

How Strong Metamerism Disturbs Color Spaces

William A. Thornton

27 Harvard Road, Cranford, NJ 07016

Received 4 December 1997; accepted 18 April 1998

*Abstract: Systems for arranging and describing color include "color spaces" and "color order systems." In a color space, tristimulus values R, G, and B are computable for every light (every point in the space). In familiar color spaces, such computation makes use of three functions of wavelength (the color-matching functions that define one of the CIE Standard Observers), one function corresponding to each of R, G, and B. In the presence of strong metamerism (marked spectral difference between the spectral power distributions of a pair of visually matching lights), the color-matching functions may report that one light of the pair has an entirely different color from that of the other member of the visually matching pair of lights. The CIE Standard Observer embodying those color-matching functions "sees" the two visually matching lights as entirely different in color, that is, it reports entirely different sets of R, G, and B for the two visually matching lights, and, thus, an entirely different chromaticity. In an example given here, each of the CIE Standard Observers assigns a strong green color to lights that are seen by normal human observers as a visual match to a hueless reference white. On the other hand, color order systems comprising sets of real objects in a specified illuminant, and which are assembled (visually arranged) by normal observers, as are the Munsell and OSA sets, do not suffer from the type of trouble discussed here. Color spaces depending on mathematical functions of R, G, and B are at risk: both Standard Observers are shown to plot visually identical lights at widely varying points in familiar color spaces (e.g., $\Delta E^*_{ab} = 40\text{--}50$). © 1998 John Wiley & Sons, Inc. Col Res Appl, 23, 402–407, 1998*

Key words: color order systems; color spaces; strong metamerism; Standard Observer error; breakdown of order in the color space

INTRODUCTION

At the 1997 Annual Meeting of the Inter-Society Color Council, a discussion was held entitled, "The Colors of Colored Things," on aspects of basic and applied research in the field of color order systems. This article is based on a contributed lecture on that subject.

Billmeyer and Saltzman¹ distinguish between collections of physical samples and systems that are not based on actual samples. The former are now designated color order systems, referring to "a rational method or plan of ordering and specifying all object colors, or all within a limited domain, by means of a set of material standards selected and displayed so as to represent adequately the whole set of object colors under consideration."^{2,3} In this definition, the term "object colors" refers to the set of lights reflected from the objects (material standards) while those objects are in a specified illumination. The specified illumination, together with the spectral reflectance of the object, determines the reflected light: the spectral power distribution (SPD) of the illumination is multiplied by the spectral reflectance of the object, wavelength by wavelength, to obtain the SPD of the reflected light — the "object color." Thus, basically, the colors in a color order system are these reflected lights, to each of which is assigned a point in the system. A system not based on physical samples may be termed a color space, "a geometric space, usually of three dimensions, in which colors are arranged systematically."³ "Colors" are again reflected lights, if *object* colors are meant in this definition also. This article deals with a problem afflicting color spaces.

Like color order systems, color spaces are, of course, distinctively visual. They owe every useful characteristic to an attempt to mirror the complex manner in which the normal human visual system works. Much of that complexity is so far poorly understood. Improvement of that understanding will eventually allow color spaces to approach complete (visual) dependability; that is, complete accord with what is seen by a normal human observer. The purpose of this article is to call attention to a little-recognized

weakness in present color spaces (of which the CIE systems are “by far the most important”¹⁾) that detracts from their dependability. The weakness is in \bar{r} , \bar{g} , and \bar{b} , the set of three functions of wavelength defining one or the other of the CIE Standard Observers.* The SPD of each light is multiplied, wavelength by wavelength, by one of these functions to yield tristimulus values R, G, and B for each light in the color space. This weakness has across-the-board importance in color science, because the three functions are always, at least implicitly, taken to represent the three spectral sensitivities of the normal trichromatic human visual system.

In the last ten years, experimental visual color-matching has revealed⁴ that traditional color-matching functions may completely fail in their role of representing normal human vision: in the presence of strong metamerism (marked spectral difference between the spectral power distributions of a pair of visually matching lights), the traditional color-matching functions may report one light of the pair as having an entirely different color from that of the other member of the visually matching pair of lights. Otherwise stated, the CIE Standard Observer embodying those color-matching functions “sees” the two visually indistinguishable lights as entirely different in color. Obviously, in a useful color space, lights that appear the same (visually match) must occupy essentially the same point in the color space, i.e., the values of R, G, and B computed via the color-matching functions for the two visually matching lights must be essentially the same: $R_1 = R_2$, $G_1 = G_2$, and $B_1 = B_2$. Quantities R, G, and B (which I shall call the red content, green content, and blue content of a viewed light) also are the basis of other popular color spaces such as CIELAB.

Unlike color order systems comprised of collections of physical samples, each currently important color space demands calculation of RGB for every point in it. In the CIE Standard Observer systems, and related systems such as CIELAB, RGB are calculated by multiplying the SPD of a constituent light by the appropriate function from \bar{r} , \bar{g} , and \bar{b} , wavelength by wavelength, and summing over the visible spectrum, as follows:

$$\text{red-content } R = \text{Sum of lightSPD} \times \bar{r}$$

$$\text{green-content } G = \text{Sum of lightSPD} \times \bar{g}$$

$$\text{blue-content } B = \text{Sum of lightSPD} \times \bar{b}$$

Then, for example, the coordinates of the light, in the CIE chromaticity diagram, are:

* I point out in this article, and in the articles of Ref. 4, that both CIE Standard Observers and their associated color spaces are shown to be capable of large and troublesome tristimulus errors (large differences in tristimulus values computed for visually matching lights). It is necessary, therefore, to speak of color-matching functions and their associated color spaces generally, to include both the present flawed examples (the familiar CIE ones that are troublesome) and improved varieties soon to come; I use R, G, B and \bar{r} , \bar{g} and \bar{b} to do so.

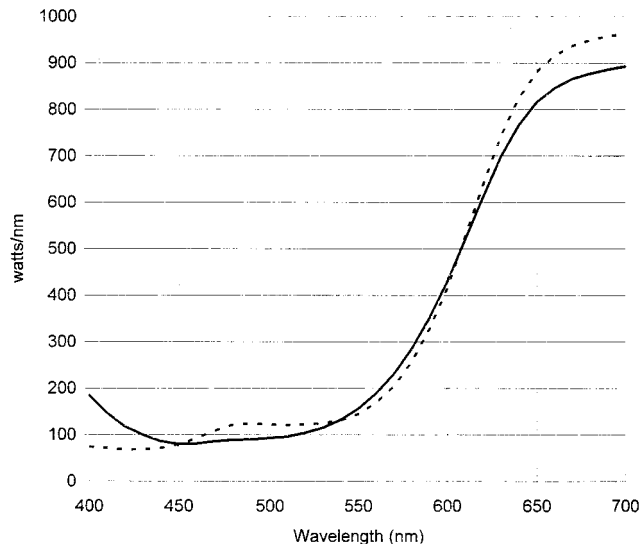


FIG. 1. Spectral power distributions of two reddish lights that are to normal human observers an approximate visual match: an example of rather weak metamerism, defined as the case in which the spectral composition of one metamer (a light entering the visual system) is little different from that of a second metamer that visually matches the first.

$$g = \frac{G}{R + G + B} \quad \text{and} \quad r = \frac{R}{R + G + B}, \quad \text{and,}$$

in the CIELAB diagram, are $a^* = f(R, G, B)$ and $b^* = f'(R, G, B)$, that is, different functions of RGB.

STRONG METAMERISM

Given a single normal human observer and a certain constant visual environment, the appearance of a viewed light depends upon its SPD. But, even with that single observer and visual environment, there is an enormous variety of lights of different SPD, each of which yields the same appearance to that observer in that environment. Members of this group of visually matching lights are called metamers. “Weak metamerism” refers to the cases in which the spectral composition of one metamer is little different from that of a second metamer. The SPDs of a pair of rather weakly metameric visually matching lights are shown in Fig. 1. “Strong metamerism” refers to cases in which the difference is great. For a color space to be broadly applicable, the order of an array of lights in that color space should be robust to occurrences of strong metamerism. That is, as plotted in that color space, two lights that appear the same to a normal human observer should occupy essentially the same point of the color space no matter how spectrally dissimilar their SPDs. Let me illustrate with one example a case of strong metamerism badly mishandled by two familiar color spaces.

For centuries, coloring materials of all kinds (inks, paints, dyes, pigments, and so on) have been improving so as to extend the usable gamut of coloration to brighter and more intense colors. Suppose that the present year is in the 21st

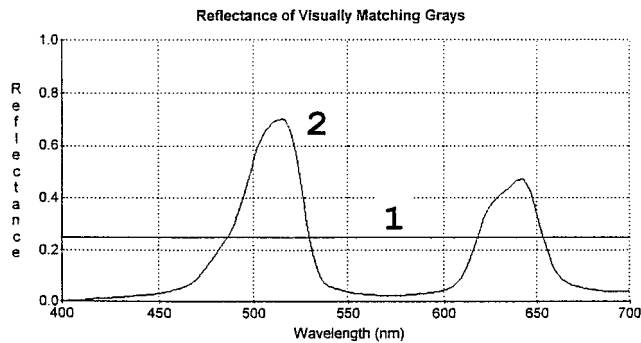


FIG. 2. Hypothetical reflectance spectra of two gray objects. The white lights reflected from the objects in sunlight are an approximate visual match to normal human observers. These reflected lights are an example of strong metamerism, defined as the case in which the spectral compositions of the visually matching lights are markedly different.

century. The chemistry of colorants has continued to become more sophisticated. The rise and fall of reflectance spectra of colorants have become sharper. In Fig. 2 are shown hypothetical reflectance spectra of two gray objects 1 and 2 that would visually match* to normal human observers in average daylight, once the reflectance spectra are reduced to practice. The average daylight spectrum is, from the point of view of the human visual system, quite smooth and flat. The reflectance spectrum of object 1 is smooth and flat. Therefore, the spectrum of the white light reflected from and entering the eye from object 1 will be smooth and flat. In contrast, object 2 reflects predominantly blue-green light (due to the left-hand concentration of reflectance centered near 510 nm) and deep red light (due to the right-hand concentration of reflectance centered near 640 nm). In average daylight, object 2 looks the same as object 1 to a normal human observer. This means that the white light reflected from object 2, although it is a mixture only of reflected components of blue-green light and deep-red light, looks the same as the simpler spectrum of light from object 1; the two objects visually match. The shapes of the two spectra of Fig. 2 approximate the spectral power distributions of the two visually matching white lights reflected from the objects.

STRONG METAMERISM AND THE CIE STANDARD OBSERVERS

In colorimetry, the CIE Standard Observer is meant to substitute for the normal human visual system. The two white lights reflected from objects 1 and 2 of Fig. 2 are very different in their spectral composition, but match visually to

* Reflectance spectrum 2 of Fig. 2 is the spectral shape of the SPD of a white light that, by consensus of our 8 normal human observers, matches visually the reference daylight-color white light always in the bottom field of our visual colorimeter. Thus, that reflectance, when physically achieved and viewed in average daylight by a normal human visual system, will appear a hueless gray.

normal human observers. Does the Standard Observer report that their red, green, and blue contents R, G, and B calculate to be the same? That is, do the two lights match also to the Standard Observer, as they are expected to do?

The answer is, "No." To the CIE Standard Observer, the two lights appear very different. In Fig. 3 is reproduced a small area strongly colored green. This color and its white background, illuminated by daylight (hold your copy of CR&A to the window), are good approximations to how either CIE Standard Observer "sees" objects 2 and 1, respectively, also in daylight. It is clear that the two white lights reflected from the objects of Fig. 2, which visually match to normal human observers, are seen as very different to a CIE Standard Observer. In other words, the two chromaticities computed by means of the Standard Observer color-matching functions \bar{r} , \bar{g} , and \bar{b} are very different, one of them representing a strong green color and the other a white color, assuming adaptation to average daylight.

Let me now support these statements. For ten years my associate researchers and I have been documenting⁴ instances in which the CIE color-matching functions \bar{r} , \bar{g} , and \bar{b} defining the Standard Observer, lead, in the presence of strong metamerism, to large errors in computed tristimulus values RGB of lights that match to our normal human observers. In all of the relevant parts of our visual experimentation, the same reference white light is always presented to the (normal human) observer in the bottom visual field, corresponding approximately to daylight illumination of object 1 of Fig. 2. The strongly metameric visually matching white light is always in the top field, corresponding to daylight illumination of object 2 of Fig. 2.

From this work, some particular examples of visual matches are taken. Part of the 1931 CIE chromaticity diagram is shown in Fig. 4, including the Planckian locus, the plotted chromaticity of CIE illuminant D65, and the MacAdam 4-step ellipses used in the lighting industry to specify the chromaticities of the common fluorescent lamplights Daylight and Cool White, all to show the scale of that part of the color space. Point A in Fig. 4 marks the chromaticity of the reference white light always in the bottom visual field; the five solid squares constitute a series of white-light mixtures each visually matching, in a 10° field, the constant reference white light (point A); the topmost black square marks the most deviant chromaticity of that series of visu-



FIG. 3. An area of colored paper that, illuminated by daylight, with its white surround, well approximates the appearance, to one or other of the CIE Standard Observers, of objects 2 and 1 of Fig. 2. That is, to a normal human observer, the lights reflected in daylight by objects 2 and 1 are an acceptable white-light match, while to either CIE Standard Observer, they appear as strong green and white, respectively.

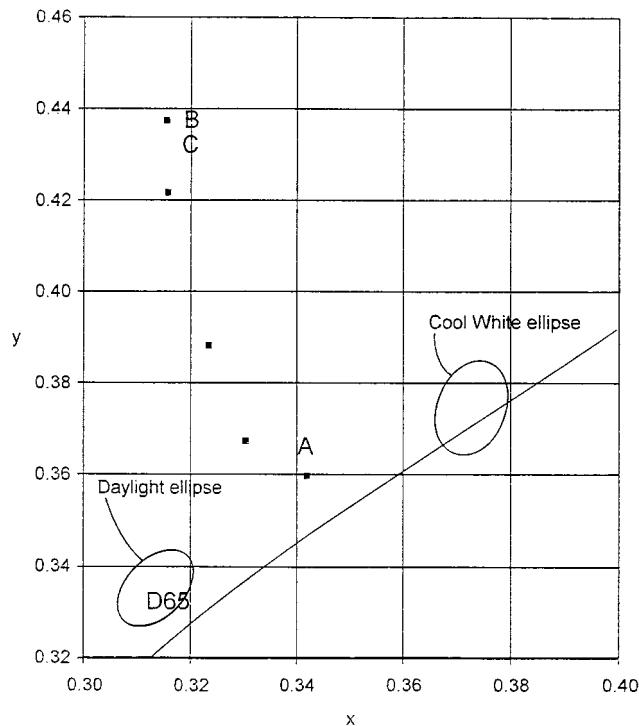


FIG. 4. Part of the 1931 CIE chromaticity diagram, including the Planckian locus, the plotted chromaticity of CIE illuminant D65, and the MacAdam 4-step ellipses used in the lighting industry to specify the chromaticities of the common fluorescent lamplights Daylight and Cool White, all to show the scale of that part of the color space. Symbols A, B, C, D65 are not labels; each symbol locates the computed chromaticity of the light it represents. Point (A): the computed chromaticity of the reference white light. Solid squares: a series of five white lights each visually matching, in a 10° field, the constant reference white light (point A); the topmost black square marks the most deviant chromaticity of that series of visually matching white lights. Point (B): the chromaticity of an averaged white light (composed of blue-green and deep-red components) pronounced a visual match to the reference white by the 8 observers of 1994 (reference 3, Part IV), using a much brighter 1.3° visual field; the SPD of this visually matching light is identical in shape to the “reflectance spectrum” of object 2 in Fig. 2. Point (C): the computed chromaticity of the strong green light reflected from the green colored paper of Fig. 3 in daylight.

ally matching white lights. Point B* marks the chromaticity of an averaged white light (composed of blue-green and deep-red components) pronounced a visual match to the reference white by our 8 observers of 1994, using a 1.3° visual field much brighter than the 10° fields of earlier work. The spectral power distribution of this visually matching light of 1994 (point B) is identical to the “reflectance spectrum” of object 2 in present Fig. 2. It is on this basis that hypothetical object 2 in average daylight will match object 1 to normal human observers.

* The spectral power distributions of these viewed lights of present Fig. 4 appear in Ref. 4 as follows: Point A: Fig. 3 of Part I. Topmost solid square: Lower right in Fig. 10 of Part I. Point B: Upper left in Fig. 108 of Part IV.

Finally, I measured the spectral reflectance spectra of a selection of Color-Aid⁵ colored papers. One of these was the colored paper from which the reproduction Fig. 3 was made. Once its spectral reflectance is multiplied by the SPD of CIE daylight D65, a reflected light results, with chromaticity at point C of Fig. 4, as calculated by the CIE Standard Observer.

Lights A, B, and C are very differently evaluated, on the one hand, by normal human observers and, on the other hand, by the two CIE Standard Observers. Let the symbol “= ” stand for “approximately matches” and the symbol “≠ ” stand for “is a bad mismatch to”. Then, in those terms, lights A, B, and C of Fig. 4 are related as follows:**

For normal human observers:

$$A = B$$

$$B \neq C$$

$$A \neq C$$

and for both CIE Standard Observers:

$$B = C$$

$$A \neq B$$

$$A \neq C.$$

The example shown in Figs. 2–5 represents only one of eight types of strongly metameric lights whose spectral power distributions are presented in Ref. 4, Part IV. Seven of these eight types of spectral composition show tristimulus errors leading, for visually matching lights, to discrepancies in ΔE^*_{ab} from 20–70.

To summarize these experimental results:

1. If Fig. 3 (the green patch) is illuminated by daylight, its computed chromaticity (point C of Fig. 4) is essentially the same as the computed chromaticities of the two visual color-matching results: the comparatively dim top-field white light of the 10° visual field of 1988, marked by the topmost solid square; and the much brighter top-field white light of the 1.3° visual field of 1994, marked by point B.
2. But the dim top-field match of the 10° visual field of 1988, marked by the topmost solid square, and the much brighter top-field match of the 1.3° visual field of 1994, marked by point B, are seen by all the normal human observers as white, as a visual match to the reference white (point A) always in the bottom field of view.
3. However, the top-field light in both cases was “seen” by the Standard Observer as strong green in color. Yet all the normal human observers saw the same two lights as exactly matching the reference white light (point A of Fig. 4).
4. The chromaticity of the 10° top-field white light (top-

** Thanks to Dr. Michael H. Brill for this aid to clarification.

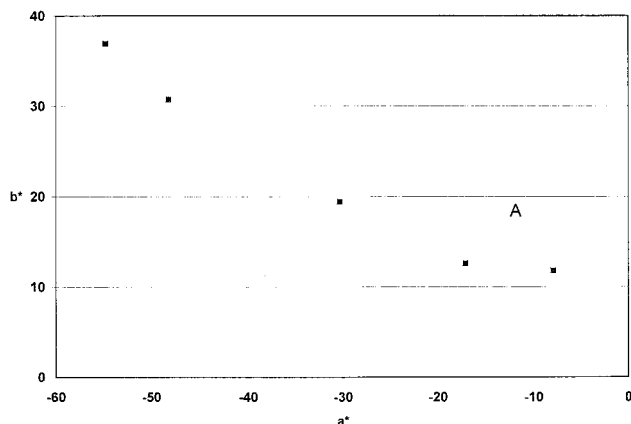


FIG. 5. As Fig. 4, but the CIELAB color space, including plots of six visually identical white lights. The solid squares and the symbol A locate the computed chromaticities of the six lights.

most solid square) was computed by the 1964 CIE 10° Standard Observer, and the chromaticity of the 1.3° top-field white light (point B) was computed by the 1931 2° Standard Observer. It is not legitimate to plot both chromaticities on the same chromaticity diagram, as in Fig. 4. However, the CIE 2° and 10° chromaticity diagrams, if superposed, are seen to be very little different.⁶ The magnitude of the discrepancy between the computed chromaticities of the two top-field lights and that of the visually matching reference white light (point A), is so large as to make the difference in diagrams inconsequential.

5. It follows that *both 2° and 10° CIE Standard Observers “see” the visually white top-field matching lights, of complex SPD strongly metameric to that of the reference white, as strong green.* That is, they report sets of RGB very different from the set reported for light A, despite the fact that all these are visually matching white lights to all observers.

BREAK-UP OF ORDER IN COLOR SPACES

When, as above, the CIE Standard Observer computes very different chromaticities for lights (e.g., point A vs. point B of Fig. 4) that by consensus are a visual match, the tristimulus values RGB obviously fail to fulfil the criterion $R1 = R2; G1 = G2; B1 = B2$ required of a competent colorimetry or Standard Observer. The failure I refer to as “tristimulus error.” The examples above comprise the worst failures of the 2° and 10° CIE Standard Observers that we have found. Note that the Standard Observers suggest a color difference of 20–30 Simon–Goodwin jnds ($60-80 \Delta E^*_{ab}$), where no visible color difference exists. The Standard Observers assign a strong green color to lights that are seen by all our normal observers as white, as a visual match to the hueless reference white of point A.

Figure 5 shows six of the visually matching white lights of Fig. 4, plotted in the CIELAB diagram. So we see here two familiar color spaces, the chromaticity diagram and the

CIELAB diagram, “populated” with real visually matching white lights. The large and very practical problem is that the visually matching lights are spread alarmingly across both diagrams, instead of plotting at or very near the same point, as they should.

I would like to close with several comments:

1. Metamerism (spectral discrepancies between visually matching viewed lights) will continue, in the coloration industry, to become progressively stronger. While the extreme metamerism of the implied reflected lights of Fig. 2 may not soon become common in industrial coloration, our preference for bright colors in many applications is a force in that direction. Figure 6, taken from Rich and Jalijali,⁷ and Fig. 7, from Ralph Stanzola,⁸ represent current examples of reflectance spectra of colored surfaces that already generate highly metameric pairs of visually matching lights.
2. The color-matching functions that at present form the basis of our colorimetry (now \bar{r} , \bar{g} , and \bar{b} , or their transformed equivalents \bar{x} , \bar{y} , and \bar{z}) are, as demonstrated in this article and those of Ref. 4, stressed by strong metamerism to the point where they clearly fail. The troublesome effects of that stress on our present functions are seen in Figs. 4 and 5. Although the present article deals with bright, visually matching white lights, the failures of the CIE systems occur throughout color space: high and low visual brightness, white and saturated visually matching colored lights; see for example Fig. 24 of Part I of Ref. 4. The visual system judges

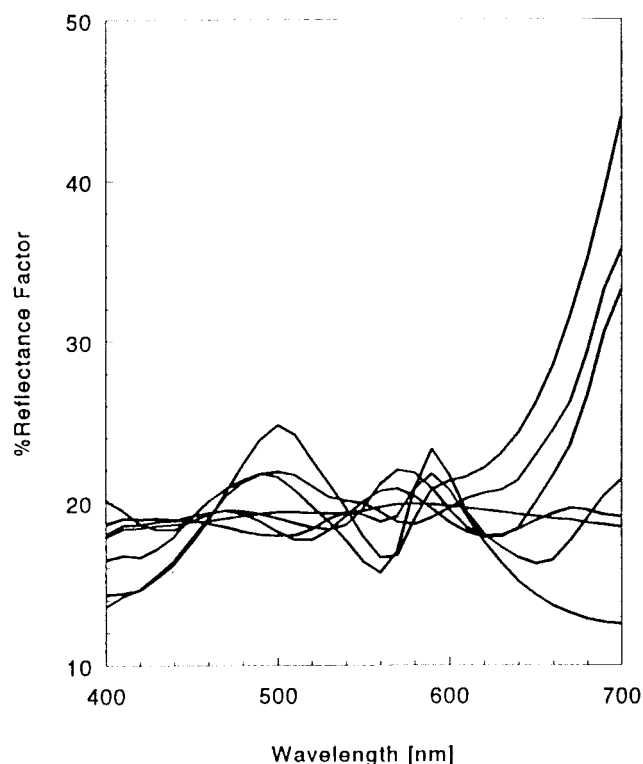


FIG. 6. Reflectance spectra for seven complex metameric gray painted papers. (From Rich and Jalijali, Ref. 8.)

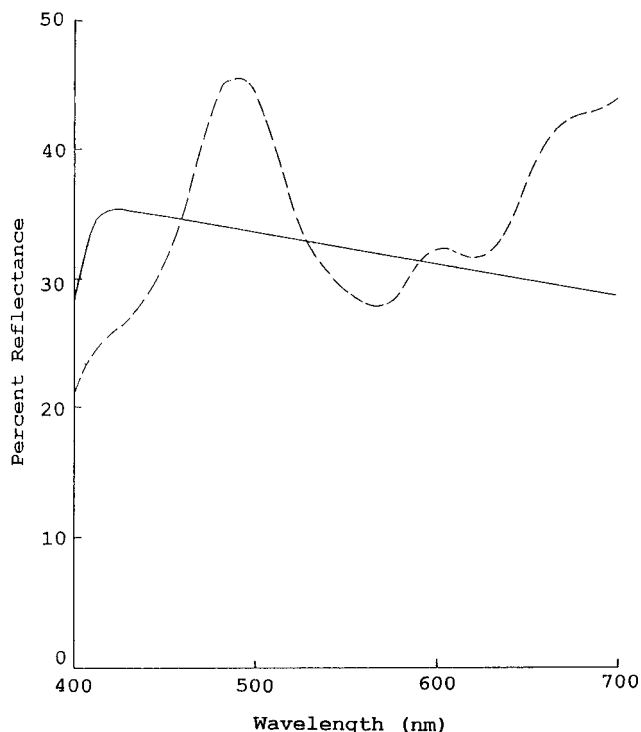


FIG. 7. Reflectance spectra for two complex metameric gray painted papers. (From Ralph Stanziola, Ref. 9.)

matches of incoming lights, whether they are known to be reflected from, and thus associated with, objects or not; “object colors” are also lights. The failures apply as well to object colors away from the white point.

3. Therefore, it is necessary that those color-matching functions be significantly improved, in the sense of substituting more accurately for the normal human visual system. That is the goal of the work of Ref. 4.
4. Only then will a point in a color space dependably represent a certain viewed (perceived) color, no matter how strong the metamerism between pairs of lights having that viewed color.
5. Color order systems comprising sets of real objects in a specified illuminant, and which are assembled (visually arranged) by normal observers as are the Munsell and OSA sets,* are safe from the type of trouble discussed here.
6. Color spaces, depending on mathematical functions of RGB, and in which RGB are computed by prevailing sets of color-matching functions (i.e., by prevailing

* Note that, although Wyszecki states⁹ that the Munsell set is “calibrated in terms of CIE 1931 (x, y, Y) — color coordinates” and the OSA set “in terms of CIE 1964 (x, y, Y)— color coordinates,” the spacing in both of these sets were selected visually and not by recourse to the CIE Standard Observers.

Standard Observers), are at great risk, which will worsen as metamerism in lights reflected from industry-colored objects strengthens.

7. In the presence of strong metamerism, both CIE Standard Observers do poorly. The three functions that define a Standard Observer fall too far short of properly representing the normal human visual system.
8. If the CIE Standard Observers fail in the presence of strong metamerism, they also fail where metamerism is weak, even if the failure is, in a given circumstance, hardly detectable.

My group, now ten in number, has been at work on this problem for nearly 10 years. We are progressing. We have 6 of a series of articles now completed, and four of these have been published.⁴ Wouldn't more of you like to join the work?

ACKNOWLEDGMENTS

I am grateful to Dr. Fred W. Billmeyer, Jr., Dr. Michael H. Brill, Mr. Hugh S. Fairman, Mr. William N. Hale, Dr. Danny C. Rich, and Mr. Ralph Stanziola for help and counsel in guiding me in regard to the complexities in the realm of color spaces and color order systems, and for critiquing the manuscript. I also wish to thank Danny Rich and Ralph Stanziola for allowing me to use, in Figs. 6 and 7, their examples of industrially colored surfaces that can lead to highly metameric visually matching lights.

1. Fred W. Billmeyer, Jr. and Max Saltzman, *Principles of Color Technology*, 2nd ed., Wiley, New York, 1981, p. 25.
2. G. Wyszecki, and W. S. Stiles, *Color Science*, First Edition (1967) and Second Edition (1982), Wiley, New York.
3. *Standard Terminology of Appearance*, ASTM Designation: E 284 – 96c (1996), American Society for Testing and Materials, West Conshohocken, PA.
4. W. A. Thornton, Toward a more accurate and extensible colorimetry. Part I. Introduction. The visual colorimeter- spectroradiometer. Experimental results. *Color Res. Appl.* **17**, 79–122 (1992); Part II. Discussion. *Color Res. Appl.* **17**, 162–186 (1992); Part III. Discussion (continued). *Color Res. Appl.* **17**, 240–262 (1992); Part IV. Visual experiments with bright fields and both 10° and 1.3° field sizes. *Color Res. Appl.* **22**, 189–198 (1997); Part V. Testing visually matching pairs of lights for possible rod participation on the Aguilar–Stiles model. *Color Res. Appl.* **23**, 92–103 (1998); Part VI. Improved weighting functions. Preliminary results. *Color Res. Appl.* **23**, 226–233(1998).
5. Color Aid Corporation, 37 East 18th Street, New York, NY 10003.
6. Fred W. Billmeyer, Jr. and Max Saltzman, *Principles of Color Technology*, 2nd ed., Wiley, New York, 1981, p. 49.
7. Danny C. Rich and Jocelyn Jalijali, Effects of observer metamerism in the determination of human color-matching functions, *Color Res. Appl.* **20**, 29–35 (1995).
8. Private communication, Ralph Stanziola, Industrial Color Technology.
9. Gunter Wyszecki, *Uniform color spaces, Golden Jubilee of Colour in the CIE*, Bradford, The Society of Dyers and Colourists, p. 52–77 (1981).