} < \epsilon\} && (16)\\
\overline{b} &= 1st(\min_{2nd} \mathcal{S}) && (17)\\
\overline{w} &= 1st(\max_{2nd} \mathcal{S}) && (18)
\end{aligned}$$

where $\mathcal{I}$ is the set of image pixels, represented as normalized RGB vectors $\langle r,g,b\rangle$ in $[0,1]$, $\mathcal{S}$ is the subset of $\mathcal{I}$ containing potential gray axis pixels, and $\overline{b}$ and $\overline{w}$ are the chosen black and white representatives, respectively. The criterion used to select potential gray axis pixels is another RMS measure, enforcing a summed square error between pixel components of less than ϵ. We have obtained good results with $\epsilon = 0.01$. The black and white representatives, then, are the potential gray axis pixels with the lowest and highest brightness, respectively, as determined by a simple sum-of-components measure. A visual way to think of this process is that it stretches the image's actual gray axis and realigns it with the theoretical gray axis for perfectly homogeneous and flat-spectrum lighting, i.e. the line segment $[\langle 0,0,0\rangle, \langle 1,1,1\rangle]$. The procedure is illustrated in Figure 13.

4.5 Some Empirical Results

The best way to get a good feel for how the algorithm performs is to run it for oneself (See Appendix A for downloading instructions). Barring that, the best we can do is to look at an image and the judgments the algorithm makes on it. We must caution against taking the color reproduction on paper too seriously:[7]

[7] If the figures are reproduced in color at all.

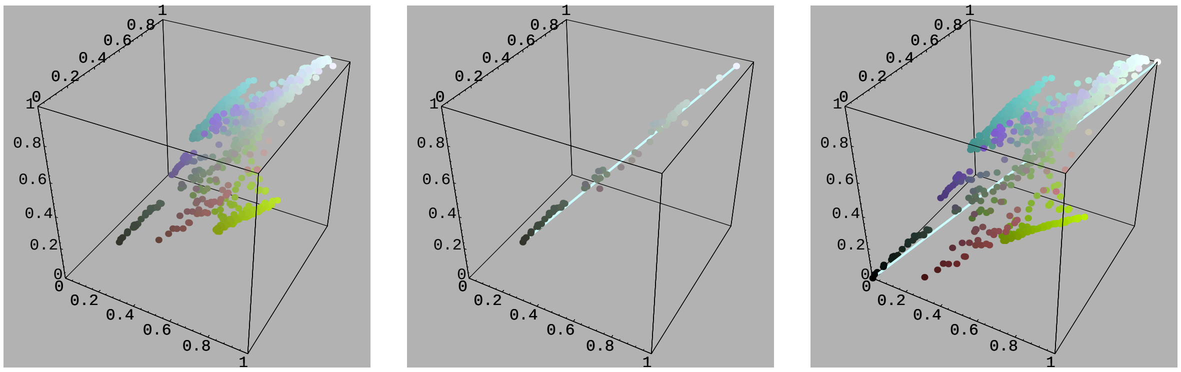

Figure 13: An illustration of the adaptation procedure, run on the image from Figure 14. Left: the image samples in RGB space. Middle: the set of potential gray axis pixels $\mathcal{S}$, with a line connecting the black and white representatives $\overline{b}$ and $\overline{w}$. Right: the set of adapted image samples, and a line connecting the adapted black and white representatives.

too many unknowns are involved, but it will do for a rough impression. Figure 14 shows a particular test image (which has not been involved in "training" or optimizing the category models at all) taken with a hand-held regular home camcorder whose composite NTSC[8] signal was hooked up to a cheap Sun VideoPix frame grabber. This equipment is certainly not of lab grade quality, not calibrated, and the lighting for the scene was just the available office TL lighting, not controlled or adjusted in any way.[9] The image shows some toy utensils and other objects of various colors, none particularly good examples of basic colors.

Due to space limits we cannot analyze the results in detail, for which we refer to [14], so we will merely summarize them here. The color space used is L*a*b*, and the constancy algorithm is applied as desribed above. Thin-bordered rectangular boxes with numeric labels represent input to the naming function, first in simple name mode with a membership threshold of 0.35. The responses are (name, membership value): (1) black, 0.57; (2) white, 0.87; (3) white, 0.65; (4) white, 0.55; (5) none; (6) gray, 0.87; (7) green, 0.38; (8) green, 0.38; (9) yellow, 0.45; (10) blue, 0.41; (11) blue, 0.41; (12) blue, 0.40; (13) blue, 0.42; (14) blue, 0.42; (15) blue, 0.41; (16) gray, 0.44; (17) gray, 0.41. Note the different membership values for the white samples; the "none" value returned for sample 5, meaning it is not a member of any of the known categories (to a large enough extent); the ambiguity between yellow and green of the bowl in the upper left; and the two kinds of blue of the plate and knife/fork/spoon. The following responses are for the same samples, but now in complex name mode (name, (secondary, primary membership value)): (18) somewhat green-ish black, (0.34, 0.57); (19) somewhat pink-ish white, (0.32, 0.88); (20) somewhat blue-ish white, (0.30, 0.65); (21) somewhat blue-ish white, (0.31, 0.53); (22) none; (23) blue-ish gray, (0.35, 0.87); (24) green, 0.38; (25) somewhat yellow-ish green, (0.34, 0.37); (26) green-ish yellow, (0.42, 0.44); (27) somewhat gray-ish blue, (0.26, 0.41); (28) somewhat gray-ish blue, (0.30, 0.41); (29) somewhat gray-ish blue, (0.26, 0.40); (30) blue, 0.42; (31) somewhat pink-ish blue, (0.27, 0.42); (32) somewhat pink-ish blue, (0.29, 0.41); (33) somewhat pink-ish gray, (0.28, 0.43); (34) pink-ish gray, (0.37, 0.44). Note the qualifiers for the white and yellow/green and blue samples. Next the pointing-out function is called, which in response to a name draws boxes around the best three examples of the corresponding category in the image. A low threshold of 0.01 is used, to make the algorithm very cooperative. Only the best examples are indicated in the figure by thick-bordered boxes labeled with the abbreviated names: Wh=white, Bk=black, R=red, G=green, Y=yellow, B=blue, Br=brown, Pk=pink, Pl=purple, Or=orange, and Gy=grey. Composite names are indicated by a modifier category preceding the main category, e.g. Y-G for yellowish green. Note that X-Y does not necessarily refer to the same thing as Y-X, and that some samples are best examples (within the image) of more than one category, e.g. the black/brown sample in the upper left corner.[10] Finally, the object-selection function is called, with the same threshold of 0.01. The samples to choose from are

[8] Never The Same Color.

[9] TL lighting is in general pretty bad for color rendition, as becomes most painfully obvious in front of institutional bathroom mirrors.

[10] In the "aperture mode" of viewing, e.g. using CIE chromaticity values only, brown would not be distinguishable from red and/or orange.

Figure 14: A frame grabbed from a regular home camcorder, in ordinary office lighting conditions. Note the shadows and the varying brightness across the background, and the fuzziness of the objects' outlines. The objects are placed on a regular piece of white paper, on a real simulated woodgrain desk top.

collected from each of the objects in the image as well as the paper background and the desk top, and are indicated by circles in the figure. The resulting selections from the set of samples in response to a category label are indicated in the figure by the labels accompanying some of the circles. Note that given the limited choice, all of gray, red, purple, and pink refer to the same sample, viz. the one from the cup in the upper right of the image, albeit with different membership values.

Some tests on rather different images, for instance color cartoons acquired with a scanner, show very comparable results.

5 Conclusion

We have presented a computational model of color categorization using a normalized Gaussian function of perceptual color space coordinates as the basic category model. This models allows an application to *name, point out, and select* colors in images, and hence in its environment, and provide a *goodness* or *membership* value for each categorial judgement. We have experimented with different color spaces, including a novel one derived from neurophysiological measurements, and have found the CIE L*a*b* space to provide the best categorization results. We have also presented a working application based on the computational model, which is to some extent able to deal with the *color constancy* problem in real images made in an unmodified office environment with low-grade uncalibrated equipment.

The novelty of the approach lies in the fact that it models *graded* or *fuzzy* color categories as found in anthropological, psychological and psychophysical color perception work, and relates these categories in a simple way to the symbolic labels that natural languages use for them. As such the approach can be seen as a particular case study in *symbol grounding*, *embodiment*, or *situated cognition*.

Although the application presented in the last section is an interesting demonstration of what can be accomplished with the color perception and naming model, it is certainly open for improvement. For instance, it would be interesting to explore an active version of the adaptation algorithm that could move the camera around searching for appropriate lighting conditions, and to make the adaptation tracking over time.

A far more interesting, although much more difficult, problem is how to extend a similar approach to other categories of perception (and action). Some thoughts about this can be found in [15].

A Getting the software

The software for the application described in Section 4, both the select/display X windows client and the Mathematica code, is available by anonymous ftp from

`ftp.cs.buffalo.edu:/ftp/pub/colornaming`

All software is provided as is, without any express or implied warranty. If you have problems with it, feel free to contact the author.

References

[1] B. Berlin and P. Kay. *Basic Color Terms: Their Universality and Evolution.* University of California Press, Berkeley CA, 1991 edition, 1969.

[2] F. Birren. *Munsell: A grammar of color.* Van Nostrand Reinhold, New York, 1969.

[3] R. M. Boynton. *Human Color Vision.* Holt, Rinehart and Winston, 1979.

[4] R. M. Boynton. Human color perception. In K. N. Leibovic, editor, *Science of Vision*, chapter 8, pages 211–253. Springer, New York, 1990.

[5] C. M. Brown. Color vision and computer vision. Technical Report 108, University of Rochester, Computer Science department, June 1982.

[6] J. E. Cairo. *The Neurophysiological Basis of Basic Color Terms.* PhD thesis, State University of New York at Binghamton, 1977.

[7] R. L. De Valois, I. Abramov, and G. H. Jacobs. Analysis of response patterns of LGN cells. *Journal of the Optical Society of America*, 56(7):966–977, July 1966.

[8] S. Harnad. The symbol grounding problem. *Physica D*, 42(1-3):335–346, 1990.

[9] H. Hexmoor, J. Lammens, and S. C. Shapiro. Embodiment in GLAIR: A grounded layered architecture with integrated reasoning. In *Florida AI Research Symposium*, pages 325–329, 1993. Also available as TR-93-10, CS dept., SUNY/Buffalo.

[10] L. M. Hurvich and D. Jameson. An opponent-process theory of color vision. *Psychological Review*, 64(6):384–404, 1957.

[11] P. Kay and W. Kempton. What is the Sapir-Whorf Hypothesis? *American Anthropologist*, 86:65–79, 1984.

[12] P. Kay and C. K. McDaniel. The linguistic significance of the meaning of Basic Color Terms. *Language*, 54(3):610–646, 1978.

[13] G. Lakoff. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* University of Chicago Press, Chicago, IL, 1987.

[14] J. M. Lammens. *A Computational Model of Color Perception and Color Naming.* PhD thesis, State University of New York at Buffalo, 1994. Also available as TR-94-26, CS dept., SUNY/Buffalo.

[15] J. M. Lammens, H. H. Hexmoor, and S. C. Shapiro. Of elephants and men. In *Proceedings of the NATO ASI on Intelligent Autonomous Agents (Trento, Italy).* Springer Verlag, 1994. (To appear. Preprint available CS TR-93-13, CS dept., SUNY/Buffalo).

[16] L. T. Maloney. Color constancy and color perception: The linear-models framework. In D. E. Meyer and S. Kornblum, editors, *Attention and Performance XIV*, chapter 3, pages 60–78. MIT Press, 1993.

[17] L. T. Maloney and B. A. Wandell. Color constancy: a method for recording surface spectral reflectance. *Journal of the Optical Sociey of America*, pages 29–33, 1986.

[18] S. M. Newhall, D. Nickerson, and D. B. Judd. Final report of the OSA subcommittee on the spacing of the Munsell colors. *Journal of the Optical Sociey of America*, 33(7):385–418, July 1943.

[19] R. N. Shepard. Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323, September 1987.

[20] L. A. Suchman. *Plans and Situated Actions: The Problem of Human Machine Communication.* Cambridge University Press, 1988.

[21] G. Wyszecki and W. S. Stiles. *Color Science.* Wiley, New York, 1982.

[22] L. A. Zadeh. Quantitative fuzzy semantics. *Information Sciences*, 3:159–176, 1971.